



Patrícia Aparecida de Souza

Estudo e Análise Comparativa das Técnicas de Redução de Dimensionalidade Lineares e Não Lineares

São José dos Campos, SP

Patrícia Aparecida de Souza

Estudo e Análise Comparativa das Técnicas de Redução de Dimensionalidade Lineares e Não Lineares

Trabalho de conclusão de curso apresentado ao
Instituto de Ciência e Tecnologia – UNIFESP,
como parte das atividades para obtenção do tí-
tulo de Bacharel em Ciência da Computação.

Universidade Federal de São Paulo – UNIFESP

Instituto de Ciência e Tecnologia

Bacharelado em Ciência da Computação

Orientador: Prof. Dr. Marcos Gonçalves Quiles

Coorientador: Prof. Dr. Thadeu Alves Senne

São José dos Campos, SP

Dezembro de 2018

Elaborado por sistema de geração automática com os dados fornecidos pelo(a) autor(a).

Aparecida , Patrícia

Estudo e Análise Comparativa das Técnicas de Redução de Dimensionalidade Lineares e Não Lineares/ Patrícia Aparecida

Orientador(a) Marcos Gonçalves Quiles; Coorientador(a) Thadeu Alves Senne- São José dos Campos, 2018.

142 p.

Trabalho de Conclusão de Curso-Bacharelado em Ciência da Computação- Universidade Federal de São Paulo-Instituto de Ciência e Tecnologia, 2018.

1. métodos de redução de dimensionalidade. I. Gonçalves Quiles, Marcos , orientador(a). II. Alves Senne, Thadeu,coorientador(a). III. Título.

Universidade Federal de São Paulo

Campus São José dos Campos
Departamento de Ciência e Tecnologia – DCT
Bacharelado em Ciência da Computação

TRABALHO DE CONCLUSÃO DE CURSO II

FORMULÁRIO DE AVALIAÇÃO

Aluno: Patrícia Aparecida de Souza

Orientador: Marcos Gonçalves Quiles

Coorientador: Thadeu Alves Senne

Título: Estudo e Análise Comparativa das Técnicas de Redução de Dimensionalidade Lineares e Não Lineares

Data da Apresentação: 06/12/2018

	0	1	2	3	4	5	6	7	8	9	10
Qualidade do texto											
Revisão bibliográfica											
Metodologia utilizada											
Resultados obtidos											
Assiduidade às reuniões com orientador											
Apresentação oral											

Nota Final: 10

Aprovado [X]

Aprovado condicionalmente¹ []

Reprovado []

Comentários adicionais: _____



Marcos Gonçalves Quiles



Márcio Porto Basgalupp

Thadeu Alves Senne



Ana Luísa Dine Martins Lemos

Este trabalho é dedicado em memória de minha amada mãe.

Agradecimentos

Primeiramente a Deus por me conceder o dom da vida e a oportunidade de desfrutar de toda sua criação. À minha mãe pelo amor e apoio incondicionais, por sua presença nos momentos difíceis pelo incentivo a continuar a caminhada. Aos amigos e colegas pelo incentivo e apoio. A Universidade Federal de São Paulo e a todos os docentes que me acompanharam durante toda a trajetória acadêmica com competência e profissionalismo. Ao meu caro orientador, Prof. Dr. Marcos Gonçalves Quiles, por sua paciência, dedicação e disponibilidade, ao querido coorientador, Prof. Dr. Thadeu Senne, e ao Prof. Dr. Marcelo Cristino Gama por seu auxílio na resolução de dúvidas quanto aos fundamentos matemáticos que nortearam e possibilitaram o desenvolvimento deste trabalho. Enfim, muito obrigada a todos.

*“Não temas porque eu sou contigo,
não te assombres, porque eu sou teu Deus,
eu te esforço, e te ajudo, e te sustento com a destra
da minha justiça.
(Bíblia Sagrada, Isaías 41,10)*

Resumo

O trabalho proposto, tem como objetivo, o estudo dos principais métodos de redução de dimensionalidade não lineares e dois dos métodos de redução lineares mais amplamente usados na literatura que são o Principal Components Analysis (PCA) e o Metric Multidimensional Scaling (MDS). Serão discutidos os fundamentos matemáticos subjacentes à cada um dos métodos e uma análise comparativa entre as técnicas será realizada utilizando conjuntos de dados naturais e artificialmente gerados.

Palavras-chaves: métodos de redução de dimensionalidade, Principal Components Analysis, Metric Multidimensional Scaling, Locally Linear Embedding, Hessian Locally Linear Embedding, Maximum Variance Unfolding, Diffusion Maps, Local Tangent Space Alignment, Laplacian Eigenmaps.

Abstract

The proposed work aims to study the main nonlinear dimensional reduction methods and two of the most widely used linear reduction methods in the literature, PCA and MDS. The mathematical foundations underlying each method will be discussed and a comparative analysis between the techniques will be performed using both natural and artificially generated data sets. Among the non-linear dimensionality reduction methods will be approached Isomap, Maximum Variance Unfolding, Locally Linear Embedding, Hessian Locally Linear Embedding, Local Tangent Space Alignment, Diffusion Maps and Laplacian Eigenmaps.

Key-words: methods of dimensionality reduction, Principal Components Analysis, Metric Multidimensional Scheduling, Locally Linear Embedding, Hessian Locally Linear Embedding, Maximum Variance Unfolding, Diffusion Maps, Local Tangent Space Alignment, Laplacian Eigenmaps.

Lista de ilustrações

Figura 1 – Matriz de <i>Coranking</i>	82
Figura 2 – Superfície <i>cylinder</i> reduzida com método <i>Laplacian Eigenmaps</i>	90
Figura 3 – Superfície <i>broken swiss</i> reduzida com método <i>Hessian Locally Linear Embedding</i>	91
Figura 4 – Superfície <i>swiss roll</i> reduzida com método <i>Isometric Mapping</i>	91

Lista de tabelas

Tabela 1 – Descrição dos conjuntos de dados naturais	85
Tabela 2 – Dimensionalidade para redução aplicada aos conjuntos naturais	87
Tabela 3 – Medidas de confiabilidade para superfícies artificiais	88
Tabela 4 – Medidas de continuidade para superfícies artificiais	88
Tabela 5 – Erro de generalização para superfícies artificiais	89
Tabela 6 – Medida independente de escala para superfícies artificiais	90
Tabela 7 – Teste de Friedman para conjuntos naturais	92
Tabela 8 – p -valores corrigidos	93
Tabela 9 – Menor dimensão de redução	94
Tabela 10 – Confiabilidade para <i>Hessian Locally Linear Embedding</i>	101
Tabela 11 – Continuidade para <i>Hessian Locally Linear Embedding</i>	102
Tabela 12 – Erro de generalização para <i>Hessian Locally Linear Embedding</i>	103
Tabela 13 – Medida independente de escala para <i>Hessian Locally Linear Embedding</i>	104
Tabela 14 – Confiabilidade para <i>Isometric Mapping</i>	105
Tabela 15 – Continuidade para <i>Isometric Mapping</i>	106
Tabela 16 – Erro de generalização para <i>Isometric Mapping</i>	107
Tabela 17 – Medida independente de escala para <i>Isometric Mapping</i>	108
Tabela 18 – Confiabilidade para <i>Laplacian Eigenmaps</i>	109
Tabela 19 – Continuidade para <i>Laplacian Eigenmaps</i>	110
Tabela 20 – Erro de generalização para <i>Laplacian Eigenmaps</i>	112
Tabela 21 – Medida independente de escala para <i>Laplacian Eigenmaps</i>	113
Tabela 22 – Confiabilidade para <i>Locally Linear Embedding</i>	114
Tabela 23 – Continuidade para <i>Locally Linear Embedding</i>	115
Tabela 24 – Erro de generalização para <i>Locally Linear Embedding</i>	116
Tabela 25 – Medida independente de escala para <i>Locally Linear Embedding</i>	117
Tabela 26 – Confiabilidade para <i>Local Tangent Space Alignment</i>	118
Tabela 27 – Continuidade para <i>Local Tangent Space Alignment</i>	119
Tabela 28 – Erro de generalização para <i>Local Tangent Space Alignment</i>	120
Tabela 29 – Medida independente de escala para <i>Local Tangent Space Alignment</i>	121
Tabela 30 – Confiabilidade para <i>Diffusion Maps</i>	123
Tabela 31 – Continuidade para <i>Diffusion Maps</i>	124
Tabela 32 – Erro de generalização para <i>Diffusion Maps</i>	125
Tabela 33 – Medida independente de escala para <i>Diffusion Maps</i>	126
Tabela 34 – Confiabilidade para <i>Maximum Variance Unfolding</i>	127

Tabela 35 – Continuidade para <i>Maximum Variance Unfolding</i>	128
Tabela 36 – Erro de generalização para <i>Maximum Variance Unfolding</i>	129
Tabela 37 – Medida independente de escala para <i>Maximum Variance Unfolding</i>	130
Tabela 38 – Confiabilidade para <i>Classical Multidimensional Scaling</i>	131
Tabela 39 – Continuidade para <i>Classical Multidimensional Scaling</i>	132
Tabela 40 – Erro de generalização para <i>Classical Multidimensional Scaling</i>	133
Tabela 41 – Medida independente de escala para <i>Classical Multidimensional Scaling</i> . .	134
Tabela 42 – Confiabilidade para <i>Principal Components Analysis</i>	136
Tabela 43 – Continuidade para <i>Principal Components Analysis</i>	137
Tabela 44 – Erro de generalização para <i>Principal Components Analysis</i>	138
Tabela 45 – Medida independente de escala para <i>Principal Components Analysis</i>	139

Lista de abreviaturas e siglas

MDS	Metric Multidimensional Scaling Escalonamento multidimensional métrico
CMDS	Classical Multidimensional Scaling Escalonamento multidimensional clássico
SDP	SemiDefinite Programming Programação semidefinida
DR	Dimensionality reduction Redução de dimensionalidade
LLE	Locally Linear Embedding Representação localmente linear
HLLE	Hessian Locally Linear Embedding Representação hessiana localmente linear
MVU	Maximum Variance Unfolding Desdobramento de máxima variância
LTSA	Local Tangent Space Alignment Alinhamento do espaço tangente local
PCA	Principal Components Analysis Análise de componentes principais
SDE	Semidefinite embedding Representação semidefinida
SVD	Singular Value Decomposition Decomposição em valores singulares
Dmaps	Diffusion Maps

	Mapas de difusão
Isomaps	Isometric Maps
	Mapas isométricos
LB operator	Laplace Beltrami operator
	Operador de Laplace Beltrami
Leigs	Laplacian Eigenmaps
	Auto mapas Laplacianos
i.i.d	independente e identicamente distribuído
o.n	ortonormal
o.g	ortogonal
psd	positive semidefinite
	semidefinida positiva
pd	positive definite
	definida positiva

Lista de símbolos

\mathcal{D}_n	matrizes diagonais reais $m \times n$
\mathcal{D}_n	matrizes diagonais reais $n \times n$
$\mathcal{M}_{m, n}$	matrizes reais $m \times n$
$\mathcal{M}_{m, n}(r)$	matrizes reais $m \times n$ de posto (r)
\mathcal{M}_n	matrizes reais $n \times n$
$\mathcal{M}_n(r)$	matrizes reais $n \times n$ de posto (r)
$\mathcal{O}_{m, n}$	matrizes reais $m \times n$ com colunas ortonormais se $m \geq n$ e com linhas ortonormais se $m < n$
\mathcal{O}_n	matrizes ortogonais reais $n \times n$
\mathcal{S}_n	matrizes semidefinidas positivas $n \times n$
$\mathcal{SP}_n(r)$	matrizes semidefinidas positivas $n \times n$ de posto r
\mathcal{P}_n	matrizes positivas definidas $n \times n$
\mathbf{R}^n	espaço Euclidiano n -dimensional, vetores reais de dimensão n
\mathbf{R}_+	números reais não negativos
\mathbf{Z}	números inteiros
\mathbf{Z}_+	números inteiros não negativos
\mathbf{N}	números naturais

Sumário

1	Introdução	19
1.1	Contextualização e motivação	19
1.2	Objetivos	24
1.3	Organização do trabalho	24
2	Fundamentos Matemáticos	25
2.1	Variedade Linear	25
2.1.1	Subespaço e Projeção	25
2.1.2	Funções sobre Espaços Euclidianos	27
2.2	Variedades Diferenciáveis	27
2.2.1	Espaço Tangente e Vetores Tangentes	28
2.2.2	Métrica Riemanniana	32
2.2.3	Distância Geodésica	33
2.3	Funções e Operadores sobre Variedades	34
2.3.1	Funções sobre Variedades	34
2.3.2	Operadores sobre Variedades	35
3	Métodos de Redução de Dimensionalidade	37
3.1	Principal Components Analysis	37
3.1.1	Descrição Geométrica do <i>Principal Components Analysis</i>	37
3.1.2	Descrição Estatística do PCA	40
3.1.3	Algoritmo <i>Principal Components Analysis</i>	42
3.2	<i>Classical Multidimensional Scaling</i>	44
3.2.1	Matrizes de Distâncias Euclidianas	44
3.2.2	Matriz de Gram sobre o Conjunto de Dados	45
3.2.3	Relação entre Distância Euclidiana e Matriz de Gram	45
3.2.4	Descrição do Método <i>Classical Multidimensional Scaling</i>	46
3.2.5	Algoritmo para <i>Classical Multidimensional Scaling</i>	47
3.3	Métodos de Redução não Lineares	48
3.3.1	Similaridade e Dissimilaridade	48
3.3.2	Definição de Vizinhança	48
3.3.3	Condições de Consistência	49
3.4	<i>Isometric Mapping</i>	50
3.4.1	Métrica Geodésica sobre Conjuntos Discretos	51
3.4.2	<i>Kernel</i> para <i>Isometric Mapping</i>	52
3.4.3	Algoritmo <i>Isometric Mapping</i>	52

3.5	<i>Locally Linear Embedding</i>	54
3.5.1	Algoritmo <i>Locally Linear Embedding</i>	55
3.6	<i>Laplacian Eigenmaps</i>	57
3.6.1	Algoritmo <i>Laplacian Eigenmaps</i>	58
3.7	<i>Local Tangent Space Alignment</i>	60
3.7.1	Coordenadas Tangentes e Coordenadas em Variedades	60
3.7.2	Representação de Coordenadas Local	61
3.7.3	Alinhamento Global	63
3.7.4	Algoritmo <i>Local Tangent Space Alignment</i>	64
3.8	<i>Hessian Locally Linear Embedding</i>	66
3.8.1	Hessiano sobre Variedades	66
3.8.2	Hessiano sobre Espaço Tangente	67
3.8.3	Construção do Funcional Hessiano	68
3.8.4	Construção do <i>kernel Hessian Locally Linear Embedding</i>	71
3.8.5	Algoritmo <i>Hessian Locally Linear Embedding</i>	72
3.9	<i>Diffusion Maps</i>	73
3.9.1	Operador de Difusão sobre a Variedade	73
3.9.2	Normalização de <i>kernel</i> de Difusão	74
3.9.3	Algoritmo <i>Diffusion Maps</i>	76
3.10	<i>Maximum Variance Unfolding</i>	77
3.10.1	Descrição do Método <i>Maximum Variance Unfolding</i>	77
3.10.2	Algoritmo <i>Maximum Variance Unfolding</i>	79
3.11	Métricas para avaliação da qualidade de redução de dimensionalidade	81
4	Experimentos	85
4.1	Seleção de conjuntos de dados naturais	85
4.2	Geração de conjuntos de dados artificiais	86
4.3	Seleção de parâmetros	86
5	Resultados e discussão para conjuntos de dados artificiais	88
5.1	Medidas de confiabilidade	88
5.2	Medidas de continuidade	88
5.3	Erro de generalização	88
5.4	Medida independente de escala	90
6	Resultados e discussão para conjuntos de dados naturais	92
7	Conclusão	95

Apêndices

98

.1	Método de redução <i>Hessian Locally Linear Embedding</i>	100
.1.1	Medidas de confiabilidade	100
.1.2	Medidas de continuidade	101
.1.3	Erro de generalização	102
.1.4	Medida independente de escala	103
.2	Método de redução <i>Isometric Mapping</i>	104
.2.1	Medidas de confiabilidade	104
.2.2	Medida de continuidade	105
.2.3	Erro de generalização	106
.2.4	Medida independente de escala	107
.3	Método de redução <i>Laplacian Eigenmaps</i>	108
.3.1	Medida de confiabilidade	108
.3.2	Medida de continuidade	110
.3.3	Erro de generalização	111
.3.4	Medida independente de escala	112
.4	Método de redução <i>Locally Linear Embedding</i>	113
.4.1	Medida de confiabilidade	113
.4.2	Medida de continuidade	114
.4.3	Erro de generalização	115
.4.4	Medida independente de escala	116
.5	Método de redução <i>Local Tangent Space Alignment</i>	117
.5.1	Medida de confiabilidade	117
.5.2	Medida de continuidade	118
.5.3	Erro de generalização	119
.5.4	Medida independente de escala	120
.6	Método de redução <i>Diffusion Maps</i>	122
.6.1	Medida de confiabilidade	122
.6.2	Medida de continuidade	123
.6.3	Erro de generalização	124
.6.4	Medida independente de escala	125
.7	Método de redução <i>Maximum Variance Unfolding</i>	126
.7.1	Medida de confiabilidade	126
.7.2	Medida de continuidade	127
.7.3	Erro de generalização	128
.7.4	Medida independente de escala	129
.8	Método de redução <i>Classical Multidimensional Scaling</i>	130
.8.1	Medida de confiabilidade	130
.8.2	Medida de continuidade	131

.8.3	Erro de generalização	132
.8.4	Medida independente de escala	134
.9	Método de redução <i>Principal Components Analysis</i>	135
.9.1	Medida de confiabilidade	135
.9.2	Medida de continuidade	136
.9.3	Erro de generalização	137
.9.4	Medida independente de escala	138
Referências	140

1 Introdução

1.1 Contextualização e motivação

Métodos de redução de dimensionalidade surgiram a partir da necessidade de se tratar, manipular e processar dados de alta dimensionalidade como são comuns os provenientes de áreas como análise de sinais, processamento de imagens, visão computacional e aprendizado de máquina. Os métodos de redução de dimensionalidade têm várias aplicações, como por exemplo, a compressão e a visualização de dados multidimensionais. Estes métodos baseiam-se na aplicação de transformações sobre os dados projetando-os em espaços de menor dimensão com a máxima manutenção da topologia original. Diversos métodos de redução foram propostos com o objetivo de solucionar a seguinte questão: como descobrir a estrutura geométrica, de baixa dimensionalidade, subjacente a dados de alta dimensionalidade, de modo que, os dados originalmente em um espaço de alta dimensão possam ser projetados em um novo espaço de menor dimensão, de forma a minimizar a perda de informação? Para responder a esta questão, foram desenvolvidos vários métodos que podem ser agrupados em duas categorias principais: métodos lineares e não lineares. Os métodos lineares assumem como hipótese que os dados estão sobre um subespaço linear de baixa dimensão, enquanto os não lineares assumem que os dados estão sobre uma variedade de baixa dimensão. Matematicamente, o problema de redução de dimensionalidade pode ser formulado da seguinte forma. Dado um conjunto de dados representado em uma matriz \mathbf{X} , de dimensão $D \times n$, consistindo de n vetores \mathbf{x}_i , $i \in \{1, \dots, n\}$ com dimensionalidade D . Assumimos que a dimensionalidade intrínseca do conjunto de dados é d (onde frequentemente $d \ll D$). Supondo que o conjunto de dados \mathbf{X} esteja sobre uma variedade M de dimensão d a qual está imersa em \mathbb{R}^D : $M \subset \mathbb{R}^D$. Então chamamos D de dimensão extrínseca de \mathbf{X} e d de dimensão intrínseca partindo do pressuposto de que dados de alta dimensão residem sobre uma variedade de baixa dimensão. Os métodos de redução transformam um conjunto de dados de dimensionalidade D para um novo conjunto de dados com dimensionalidade reduzida d onde a estrutura geométrica dos dados originais seja preservada ao máximo possível. De forma geral, a dimensão intrínseca e a estrutura geométrica dos dados é desconhecida. A dimensionalidade intrínseca dos dados, ou dimensionalidade topológica, é a dimensão base da aproximação local linear de uma hiper superfície sobre a qual os dados estão imersos, por exemplo, o espaço tangente. Se o conjunto de dados está sobre uma subvariedade de dimensão d então seu espaço tangente tem dimensão d para cada ponto do conjunto. Métodos para estimação da dimensionalidade intrínseca ou topológica dos dados podem ser encontrados em (FUKUNAGA, 1982) e (CAMASTRA, 2003). No trabalho proposto, serão estudados dois dos principais métodos de redução lineares que são o *Principal Components Analysis* (PCA) e o *Metric Multidimensional Scaling* (MDS). Entre os métodos de redução não lineares, serão abor-

dados os seguintes: *Isometric Maps* (ISOMAP), *Maximum Variance Unfolding* (MVU), *Diffusion Maps*, *Locally Linear Embedding* (LLE), *Hessian Locally Linear Embedding* (HLLE), *Local Tangent Space Alignment* (LTSA) e *Laplacian Eigenmaps*. No capítulo 3 é apresentada uma descrição detalhada das técnicas de redução abordadas. O *Principal Components Analysis* (PCA) foi inicialmente proposto por (PEARSON, 1901) e o termo componentes principais foi cunhado pela primeira vez por (HOTELLING, 1933). O objetivo do PCA é encontrar projeções lineares que maximizem a preservação da variância nos dados. As direções das projeções são encontradas pela decomposição espectral da matriz de covariância dos dados. Os autovetores da matriz de covariância correspondem às direções, nas quais os dados originalmente em alta dimensão, apresentam maior variabilidade. Assim, os autovetores correspondentes aos d maiores autovalores descrevem o subespaço de dimensão d que contém a maior variância. O *Metric Multidimensional Scaling* pode ser descrito como uma família de métodos para análise de proximidade de dados multidimensionais. Os métodos de *Multidimensional Scaling* tomam como entrada uma matriz de medidas de proximidades ou afinidades e como saída fornecem uma configuração de pontos, em um espaço Euclidiano de baixa dimensão, de forma que as medidas de proximidades dadas sejam o máximo possível aproximadas pelas distâncias entre os pontos da configuração de saída. Os métodos de *Multidimensional Scaling* originaram-se de pesquisas realizadas na área de psicometria. Psicometria é um campo da psicologia cujo objetivo é a medida dos processos psíquicos. Vários foram os pesquisadores que fizeram importantes contribuições para o desenvolvimento de abordagens para os métodos de escalonamento multidimensionais. Os métodos de *Multidimensional Scaling* são classificados em métrico e não métrico de acordo com o tipo de variáveis para qual as medidas de proximidades são computadas. Os métodos de *Multidimensional Scaling* métrico operam com variáveis quantitativas enquanto os métodos de *Multidimensional Scaling* não métrico operam com variáveis ordinais.

O primeiro método de *Multidimensional Scaling* métrico, foi proposto e descrito por (TORGERSON, 1952), no qual o Torgerson (1952), assumia que as medidas de proximidades eram aproximadamente uma função linear de um conjunto de distâncias. A formulação do *Multidimensional Scaling* métrico teve sua base teórica fundamentada no trabalho de (YOUNG; HOUSEHOLDER, 1938) que demonstrou como determinar a dimensionalidade mínima necessária para acomodar um conjunto de distâncias euclidianas computadas entre n pontos. O trabalho de Young e Householder (1938) também propôs um método para construção do espaço e também foi capaz de determinar quando as distâncias podiam ser representadas em qualquer espaço Euclidiano. O procedimento de *Multidimensional Scaling*, criado por (TORGERSON, 1952), pode ser resumido na execução de três etapas. Na primeira, uma escala de distâncias comparativas entre todos os pares de estímulos era obtida. A segunda etapa envolvia a estimativa de uma constante aditiva e o seu uso na conversão das distâncias comparativas em distâncias absolutas. A terceira e última etapa, a dimensionalidade do espaço psicológico necessário para acomodar as distâncias absolutas era determinada e as projeções dos estímulos sobre os eixos desse espaço eram obtidas.

Para o desenvolvimento de métodos *Multidimensional Scaling* não métrico dois pesquisadores contribuíram de forma significativa. O primeiro, [Shepard \(1962a\)](#), propôs uma abordagem para o problema do escalonamento multidimensional em que nenhuma hipótese era feita sobre a forma da função de distância utilizada para o cálculo das medidas de proximidades. A única exigência era que a função fosse monotônica. Seu trabalho visou atingir três objetivos: o primeiro encontrar o número mínimo de dimensões em um espaço Euclidiano dado que as distâncias nesse espaço fossem monotonicamente relacionadas com as medidas de proximidades dadas; o segundo objetivo encontrar um conjunto de coordenadas ortogonais para os pontos nesse espaço de baixa dimensão; e o terceiro objetivo mostrar a forma verdadeira da função de distância inicialmente desconhecida relacionando as medidas de proximidades com as medidas de distâncias. Em [Shepard \(1962b\)](#), foi proposto um método geral de escalonamento multidimensional no qual a convergência para uma configuração final, no espaço de baixa dimensionalidade, era atingida através de um processo iterativo regido por dois parâmetros principais: α que era responsável pelo controle da taxa de abordagem para a condição de monotonicidade, por exemplo, a condição na qual a classificação das distâncias fosse o inverso da classificação das medidas de proximidades e β que controlava a taxa na qual a configuração inicial era colapsada em um espaço de dimensão mínima.

Os trabalhos de [Kruskal e Hill \(1964\)](#) e [Kruskal \(1964\)](#), também fizeram relevantes contribuições para o desenvolvimento de métodos de *Multidimensional Scaling* não métrico. A técnica proposta pelo [Kruskal e Hill \(1964\)](#), tinha seus fundamentos no problema de regressão de mínimos quadrados. O problema do escalonamento multidimensional era visto como uma questão de ajuste estatístico, onde as medidas de proximidades eram fornecidas e o objetivo era encontrar uma configuração cujas distâncias melhor se "ajustassem" às medidas de proximidades experimentais fornecidas. Para alcance de tal objetivo adotou-se uma relação monótona entre as medidas de proximidades experimentais e as distâncias. Para resumo da técnica proposta, dada uma configuração em um espaço t -dimensional, com t fixado e conhecido, realizava-se uma regressão monótona de distâncias sobre as proximidades e usava-se a variância residual, adequadamente normalizada, como medida quantitativa. Essa medida foi chamada de *stress*. Dessa forma, para qualquer configuração fornecida, o *stress* media o qual bem essa configuração correspondia aos dados originais. Quanto menor fosse o valor da medida *stress* melhor a configuração se ajustaria aos dados, o que significava um relacionamento monótono perfeito entre medidas de proximidades e distâncias. Para o cálculo do *stress*, admitia-se como hipóteses que os dados eram inerentemente simétricos, auto similaridades eram ignoradas e empates desconsiderados. O objetivo da técnica era representar os n pontos no espaço t -dimensional e avaliar o quão bem a configuração representava os dados. Neste trabalho, foi abordado o método de *Multidimensional Scaling* métrico, também conhecido como *Multidimensional Scaling* clássico, o qual toma como entrada uma matriz de similaridades ou afinidades computada a partir dos dados de entrada, utilizando para tal a distância euclidiana tomada entre todos os pares de dados. O método de *Multidimensional Scaling* clássico e o *Principal Components Analysis*

possuem algumas características em comum, como por exemplo, o fato da representação dos dados no espaço de baixa dimensão ser obtida por meio da decomposição espectral da matriz de Gram para o *Multidimensional Scaling* e da matriz de covariância para o *Principal Components Analysis*. Enquanto o *Principal Components Analysis* procura uma representação na qual seja maximizada a preservação da variância o *Multidimensional Scaling* procura a preservação das distâncias dentro do espaço de baixa dimensão.

Os métodos de redução não lineares, que foram abordados, partem do pressuposto de que os dados estão imersos em uma variedade de baixa dimensionalidade. Para desempenhar a análise dos dados que residem sobre essa variedade de baixa dimensão, as matrizes (*kernel*) são construídas a partir de grafos ponderados cujos vértices representam os exemplos de entrada e cujas arestas definem relações de vizinhança.

O *Isometric Maps* (ISOMAP), proposto por (TENENBAUM; SILVA; LANGFORD, 2000), é um método que procura preservar a distância geodésica entre os pontos enquanto mapeia os dados observados em menos dimensões. A distância geodésica é definida como o comprimento do caminho mais curto entre dois pontos sobre a superfície de uma variedade. Como os dados em processamento são discretos o algoritmo faz uma estimativa das distâncias geodésicas encontrando o caminho mais curto sobre o grafo entre todos os pares de pontos, que podem ser computados eficientemente utilizando o algoritmo de Dijkstra (DIJKSTRA, 1959).

O *Maximum Variance Unfolding* (MVU), (WEINBERGER; PACKER; SAUL, 2005) e (WEINBERGER; SAUL, 2004), formalmente conhecido como representação semidefinida é um método baseado sobre o pressuposto de que quando a variedade no qual os dados estão imersos é "desdobrada" a variância sobre os pontos é maximizada. Inicialmente o algoritmo constrói um grafo a partir do conjunto de dados e define uma vizinhança para cada ponto e então maximiza a variância sobre o conjunto de dados completo de forma que as distâncias e ângulos locais entre os pares de todos os vizinhos de cada ponto sejam preservados.

Proposto por Roweis e Saul (2000), o método de redução *Locally Linear Embedding* (LLE) apresenta uma abordagem distinta com relação aos métodos ISOMAP E MVU descritos anteriormente. Enquanto estes últimos são baseados na preservação da estrutura global da variedade, o LLE propõe uma projeção local dos dados de forma a preservar a geometria e a linearidade local subjacente à variedade M . Inicialmente o LLE constrói um grafo de vizinhança G cujo objetivo é conectar apenas os pontos mais próximos de cada vértice x_i , de modo que x_i e seus vizinhos estejam contidos em uma pequena vizinhança próxima ao espaço tangente da variedade M . De maneira a usufruir da linearidade local, o método computa para cada vértice do grafo sua melhor aproximação como uma combinação linear de seus vizinhos.

O método *Local Tangent Space Alignment* (LTSA), proposto por Zhang e Zha (2004), da mesma forma que o LLE, considera que o conjunto de dados é amostrado de uma variedade suave, então os vizinhos de cada ponto permanecem próximos quando projetados em um espaço de dimensão inferior. A abordagem adotada pelo LTSA é ligeiramente distinta da utilizada

pelo LLE. No LLE, cada ponto no conjunto de dados é linearmente representado como uma combinação linear de seus vizinhos mais próximos. Já para o LTSA, as coordenadas locais tangentes são construídas aplicando-se o método PCA sobre a vizinhança de cada vértice do grafo de vizinhança G . As coordenadas locais tangentes são consideradas como uma aproximação do espaço tangente em cada ponto da variedade M . Como o espaço tangente provê uma representação em coordenadas da vizinhança, as coordenadas dadas são uma representação dos dados no espaço de baixa dimensão. Para que sejam obtidas as coordenadas, em baixa dimensão, para todos os pontos do conjunto de dados, o LTSA executa um procedimento de alinhamento de coordenadas de forma a tornar a representação local em global.

O método *Laplacian Eigenmaps*, proposto por, [Belkin e Niyogi \(2003\)](#), de maneira análoga aos métodos anteriores, inicia-se pela construção do grafo ponderado G . Os pesos definem uma medida de proximidade entre os pontos e são computados usando o *kernel* gaussiano. A solução fornecida pelo método é obtida a partir da discretização do operador de Laplace Beltrami, onde as coordenadas do espaço dimensionalmente reduzido são dadas pela resolução de um problema generalizado de autovetores e autovalores envolvendo o Laplaciano do grafo.

O método *Hessian Locally Linear Embedding* (HLLE) proposto por, [Donoho e Grimes \(2003\)](#), assume como hipótese que a variedade sobre a qual os dados residem é localmente isométrica a um subconjunto aberto e conectado de \mathbb{R}^d . A solução fornecida pelo método é obtida a partir de uma aproximação para o espaço nulo do operador Hessiano. É importante destacar que o método é custoso computacionalmente, uma vez que estimativas de diferenciais de segunda ordem são computadas. Para que o método seja eficiente na redução de dimensionalidade é necessário que os dados sejam suficientemente densos.

O método *Diffusion Maps* proposto por, [Coifman e Lafon \(2006\)](#), é semelhante ao método *Laplacian Eigenmaps*. Inicialmente é construído um grafo ponderado, onde os pesos das arestas são computados por meio da discretização do *kernel* do calor de Neumann, que pode ser aproximado pelo *kernel* Gaussiano. Em um passo seguinte, é realizada a normalização dos pesos e define-se o que chamamos de distâncias de difusão. O objetivo do método é realizar a redução de dimensionalidade de modo a minimizar as distâncias de difusão computadas entre todos os pares de vértices do grafo. Os princípios utilizados para a construção teórica do método *diffusion maps* possuem conexões com diversas outras áreas do conhecimento como análise de *clustering* espectral e análise espectral sobre variedades. É uma área de pesquisa que ainda possui diversas questões em aberto. Para maiores detalhes sobre as questões pertinentes ao arcabouço matemática subjacente ao método o leitor pode consultar as seguintes referências ([COIFMAN et al., 2005](#)), ([LAFON; LEE, 2006](#)), ([LAFON, 2004](#)).

1.2 Objetivos

O trabalho proposto teve como objetivo o estudo dos métodos de redução de dimensionalidade lineares e não lineares, e a análise comparativa do desempenho destes métodos quando aplicados a conjuntos de dados reais e artificiais. Para a avaliação do desempenho, dos diversos métodos estudados foi utilizado as métricas de confiabilidade, continuidade, erro de generalização de um classificador 1-NN e a medida independente de escala. Dentre os métodos lineares, foram estudados o *Principal Components Analysis* e o *Metric Multidimensional Scaling*. Dos métodos não lineares foram estudados os seguintes: *Isometric Maps* (ISOMAP), *Locally Linear Embedding* (LLE), *Laplacian Eigenmaps* (LE), *Local Tangent Space Alignment* (LTSA), *Hessian Locally Linear Embedding* (HLLE) e *Diffusion Maps*.

1.3 Organização do trabalho

O trabalho está organizado do seguinte modo: o capítulo 2 apresenta os fundamentos matemáticos subjacentes aos métodos que foram abordados e que irão dar ao leitor a fundamentação necessária para a compreensão dos métodos abordados. O capítulo 3 descreve de forma mais aprofundada cada um dos métodos discutidos durante a execução deste trabalho. O capítulo 4 apresenta as etapas de seleção dos conjuntos de dados naturais, a geração dos conjuntos de dados artificiais e a seleção de parâmetros utilizada. O capítulo 5 apresenta os resultados e discussões obtidas a partir dos experimentos realizados sobre os conjuntos de dados artificiais. O capítulo 6 apresenta os resultados obtidos a partir dos experimentos realizados sobre os conjuntos de dados naturais e por fim o capítulo 7 apresenta as conclusões decorrentes do estudo e experimentos realizados durante a execução do trabalho proposto. A fundamentação teórica do trabalho proposto baseou-se principalmente no seguinte livro texto (WANG, 2012).

2 Fundamentos Matemáticos

2.1 Variedade Linear

O exemplo mais simples de variedade é uma variedade linear conhecida como hiperplano. Uma variedade linear é uma ampliação local de uma variedade não linear. Na verdade, em cada ponto de uma variedade não linear, existe um espaço tangente, que localmente aproxima a variedade. O espaço tangente é uma variedade linear. Métodos de redução de dimensionalidade lineares são baseados na hipótese de que os dados observados residem sobre uma variedade linear. No estudo de variedades lineares, a ferramenta principal é a Álgebra Linear. A seguir são revisados alguns conceitos importantes de Álgebra Linear.

As coordenadas de um ponto \mathbf{x} no espaço Euclidiano \mathbb{R}^D são denotadas por $\mathbf{x} = [x_1 \dots x_D]'$. Um ponto também é chamado de vetor. Seja $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{R}^D$ um conjunto finito de pontos. Este conjunto será representado na forma de uma matriz $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_n] = [x_{ij}]_{i,j=1}^{D,n}$. O conjunto de dados \mathcal{X} gera um subespaço de \mathbb{R}^D denotado por $S = \text{span}\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, que é o espaço coluna da matriz de dados \mathbf{X} . Definimos o espaço coluna da matriz \mathbf{M} por $\text{Col}(\mathbf{M})$ e o espaço linha de \mathbf{M} por $\text{Row}(\mathbf{M})$. Consequentemente, $\text{span}\{\mathbf{x}_1, \dots, \mathbf{x}_n\} = \text{Col}(\mathbf{X})$.

A distância Euclidiana entre dois pontos $\mathbf{x}, \mathbf{y} \in \mathbb{R}^D$ é indicada por $d_2(\mathbf{x}, \mathbf{y})$ ou $\|\mathbf{x} - \mathbf{y}\|$, e o produto interno de \mathbf{x} e \mathbf{y} é indicado por $\langle \mathbf{x}, \mathbf{y} \rangle$ ou $\mathbf{x} \cdot \mathbf{y}$.

A transposta da matriz \mathbf{A} é indicada por \mathbf{A}' ou \mathbf{A}^T . Usando a notação transposta, o produto interno $\mathbf{x} \cdot \mathbf{y}$ também pode ser escrito como $\mathbf{x}'\mathbf{y}$ ou $\mathbf{x}^T\mathbf{y}$.

2.1.1 Subespaço e Projeção

Seja $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_k\}$ uma base ortonormal para o subespaço de dimensão k onde $S \subset \mathbb{R}^m$. Escreva $\mathbf{B} = [\mathbf{b}_1 \dots \mathbf{b}_k] \in \mathcal{O}_{m,k}$. A projeção ortogonal $\mathbf{P} : \mathbb{R}^m \rightarrow S$ pode ser representada por

$$\mathbf{y} = \mathbf{P}(\mathbf{x}) = \mathbf{B}\mathbf{B}'\mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^m.$$

Para um vetor $\mathbf{y} \in S$,

$$\mathbf{y} = \mathbf{P}(\mathbf{y}) = \mathbf{B}\mathbf{B}'\mathbf{y} = \mathbf{B}\mathbf{y}_{[\mathcal{B}]}$$

onde $\mathbf{y}_{[\mathcal{B}]} \stackrel{\text{def}}{=} \mathbf{B}'\mathbf{y}$ é chamado de \mathcal{B} -coordenadas de \mathbf{y} , ou seja, são as coordenadas de \mathbf{y} escritas na base \mathcal{B} . Como a dimensão de S é k , S é isometricamente isomorfo a \mathbb{R}^k , e a matriz

\mathbf{B} define uma representação (*embedding*) ortogonal de \mathbb{R}^k para \mathbb{R}^m por

$$\mathbf{y} = \mathbf{Bz}, \quad \mathbf{z} \in \mathbb{R}^k.$$

Temos que

$$d_2(\mathbf{Bz}_1, \mathbf{Bz}_2) = d_2(\mathbf{z}_1, \mathbf{z}_2), \quad \mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^k,$$

onde a distância sobre o lado esquerdo é medida em \mathbb{R}^m enquanto a distância sobre o lado direito é medida em \mathbb{R}^k . Seja $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{R}^m$ e $S = \text{span}\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$. Então a projeção ortogonal \mathbf{P} tem a seguinte forma

$$\mathbf{y} = \mathbf{P}(\mathbf{x}) = \mathbf{X}\mathbf{X}^-\mathbf{x},$$

onde \mathbf{X}^- define a inversa generalizada de \mathbf{X} . Quando o conjunto de vetores \mathcal{X} é linearmente dependente, a inversa generalizada \mathbf{X}^- não é única. Caso contrário, \mathbf{X}^- é única e igual à inversa esquerda de \mathbf{X} , que pode ser representada por $\mathbf{X}^- = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

Para encontrar uma base ortonormal de $S = \text{Col}(\mathbf{X})$, nós fazemos a fatoração \mathbf{QR} de \mathbf{X} :

$$\mathbf{X} = \mathbf{QR}, \tag{2.1}$$

onde $\mathbf{Q} = \mathbf{u}_1 \dots \mathbf{u}_k \in \mathcal{D}_{m,k}$ e \mathbf{R} é uma matriz $k \times n$ triangular superior. O conjunto $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ é uma base ortonormal de S . Portanto, $\mathbf{Q}\mathbf{Q}'$ é a projeção ortogonal de \mathbb{R}^m para S . A inversa generalizada de \mathbf{X} pode ser obtida por

$$\mathbf{X}^- = (\mathbf{R}'\mathbf{R})^{-1}\mathbf{X}'.$$

Seja $\mathbf{a} \in \mathbb{R}^m$ conhecido. O produto interno $\mathbf{a} \cdot \mathbf{x}$ define uma função linear $L_{\mathbf{a}}(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x}$ sobre \mathbb{R}^m , que também é conhecido como um funcional linear sobre \mathbb{R}^m . O espaço de todos os funcionais lineares sobre um espaço linear S é chamado de espaço dual de S e é indicado por S^* .

Um hiperplano é um subespaço transladado. Seja $S \subset \mathbb{R}^m$ um subespaço de dimensão k e $\mathbf{a} \in \mathbb{R}^m$. Então a translação de S através de \mathbf{a} é o hiperplano

$$H = \mathbf{a} + S = \{\mathbf{x} + \mathbf{a} : \mathbf{x} \in S\}.$$

Por outro lado, cada hiperplano $H \subset \mathbb{R}^m$ tem um subespaço paralelo único S , que pode ser construído como segue. Escolha um vetor $\mathbf{a} \in H$, então

$$S = H - \mathbf{a} = \{\mathbf{x} - \mathbf{a} : \mathbf{x} \in H\}.$$

A dimensão do hiperplano é igual à dimensão de seu subespaço paralelo.

2.1.2 Funções sobre Espaços Euclidianos

Cada vetor $\mathbf{a} \in \mathbb{R}^m$ define um funcional linear dado por

$$L_{\mathbf{a}}(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x} \rangle, \quad \mathbf{x} \in \mathbb{R}^m, \quad (2.2)$$

que é uma função linear homogênea sobre \mathbb{R}^m . Neste trabalho, os termos função linear homogênea e funcional linear serão usados como sinônimos. A função constante $L_c(\mathbf{x}) = c$, $c \in \mathbb{R}$ é uma função linear não homogênea.

Seja $\{\mathbf{e}_1, \dots, \mathbf{e}_m\}$ a base canônica ortonormal de \mathbb{R}^m (também chamada de base coordenada). Então cada \mathbf{e}_i define um funcional com $L_{\mathbf{e}_i}(\mathbf{x}) = x_i$, que frequentemente é indicado pela δ -notação $\delta_i : \delta_i(\mathbf{x}) = L_{\mathbf{e}_i}(\mathbf{x})$, e chamado a i -ésima função coordenada. O conjunto $\{1, \mathbf{e}_1, \dots, \mathbf{e}_m\}$ é uma base do espaço de funções lineares sobre \mathbb{R}^m . Tome $S = \text{span}(\mathcal{X})$. Pode-se encontrar uma base de S usando a decomposição QR de \mathbf{X} . Ela também pode ser obtida da decomposição de valores singulares de \mathbf{X} :

$$\mathbf{X} = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T,$$

onde $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ é uma base ortonormal de S .

2.2 Variedades Diferenciáveis

Nesta seção, é apresentada algumas definições relevantes acerca da teoria de variedades. Estas definições tem o intuito de facilitar a compreensão pelo leitor dos métodos abordados.

Definição 2.1. *Seja $X \subset \mathbb{R}^k$ e $Y \subset \mathbb{R}^l$ dois conjuntos arbitrários não vazios. Se uma bijeção $f : X \rightarrow Y$ e sua inversa f^{-1} são ambas suaves, então f é chamada de homeomorfismo diferenciável (ou difeomorfismo), e X é dito ser difeomorfo a Y .*

Definição 2.2. *Seja $M \subset \mathbb{R}^m$ não vazio. Se, para cada ponto $\mathbf{x} \in M$, existe um conjunto aberto $W \subset M$ tal que W é um difeomorfismo para um conjunto aberto $U \subset \mathbb{R}^k$, então M é chamada de variedade diferenciável de dimensão k , um difeomorfismo $g : U \rightarrow W$ é chamado de parametrização de W , e sua inversa $h(= g^{-1}) : W \rightarrow U$ é chamado um mapa coordenado, W é chamado de vizinhança coordenada, e ao par (W, h) chamamos de sistema de coordenadas*

locais ou uma carta sobre M . Em um sistema de coordenadas (W, h) , um ponto $\mathbf{x} \in W$ tem coordenadas

$$h(\mathbf{x}) = [h^1(\mathbf{x}), \dots, h^m(\mathbf{x})],$$

onde h^i é chamada a i -ésima função coordenada.

Definição 2.3. Uma estrutura diferenciável (ou um atlas) \mathcal{D} sobre uma variedade M de dimensão k é uma coleção de todos os sistemas de coordenadas locais $\{(W_i, h_i)\}$ sobre M , satisfazendo as seguintes condições.

- A união de W_i cobre M : $M \subset \bigcup_i W_i$.
- Para cada par (i, j) , $h_j \circ h_i^{-1}$ é um mapeamento suave sobre $h_i(W_i \cap W_j)$.

Em topologia e ramos correlatos da matemática, um espaço de Hausdorff é definido como um espaço topológico em que pontos distintos possuem vizinhanças disjuntas. Portanto uma k -variedade pode ser considerada como um espaço de Hausdorff, que é localmente homeomórfico a \mathbb{R}^k e possui uma base contável.

Definição 2.4. Seja M uma variedade d -dimensional e \mathcal{D} uma estrutura diferenciável sobre M . Um subconjunto $S \subset M$ é chamado de k -subvariedade de M se, para cada sistema de coordenadas locais $(W, h) \in \mathcal{D}$, $h(S \cap W)$ é um difeomorfismo para um conjunto aberto $O \subset \mathbb{R}^k$.

Definição 2.5. Seja \mathcal{D} uma estrutura diferenciável sobre M . Uma função f sobre um conjunto aberto $V \subset M$ é chamado diferenciável (ou suave) se, para cada sistema de coordenadas locais $(W, h) \in \mathcal{D}$, $f \circ h^{-1}$ é diferenciável sobre $h(V \cap W)$.

Indicamos a classe de funções diferenciáveis sobre $V \subset M$ por $\mathcal{F}(V)$. Seja $f, q \in \mathcal{F}(V)$ e $c \in \mathbb{R}$. Então cf , $f + q$, e fq estão em $\mathcal{F}(V)$. A diferenciabilidade de uma função sobre V não depende da escolha de um sistema de coordenadas em particular sobre M .

Se existe um sistema de coordenadas único (M, h) sobre uma variedade M tal que para cada ponto $\mathbf{x} \in M$ tem as coordenadas $h(\mathbf{x})$, então nós chamamos M de uma variedade simples. Conjuntos de dados de alta dimensionalidade comumente podem ser modelados como subconjuntos de variedades simples. Consequentemente, é assumido que a variedade subjacente de um conjunto de dados de alta dimensão é simples, a menos que seja afirmado o contrário.

2.2.1 Espaço Tangente e Vetores Tangentes

Seja $M \subset \mathbb{R}^m$ uma variedade. Por um ponto $\mathbf{p} \in M$, é construído o hiperplano H que melhor "aproxima" M em uma vizinhança de \mathbf{p} . Intuitivamente, o espaço tangente $T_{\mathbf{p}}$ é o subespaço paralelo ao hiperplano H . Para um conjunto aberto $U \subset \mathbb{R}^k$, seu espaço tangente

$T_{\mathbf{x}}U$, $\mathbf{x} \in U$, é trivialmente definido como o espaço completo \mathbb{R}^k . Para definir o espaço tangente para uma variedade não linear, primeiramente relembremos o conceito de derivada de um mapa diferenciável $f = [f_1, \dots, f_l] : U \rightarrow V$, onde U e V são conjuntos abertos de \mathbb{R}^k e \mathbb{R}^l respectivamente. Em $\mathbf{x} \in U$, a derivada de f é uma aplicação linear $df_{\mathbf{x}} : \mathbb{R}^k \rightarrow \mathbb{R}^l$ definida por

$$df_{\mathbf{x}} = \lim_{t \rightarrow 0} \frac{f(\mathbf{x} + t\mathbf{h}) - f(\mathbf{x})}{t}, \quad \mathbf{h} \in \mathbb{R}^k, \quad (2.3)$$

que pode ser representada pela matriz

$$df_{\mathbf{x}} = \left[\frac{\partial f_i(\mathbf{x})}{\partial x_j} \right]_{i,j=1}^{l,k} = \begin{pmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_k} \\ \vdots & & \vdots \\ \frac{\partial f_l(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_l(\mathbf{x})}{\partial x_k} \end{pmatrix} \quad (2.4)$$

Particularmente, quando $l = 1$, f é reduzida para a função $f : U \rightarrow \mathbb{R}$, e a derivada de f é identificada com o gradiente de f :

$$df_{\mathbf{x}} = \nabla f_{\mathbf{x}} = \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_k} \right].$$

Derivadas de mapas suaves têm as seguintes propriedades.

Lema 2.1. 1. Se $f : U \rightarrow V$ e $q : V \rightarrow W$ ambas são mapas suaves e $f(\mathbf{x}) = \mathbf{y}$, então

$$d(q \circ f)_{\mathbf{x}} = dq_{\mathbf{y}} df_{\mathbf{x}}.$$

2. Se $U \subset \tilde{U}$ é um mapa de inclusão, por exemplo, $i(\mathbf{x}) = \mathbf{x}$ para cada $\mathbf{x} \in U$, então $di_{\mathbf{x}}$ é o mapa identidade sobre \mathbb{R}^k .

3. Se $L : \mathbb{R}^k \rightarrow \mathbb{R}^l$ é uma transformação linear, representada pela matriz \mathbf{L} tal que $L(\mathbf{x}) = \mathbf{L}\mathbf{x}$, $\mathbf{x} \in \mathbb{R}^k$, então a derivada de L em cada \mathbf{x} é \mathbf{L} : $dL_{\mathbf{x}} = \mathbf{L}$.

Pelas propriedades de derivadas, podemos verificar que se $U \subset \mathbb{R}^k$, $V \subset \mathbb{R}^l$, e $f : U \rightarrow V$ é um difeomorfismo, então $k = l$ e $df_{\mathbf{x}}$ é não degenerada, por exemplo, a matriz $df_{\mathbf{x}}$ é invertível. O resultado é parcialmente reversível de acordo com o Teorema da Função Inversa apresentado em 2.2.

Teorema 2.2. Se a derivada $df_{\mathbf{x}}$ é não degenerada, então existe uma vizinhança U em torno de \mathbf{x} tal que f é um difeomorfismo de U para $f(U)$.

O espaço tangente de uma variedade é definido de acordo com 2.6.

Definição 2.6. Seja $M \subset \mathbb{R}^m$ uma variedade k -dimensional, $U \subset \mathbb{R}^k$ um conjunto aberto, e $g : U \rightarrow M$ como a parametrização da vizinhança $g(U) \subset M$. Dado que $\mathbf{p} \in M$, $\mathbf{u} \in U$, e

$g(\mathbf{u}) = \mathbf{p}$. Então a imagem da transformação linear $\mathbf{d}g_{\mathbf{u}}$, denotada por $T_{\mathbf{p}}M \stackrel{\text{def}}{=} \mathbf{d}g_{\mathbf{u}}(\mathbb{R}^k)$, é chamado de espaço tangente de M em \mathbf{p} . Um vetor em $T_{\mathbf{p}}M$ é chamado de vetor tangente.

Denota-se o espaço tangente $T_{\mathbf{p}}M$ de forma simplificada por $T_{\mathbf{p}}$ se não é necessário enfatizar-se a variedade M . Porque $\mathbf{d}g_{\mathbf{u}}$ é uma matriz de posto completo, $T_{\mathbf{p}}M$ é um subespaço k -dimensional de \mathbb{R}^k .

Por conveniência, chamamos o hiperplano $H_{\mathbf{p}} \stackrel{\text{def}}{=} \mathbf{p} + T_{\mathbf{p}}$ o espaço tangente de M (em \mathbf{p}), de $H_{\mathbf{p}}$ que é exatamente tangente a M .

Escrevemos

$$\frac{\partial g}{\partial U^i} = \left[\frac{\partial g_1}{\partial U^i}, \dots, \frac{\partial g_m}{\partial U^i} \right]^T, \quad 1 \leq i \leq k.$$

Pela definição 2.6, o conjunto

$$\left\{ \frac{\partial g}{\partial u^1}, \dots, \frac{\partial g}{\partial u^k} \right\} \tag{2.5}$$

é uma base para $T_{\mathbf{p}}M$. A matriz base é definida por

$$\frac{\partial g}{\partial u} \stackrel{\text{def}}{=} \left[\frac{\partial g_i}{\partial u^j} \right]_{i,j=1}^{m,k} = \begin{pmatrix} \frac{\partial g_1}{\partial u^1} & \dots & \frac{\partial g_1}{\partial u^k} \\ \frac{\partial g_2}{\partial u^1} & \dots & \frac{\partial g_2}{\partial u^k} \\ \vdots & & \vdots \\ \frac{\partial g_m}{\partial u^1} & \dots & \frac{\partial g_m}{\partial u^k} \end{pmatrix}. \tag{2.6}$$

Então um vetor tangente em \mathbf{p} tem a forma

$$\mathbf{X}_{\mathbf{p}} = \sum_{i=1}^k \alpha_i \frac{\partial g}{\partial u^i} = \frac{\partial g}{\partial u} \vec{\alpha}, \tag{2.7}$$

onde $\vec{\alpha} = [\alpha_1, \dots, \alpha_k]' \in \mathbb{R}^k$. Se $\vec{\alpha} \neq 0$, ele representa a direção em \mathbb{R}^k e $\mathbf{X}_{\mathbf{p}}$ é a direção no espaço tangente $T_{\mathbf{p}}$. Seja um mapa isométrico g , então a matriz $\frac{\partial g}{\partial u}$ tem colunas ortonormais, então temos $\|\mathbf{X}_{\mathbf{p}}\| = \|\vec{\alpha}\|$.

Assumimos que f é uma função sobre M , que possui uma extensão suave sobre um conjunto aberto O de \mathbb{R}^m tal que $M \subset O$. Então a derivada de f é computada como uma função sobre $O \subset \mathbb{R}^m$. Usando g , constrói-se a função composta $f \circ g$ sobre \mathbb{R}^k . Por outro lado, como uma função sobre M , f pode também ser representada como uma função de coordenadas u ,

$F(u)$. Estas duas funções são idênticas: $f \circ g(u) = F(u)$. Então a derivada direcional de F na direção de $\vec{\alpha} \in \mathbb{R}^k$ pode ser computada por

$$\frac{\partial f \circ g}{\partial \vec{\alpha}} = df_{\mathbf{p}} \frac{\partial g}{\partial u} \vec{\alpha}. \quad (2.8)$$

Agora considerando a mudança de coordenadas sobre o espaço tangente. Assumindo como hipótese que $g_1 : U \rightarrow M$ e $g_2 : V \rightarrow M$ são duas parametrizações tal que $g_1(\mathbf{u}) = \mathbf{p}$ e $g_2(\mathbf{v}) = \mathbf{p}$. O espaço tangente de M em \mathbf{p} pode ser representado por cada uma delas. A parametrização g_1 gera a base de $T_{\mathbf{p}}M$ através da matriz

$$\frac{\partial g_1}{\partial u} \stackrel{\text{def}}{=} \left[\frac{\partial g_{1,i}}{\partial u^j} \right]_{i,j=1}^{m,k}.$$

e a parametrização g_2 gera a seguinte base para $T_{\mathbf{p}}M$ através da seguinte matriz

$$\frac{\partial g_2}{\partial v} \stackrel{\text{def}}{=} \left[\frac{\partial g_{2,i}}{\partial v^j} \right]_{i,j=1}^{m,k}.$$

Seja $\mathbf{X}_{\mathbf{p}}$ o vetor tangente sobre $T_{\mathbf{p}}M$ tal que

$$\mathbf{X}_{\mathbf{p}} = \sum_{i=1}^k \alpha_i \frac{\partial g}{\partial u^i} = \frac{\partial g}{\partial u} \vec{\alpha}.$$

Assumindo o uso da base de $\left\{ \frac{\partial g_2}{\partial v^j} \right\}_{j=1}^k$, temos

$$\mathbf{X}_{\mathbf{p}} = \sum_{i=1}^k \beta_i \frac{\partial g_2}{\partial v^i} = \frac{\partial g_2}{\partial v} \vec{\beta}.$$

Temos que

$$\vec{\beta} = \left(\frac{\partial g_2}{\partial v} \right)^{-} = \frac{\partial g_1}{\partial u} \vec{\alpha},$$

onde $\left(\frac{\partial g_2}{\partial v} \right)^{-}$ é a inversa-esquerda de $\frac{\partial g_2}{\partial v}$. Correspondentemente para g_1 e g_2 , temos os mapas coordenados h_1 e h_2 respectivamente. Então o mapa $h_2 \circ g_1 : U \rightarrow V$ é um difeomorfismo e $h_2 \circ g_1(\mathbf{u}) = \mathbf{v}$. O mapa linear $d(h_2 \circ g_1)_{\mathbf{u}}$ é uma transformação linear não degenerada sobre \mathbb{R}^k . Pelo Lema 2.1, temos que

$$d(h_2 \circ g_1)_{\mathbf{u}} = (dh_2)_{\mathbf{p}}(dg_1)_{\mathbf{u}}.$$

Como $h_2 = g_2^{-1}$,

$$(dh_2)_\mathbf{p} = \left\{ \frac{\partial g_2}{\partial v} \right\}^-.$$

Teorema 2.3. *Seja g_1 e g_2 duas parametrizações de M e $T_\mathbf{p}M$ o espaço tangente de M em \mathbf{p} . Seja $\frac{\partial g_1}{\partial u}$ e $\frac{\partial g_2}{\partial v}$ duas bases de $T_\mathbf{p}M$ correspondentes a g_1 e g_2 respectivamente. Seja $\mathbf{X}_\mathbf{p}$ o vetor tangente com as coordenadas $\vec{\alpha}$ associadas com $\frac{\partial g_1}{\partial u}$ e as coordenadas $\vec{\beta}$ associadas com $\frac{\partial g_2}{\partial v}$. Então*

$$\vec{\beta} = \frac{\partial v}{\partial u} \vec{\alpha}, \quad (2.9)$$

$$\frac{\partial g_2}{\partial v} = \frac{\partial g_1}{\partial u} \frac{\partial u}{\partial v}, \quad (2.10)$$

onde

$$\frac{\partial v}{\partial u} = d((g_2)^{-1} \circ g_1)_\mathbf{u}, \quad \frac{\partial u}{\partial v} = d((g_1)^{-1} \circ g_2)_\mathbf{v}. \quad (2.11)$$

O Teorema 2.3 mostra a importância do espaço tangente na análise local de funções sobre a variedade. Para tratarmos com uma função sobre uma variedade, é usado sua derivada e trabalha-se sobre o espaço tangente.

2.2.2 Métrica Riemanniana

Seja M uma variedade k -dimensional e $T_\mathbf{p}$ seu espaço tangente em $\mathbf{p} \in M$. Dado que g é uma parametrização de M e (W, h) é um sistema de coordenadas sobre M , onde $h = g^{-1}$. Para cada $\mathbf{p} \in M$, tome $\mathbf{u} = h(\mathbf{p}) \in \mathbb{R}^k$, então g define a base em 2.5 sobre $T_\mathbf{p}M$, que pode ser representada pela base em 2.6. A métrica $\mathbf{G}_g(\mathbf{p})$ é definida por

$$\mathbf{G}_g(\mathbf{p}) = \left(\frac{\partial g}{\partial u} \right)^T \frac{\partial g}{\partial u}, \quad \mathbf{p} \in M. \quad (2.12)$$

A $\mathbf{G}_g(\mathbf{p})$ é uma matriz positiva definida, chamada de métrica Riemanniana sobre M . Seja dois vetores tangentes $\mathbf{X}, \mathbf{Y} \in T_\mathbf{p}$ representados por $\mathbf{X} = \frac{\partial g}{\partial u} \mathbf{x}$ e $\mathbf{Y} = \frac{\partial g}{\partial u} \mathbf{y}$ respectivamente. O produto interno é definido por

$$\langle \mathbf{X}, \mathbf{Y} \rangle_\mathbf{p} = \mathbf{x}^T \mathbf{G}(\mathbf{p}) \mathbf{y}. \quad (2.13)$$

Lema 2.4. *Para quaisquer vetores tangentes $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \in T_\mathbf{p}$ e números reais $a, b \in \mathbb{R}$, as seguintes propriedades são válidas.*

$$\text{Linearidade: } \langle a\mathbf{X} + b\mathbf{Y}, \mathbf{Z} \rangle = a\langle \mathbf{X}, \mathbf{Z} \rangle + b\langle \mathbf{Y}, \mathbf{Z} \rangle.$$

Simetria: $\langle \mathbf{X}, \mathbf{Y} \rangle_{\mathbf{p}} = \langle \mathbf{Y}, \mathbf{X} \rangle_{\mathbf{p}}$,

Positividade: $\langle \mathbf{X}, \mathbf{X} \rangle_{\mathbf{p}} \geq 0$, e a igualdade é válida se e somente se $\mathbf{X} \neq \mathbf{0}$.

O produto interno define a norma de \mathbf{X} por

$$\|\mathbf{X}\|_{\mathbf{p}}^2 = \mathbf{x}^T \mathbf{G}(\mathbf{p}) \mathbf{x}. \quad (2.14)$$

Uma variedade Riemanniana é definida abaixo.

Definição 2.7. *Seja M uma variedade k -dimensional. Se M é dotada com a métrica $\mathbf{G}_g(\mathbf{p})$, e para cada par de vetores tangentes, o produto interno é definido por 2.13, então M é chamada de variedade Riemanniana e denotada por (M, g) ou (M, \mathbf{G}) .*

Neste trabalho, uma métrica Riemanniana simplesmente será chamada de métrica e uma variedade Riemanniana de variedade. Salientamos que o produto interno definido independe da métrica adotada e é independente da parametrização de M .

Teorema 2.5. *Seja g_1 e g_2 duas parametrizações de M e $T_{\mathbf{p}}M$ o espaço tangente de M em \mathbf{p} . Seja \mathbf{X} e \mathbf{Y} dois vetores tangentes sobre $T_{\mathbf{p}}M$ e tomando g_1 e g_2 , suas coordenadas são $\mathbf{x}_1, \mathbf{y}_1$, e $\mathbf{x}_2, \mathbf{y}_2$, respectivamente. Então*

$$\mathbf{x}_1^T \mathbf{G}_{g_1}(\mathbf{p}) \mathbf{y}_1 = \mathbf{x}_2^T \mathbf{G}_{g_2}(\mathbf{p}) \mathbf{y}_2, \quad (2.15)$$

e a fórmula para a métrica, agora, é dada por

$$\mathbf{G}_{g_2}(\mathbf{p}) = \left(\frac{\partial u}{\partial v} \right)_{\mathbf{p}}^T \mathbf{G}_{g_1}(\mathbf{p}) \left(\frac{\partial u}{\partial v} \right)_{\mathbf{p}}. \quad (2.16)$$

Em muitas aplicações, a parametrização de g é escolhida como sendo isométrica, então temos que $\mathbf{G}_g(\mathbf{p}) = \mathbf{I}$. Nestes casos, a métrica sobre a variedade é similar à métrica Euclidiana.

2.2.3 Distância Geodésica

O comprimento de arco sobre uma variedade M é definido em 2.17.

Definição 2.8. *Seja M uma variedade Riemanniana conectada e $\gamma \subset M$ é uma curva sobre M com a equação paramétrica $\gamma = [\gamma^1(t), \dots, \gamma^m(t)]'$, $t \in [a, b]$. Então o comprimento da curva γ é definido por*

$$\|\gamma\| = \int_a^b \|\dot{\gamma}\| dt = \int_a^b \sqrt{\sum_i^m (\dot{\gamma}^i)^2} dt, \quad (2.17)$$

onde $\dot{\gamma}$ é a derivada de γ com relação a t .

O comprimento de arco é independente da escolha da equação paramétrica da curva. Particularmente, pode-se escolher o comprimento de arco como o parâmetro da curva. A curva é parametrizada pelo comprimento de arco se e somente se $\|\dot{\gamma}(t)\| = 1$ para todo $t \in [a, b]$.

Seja \mathbf{p} e \mathbf{q} dois pontos em M . A distância geodésica entre eles é definida por

$$d_g(\mathbf{p}, \mathbf{q}) = \inf \|\dot{\gamma}\|,$$

onde o ínfimo estende sobre todas as curvas suaves conectando \mathbf{p} e \mathbf{q} .

2.3 Funções e Operadores sobre Variedades

2.3.1 Funções sobre Variedades

Seja $M \subset \mathbb{R}^m$ uma variedade k -dimensional, g uma parametrização de M e $\mathbf{u} = [u^1, \dots, u^k]$ o correspondente sistema coordenado sobre M . Então existe um domínio $U \subset \mathbb{R}^k$ tal que $h(M) = U$. Por conveniência, u é identificado com um ponto em U . Com o uso de coordenadas sobre M , pode-se representar uma função f sobre M como uma função sobre o domínio Euclidiano $U : f : U \rightarrow \mathbb{R}, f(u) = f(\mathbf{x}(u))$, onde $\mathbf{x} \in M$ com $u(\mathbf{x}) = u \in \mathbb{R}^k$. Como o domínio de f agora é Euclidiano, por exemplo, temos a fórmula de Taylor de $f(u)$ em torno do ponto $u_0 \in U$ com $g(u_0) = \mathbf{x}_0$:

$$f(u) = \sum_{j=1}^k \frac{1}{j!} (D\Delta u)^j f(u_0) + \frac{1}{(k+1)!} (D\Delta u)^{k+1} f(u_0 + \theta \Delta u), \quad 0 < \theta < 1,$$

onde $\Delta u = u - u_0$.

Denomina-se f como uma função linear (sobre as coordenadas u) se $f(u) = c + \sum_{i=1}^k a_i u^i = c + \mathbf{u}\mathbf{a}$, onde $\mathbf{a} = [a_1, \dots, a_k]' \in \mathbb{R}^k$, e f quadrática se

$$f(u) = c + \sum_{i=1}^k a_i u^i + \sum_{i,j=1}^k b_{ij} u^i u^j = c + \mathbf{u}\mathbf{a} + \mathbf{u}\mathbf{B}\mathbf{u}' \quad ,$$

onde $\mathbf{B} = [b_{ij}]$ é uma matriz $k \times k$.

A i -ésima derivada parcial de f é indicada por $\frac{\partial f}{\partial u^i}$. Em muitas aplicações, a extensão de f sobre um conjunto aberto $O \subset \mathbb{R}^m, M \subset O$, é conhecida, então pode-se computar a derivada parcial de f por 2.8. Seja $g(\mathbf{x}) = u$. Tem-se que

$$\partial_i f = Df_{\mathbf{x}} \frac{\partial g}{\partial u^i} = \sum_{j=1}^m \frac{\partial f}{\partial x_j} \frac{\partial g_j}{\partial u^i}.$$

2.3.2 Operadores sobre Variedades

Seja a métrica Riemanniana derivada de g indicada por $\mathbf{G} \stackrel{\text{def}}{=} \mathbf{G}_g$, que é uma matriz-função sobre M . Como ela é definida positiva, $|\mathbf{G}| \stackrel{\text{def}}{=} \det(\mathbf{G}) > 0$. Por conveniência, escrevamos a inversa $\mathbf{G}^{-1} = [g^{ij}]$. O gradiente de f sobre M com relação a g é um campo vetorial definido por

$$\text{grad } f = [(\text{grad } f)^1, \dots, (\text{grad } f)^k], \quad (2.18)$$

com

$$(\text{grad } f)^i = \sum_{j=1}^k g^{ij} \partial_j f.$$

O divergente de um campo vetorial $\mathbf{X} = [X^1, \dots, X^k]$ sobre M é definido por

$$\text{div } \mathbf{X} = \frac{1}{\sqrt{|\mathbf{G}|}} \sum_{i=1}^k \partial_i (\sqrt{|\mathbf{G}|} X^i). \quad (2.19)$$

Definição 2.9. O operador de Laplace-Beltrami sobre $C^2(M)$ é definido por

$$\Delta f = \text{div grad } f = \sum_{i=1}^k \sum_{j=1}^k \frac{1}{\sqrt{|\mathbf{G}|}} \partial_i (\sqrt{|\mathbf{G}|} g^{ij} \partial_j f). \quad (2.20)$$

Todos estes operadores são dependentes do sistema de coordenadas. Em métodos de redução, frequentemente é escolhido g para ser isométrico. Então $|\mathbf{G}| = 1$ e $g_{ij} = g^{ij} = \delta_{ij}$, de modo que estas fórmulas são reduzidas para expressões mais simples:

$$\text{grad } f = \left[\frac{\partial}{\partial u^1}, \dots, \frac{\partial}{\partial u^k} \right]$$

$$\text{div } \mathbf{X} = \sum_{i=1}^k \partial_i X^i, \quad e$$

$$\Delta f = \sum_{i=1}^k \frac{\partial^2 f}{\partial (u^i)^2}.$$

Finalmente, é discutido o espaço de funções sobre M . Temos como hipótese que M é conectada, limitada e orientada. Seja g uma parametrização de M , que produz a métrica \mathbf{G} , e $g(M) = U \subset \mathbb{R}^k$. O volume de M é definido por

$$\text{Vol}(M) = \int_U \sqrt{|\mathbf{G}|} du. \quad (2.21)$$

Usando a fórmula para métrica em 2.16, é possível provar que a definição de volume dada em 2.21 é independente da seleção de coordenadas. Consequentemente, é adotada a notação livre de coordenadas para reescrever 2.21 como

$$\text{Vol}(M) = \int_M 1.$$

Assumindo que M tem um único mapa coordenado h cujo domínio $U \subset \mathbf{R}^k$ é conectada, limitada e orientada. Assumindo também que h é suave sobre o conjunto fechado $U \cup \partial U$. Então M é conectada, limitada e orientada.

Em geral, para uma função f suave sobre M , a integral de f sobre M é definida por

$$\int_M f = \int_U f \sqrt{\mathbf{G}} du. \quad (2.22)$$

Como a integral é independente da escolha das coordenadas, novamente é aplicada a notação livre de coordenadas sobre o lado direito da equação 2.22. A função f definida sobre M é chamada integrável quadrada se $|f|^2$ é integrável. O espaço de todas as funções integráveis quadradas sobre M , indicado por $H(M)$, é um espaço de Hilbert munido com produto interno

$$\langle f, h \rangle = \int_M fh.$$

O teorema seguinte pode ser derivado do Teorema da Divergência.

Teorema 2.6. *Seja $\mathbf{X} = [X^1, \dots, X^k]^T$ como o campo vetorial sobre M e f uma função compactamente suportada sobre M tal que $f = 0$ sobre ∂M . Então a seguinte equação é verdadeira.*

$$\int_M \text{div } f\mathbf{X} = 0.$$

Pelo Teorema 2.6, é obtido o seguinte.

Teorema 2.7. *Seja f e q funções compactamente suportadas sobre M . Então*

$$\int_M f \Delta q = \int_M q \Delta f, \quad (2.23)$$

$$\int_M \|\text{grad } f\|^2 = - \int_M f \Delta f. \quad (2.24)$$

3 Métodos de Redução de Dimensionalidade

3.1 Principal Components Analysis

Proposto por [Pearson \(1901\)](#), o *Principal Components Analysis* dentre os métodos de redução de dimensionalidade lineares é o mais relevante e amplamente utilizado. O método assume que as afinidades entre pontos de um conjunto de dados são definidas usando-se a distância Euclidiana entre os pontos. Métodos de redução lineares assumem que o sistema de vizinhança é global, ou seja, os vizinhos de um ponto pertencente ao conjunto de dados consiste de todos os demais. Dado um conjunto de dados $\mathcal{X} = \mathbf{x}_1, \dots, \mathbf{x}_n \subset \mathbb{R}^D$ e o conjunto reduzido, após a aplicação do método, \mathcal{Y} e ainda usando como métrica a distância Euclidiana, o *Principal Components Analysis* procura por projeções lineares ortogonais $T : \mathbb{R}^D \rightarrow \mathbb{R}^d$ dado que os dados reduzidos $\mathcal{Y} = T(\mathcal{X})$ maximize a energia contida no conjunto original. O *Principal Components Analysis* pode ser analisado sob o ponto de vista geométrico e estatístico. Nas subseções 3.1.1 e 3.1.2 é abordado maiores detalhes.

3.1.1 Descrição Geométrica do *Principal Components Analysis*

Sabendo que translações não alteram a geometria dos dados, seja $\mathcal{X} = \mathbf{x}_1, \dots, \mathbf{x}_n \subset \mathbb{R}^D$ o conjunto de dados fornecido e $\mathbf{a} \in \mathbb{R}^D$. Denota-se por $\mathcal{X}_{\mathbf{a}}$, ou $\mathcal{X} - \mathbf{a}$, a translação de \mathcal{X} :

$$\mathcal{X}_{\mathbf{a}} = \{ \mathbf{x} - \mathbf{a} \in \mathbb{R}^D : \mathbf{x} \in \mathcal{X} \}$$

Qualquer translação de \mathcal{X} é geometricamente equivalente a própria \mathcal{X} , sendo assim, em *Principal Components Analysis* sempre transladamos \mathcal{X} para seu centro geométrico

$$\bar{\mathbf{x}} = \sum_{i=1}^n \mathbf{x}_i$$

O conjunto de dados centrado é definido por $\widehat{\mathcal{X}}$. Seja a energia de \mathcal{X} dada por

$$\mathcal{E}(\mathcal{X}) = \sum_{i=1}^n \|\mathbf{x}_i\|^2 = \|\mathbf{X}\|_F^2$$

onde $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_n]$ é a matriz de dados e $\|\mathbf{X}\|_F^2$ é a norma de Frobenius de \mathbf{X} . Então os dados centrados $\widehat{\mathcal{X}}$ que tem a energia mínima entre todas as translações de \mathcal{X} :

$$\mathcal{E}(\widehat{\mathcal{X}}) = \min_{\mathbf{a} \in \mathbb{R}^D} \mathcal{E}(\mathcal{X}_{\mathbf{a}})$$

Por simplicidade, é assumido que o conjunto de dados original \mathcal{X} já está centrado, então, $\widehat{\mathcal{X}} = \mathcal{X}$. Dessa forma o *Principal Components Analysis* é um método que procura um subespaço onde os dados estão mais concentrados. Para descrever a concentração de \mathcal{X} , considera-se o seguinte problema de maximização: sobre qual reta a projeção de \mathcal{X} terá a maior proporção de energia preservada? Para responder a esta questão, denotamos a reta por \mathbf{S}_1 . Como os dados \mathcal{X} estão centrados, a reta \mathbf{S}_1 deve passar pela origem, ela é um subespaço de dimensão 1 de \mathbb{R}^D . Denota-se a direção de \mathbf{S}_1 pelo vetor unitário \mathbf{v}_1 , o operador de projeção ortogonal $\mathbb{R}^D \rightarrow \mathbf{S}_1$ por $T_{\mathbf{v}_1}$, e o conjunto de todas as direções em \mathbb{R}^D pela esfera unitária $\mathbb{S}^{D-1} \subset \mathbb{R}^D$. Então o vetor \mathbf{v}_1 é a solução do seguinte problema de maximização:

$$\mathbf{v}_1 = \arg \max_{\mathbf{a} \in \mathbb{S}^{D-1}} \mathcal{E}(T_{\mathbf{a}}(\mathcal{X})) \quad (3.1)$$

Temos $\mathbf{Y}_1 = [y_{1,1}, \dots, y_{1,n}]' = \mathbf{T}_{\mathbf{a}}(\mathcal{X}) = [\mathbf{T}_{\mathbf{a}}(\mathbf{x}_1), \dots, \mathbf{T}_{\mathbf{a}}(\mathbf{x}_n)]' \in \mathbb{R}^n$.

Chama-se $\mathbf{v}_1 \in \mathbb{S}^{D-1}$ de primeira direção principal, e Y_1 é a primeira componente principal de \mathcal{X} . Por indução matemática, definimos as sucessivas direções principais e as correspondentes componentes principais como segue. Assumimos que as primeiras $s-1$ direções principais $\mathcal{V}_{s-1} = [\mathbf{v}_1, \dots, \mathbf{v}_{s-1}] \subset \mathbb{S}^{D-1}$ e as correspondentes $d-1$ componentes principais $Y_1, \dots, Y_{s-1} \subset \mathbb{R}^n$ de \mathcal{X} são bem definidas, onde \mathcal{V}_{s-1} é um sistema ortonormal em \mathbb{R}^D .

Definimos

$$\mathcal{X}_{s-1} = \left\{ \mathbf{x}_i - \sum_{j=1}^{s-1} \mathbf{v}_j y_{j,i}, 1 \leq i \leq n \right\}$$

e $S_{s-1} = \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_{s-1})$, $s = 2, 3, \dots$

Então a s -ésima direção principal é definida por:

$$\mathbf{v}_s = \arg \max_{\mathbf{a} \in S_{s-1}^\perp \cap \mathbb{S}^{D-1}} \mathcal{E}(\mathbf{T}_{\mathbf{a}}(\mathcal{X}_{s-1})) \quad s = 2, 3, \dots, \quad (3.2)$$

e a s -ésima componente principal de \mathcal{X} é $Y_s = T_{\mathbf{v}_s}(\mathcal{X})$. Portanto, a concentração pode ser explicada matematicamente pelas direções e componentes principais.

Os problemas de maximização 3.1 e 3.2 podem ser convertidos nos seguintes problemas de minimização:

$$\mathbf{v}_1 = \arg \min_{\mathbf{a} \in \mathbb{S}^{D-1}} \mathcal{E}((\mathbf{I} - \mathbf{T}_{\mathbf{a}})(\mathcal{X})) \quad (3.3)$$

e

$$\mathbf{v}_s = \arg \min_{\mathbf{a} \in S_{s-1}^\perp \cap \mathbb{R}^{D-1}} \mathcal{E}((\mathbf{I} - \mathbf{T}_a)(\mathcal{X}_{s-1})) \quad (3.4)$$

Realizando indução matemática em 3.2 sobre d , temos que a indução indica que (1) $\mathcal{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_d\}$ é uma base ortonormal do subespaço de dimensão d $S_d \subset \mathbb{R}^D$, e (2)

$$\mathbf{y}_i = \sum_{j=1}^d \mathbf{v}_j y_{i,j} \quad (3.5)$$

é a projeção ortogonal de \mathbf{x}_i sobre S_d de tal modo que o conjunto de dados $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ preserva a energia maximal de \mathcal{X} sobre todos os dados d -dimensional. Chama-se \mathcal{Y} uma d redução de dimensionalidade de \mathcal{X} . As soluções \mathcal{V} e \mathcal{Y} podem ser obtidas com o auxílio de métodos de álgebra numérica computacional.

Teorema 3.1 (Mirsky (1960), Eckart e Young (1936)). *Seja a decomposição de valor singular (SVD) da matriz $\mathbf{X} \in \mathfrak{M}_{D,n}(r)$ como*

$$\mathbf{X} = \mathbf{V}\Sigma\mathbf{U}' = \sum_{i=1}^r \sigma_i \mathbf{v}_i (\mathbf{u}_i)' \quad (3.6)$$

onde $\mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_r] \in \mathfrak{D}_{n,r}$, $\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_r] \in \mathfrak{D}_{D,r}$ e

$\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r)$ é a matriz diagonal com $\sigma_1, \sigma_2, \dots, \sigma_r (> 0)$.

Seja

$$\mathbf{X}_* = \sum_{i=1}^d \sigma_i \mathbf{v}_i (\mathbf{u}_i)', d \leq r \quad (3.7)$$

então $\mathbf{X}_* \in \mathfrak{M}_{D,n}(d)$ em 3.7 é a melhor aproximação d -posto de \mathbf{X} sobre a norma de Frobenius, dada por,

$$\|\mathbf{X} - \mathbf{X}_*\|_F = \min_{\mathbf{B} \in \mathfrak{M}_{D,n}(r)} \|\mathbf{X} - \mathbf{B}\|_F$$

com o erro dado por

$$\|\mathbf{X} - \mathbf{X}_*\|_F = \sqrt{\sum_{l=k+1}^r \sigma_l^2}$$

Do Teorema 3.1, podemos obter:

Teorema 3.2. *Seja $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \in \mathbb{R}^D$ o conjunto de dados dado (centrado) e a SVD de \mathbf{X} dada em (3.6). Escreva $\mathcal{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_d\} \in \mathbb{R}^D$, $\Sigma_d = \text{diag}(\sigma_1, \dots, \sigma_d)$, e $\mathbf{U}_d = [\mathbf{u}_1 \dots \mathbf{u}_d]$. Então as d -direções principais de \mathcal{X} são \mathcal{V} , e as \mathcal{V} -coordenadas das d -componentes principais de \mathcal{X} são os vetores linha da matriz*

$$\mathbf{Y} (\stackrel{\text{def}}{=} [\mathbf{y}_1 \dots \mathbf{y}_n]) = \Sigma_d (\mathbf{U}_d)' \quad (3.8)$$

e o conjunto de dados reduzido é dado por $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$.

Caso o leitor deseje maiores detalhes sobre a prova do Teorema 3.2 consulte a referência (WANG, 2012).

Pelo Teorema 3.2, o conjunto de dados reduzido pode ser obtido diretamente da SVD da matriz \mathbf{X} usando a fórmula 3.8. Comumente o conjunto de dados reduzido, \mathcal{Y} , é obtido pela decomposição espectral da matriz de covariância de \mathbf{X} , definida por

$$\mathbf{F} = \frac{1}{n} \mathbf{X} \mathbf{X}' \quad (3.9)$$

Seja a decomposição espectral de \mathbf{F} dada por

$$\mathbf{F} = \mathbf{V} \Lambda \mathbf{V}'$$

Então a matriz \mathbf{V} é a mesma que foi obtida em 3.6. Consequentemente, podemos extrair a submatriz \mathbf{V}_d de \mathbf{V} . Como as colunas de \mathbf{V}_d são uma base ortonormal para a melhor aproximação do subespaço \mathbf{S}_d , a projeção ortonormal de \mathbf{x}_i pertencente \mathcal{X} sobre \mathbf{S}_d é $\mathbf{V}_d' \mathbf{x}_i$, que produz

$$\mathbf{Y} = (\mathbf{V}_d)' \mathbf{X} \quad (3.10)$$

Em um grande número de aplicações, o número de pontos n é muito maior que a dimensão D : $D \ll n$. Assim sendo, a matriz de covariância \mathbf{F} tem uma dimensão relativamente pequena, por exemplo, $D \times D$, de forma que a decomposição espectral de \mathbf{F} torna-se muito rápida.

3.1.2 Descrição Estatística do PCA

Seja $\mathcal{X} \subset \mathbb{R}^D$ o conjunto de dados dados e \mathbf{X} a sua representação matricial. Agora, considerando \mathcal{X} como o vetor aleatório

$$\mathbf{X} = [X_1, \dots, X_D]' \in \mathbb{R}^D$$

para o qual cada vetor coluna $\mathbf{x}_i \in \mathcal{X}$ é uma observação de \mathbf{X} . Então \mathcal{X} é o espaço amostral do vetor aleatório \mathbf{X} e a i -ésima linha de \mathbf{X} é o espaço amostral da i -ésima variável aleatória X_i (também chamado i -ésimo componente aleatório) de \mathbf{X} . Por conveniência, assumimos que o vetor aleatório \mathbf{X} é normalizado para ter média igual a zero, o que significa que todos os seus componentes possuem média zero: $E(X_i) = 0$, $1 \leq i \leq D$.

Lembrando que a variância de uma variável aleatória Z com n amostras $\{z_i\}_{i=1}^n$ é calculada por

$$\text{Var}(Z) = E((Z - \mu)^2) = \frac{1}{n} \sum_{i=1}^n (z_i - \mu)^2,$$

onde μ é a média de Z . Particularmente, quando $E(Z) = 0$,

$$\text{Var}(Z) = \frac{1}{n} \sum_{i=1}^n (z_i)^2 = \frac{1}{n} \|z\|^2$$

Em estatística, a primeira direção principal do vetor aleatório \mathbf{X} (com média zero) é definida como a direção de variância máxima. Mais precisamente, temos o seguinte.

Definição 3.1. *Seja \mathbf{X} um vetor aleatório D -dimensional. Sua primeira direção principal é um vetor unitário \mathbf{v}_1 definido por*

$$\mathbf{v}_1 = \arg \max_{\|\mathbf{w}\|=1} \text{Var}(\mathbf{w}'\mathbf{X}) \quad (3.11)$$

Correspondentemente, a primeira componente principal é a variável aleatória definida por

$$P_1 = \mathbf{v}'_1 \mathbf{X}$$

Similarmente, assumimos que as $k - 1$ principais componentes principais de \mathbf{X} são Y_1, \dots, Y_{k-1} nas direções principais $\mathbf{v}_1, \dots, \mathbf{v}_{k-1}$, respectivamente. Então a k -ésima direção principal é o vetor unitário definido por

$$\mathbf{v}_k = \arg \max_{\|\mathbf{w}\|=1} \text{Var} \left(\mathbf{w}' \left(\mathbf{X} - \sum_{i=1}^{k-1} \mathbf{Y}_i \mathbf{v}_i \right) \right)$$

e a k -ésima componente principal é a variável aleatória $P_k = (\mathbf{v}'_k \mathbf{X})$.

Se tomarmos um vetor aleatório k -dimensional

$$\mathbf{P} = [P_1, \dots, P_k]',$$

que consiste das k principais componentes principais de \mathbf{X} , então o espaço amostral de \mathbf{P} é um subconjunto de \mathbb{R}^k . Conseqüentemente, do ponto de vista estatístico, o método *Principal Components Analysis* encontra as d principais componentes principais de um dado vetor aleatório D -dimensional. Assumindo que o posto de \mathbf{X} é r , tome a SVD da matriz de dados \mathbf{X} dada por 3.6. Então a matriz \mathbf{U} em 3.6 nos dá o vetor aleatório r -dimensional $\mathbf{U} = [U_1, \dots, U_r]'$ de forma que cada coluna é uma observação de \mathbf{U} . Então, podemos reescrever 3.6 na forma de um vetor aleatório:

$$\mathbf{X} = \sum_{i=1}^r (\sigma_i \mathbf{U}_i) \mathbf{v}_i$$

Lembrando que os valores singulares $\{\sigma_i\}_{i=1}^r$ são arranjados em ordem decrescente. Pelo Teorema 3.2, a primeira componente principal de \mathbf{X} é

$$P_1 = \sigma_1 U_1 \tag{3.12}$$

correspondendo à primeira direção principal \mathbf{v}_1 , que é a solução de 3.11. Por 3.6, P_1 pode ser escrito como

$$P_1 = (\mathbf{v}_1)' \mathbf{X}.$$

Por indução matemática, pode-se provar que as principais d direções principais de \mathbf{X} são $\mathbf{v}_1, \dots, \mathbf{v}_d$, respectivamente, e as d principais componentes principais são

$$P_i = (\mathbf{v}_i)' \mathbf{X}, \quad 1 \leq i \leq d.$$

Representando \mathbf{P} no espaço amostral, temos

$$\mathbf{Y} = (\mathbf{V}_d)' \mathbf{X},$$

que é o mesmo conjunto de dados reduzido \mathcal{Y} como obtido em 3.10.

3.1.3 Algoritmo *Principal Components Analysis*

Passo 1. Centralização dos dados.

Passo 2. Realizar a decomposição de valores singulares da matriz \mathbf{X} . A SVD pode ser substituída pela decomposição espectral da matriz de covariância, particularmente, quando

$D \ll n$. Quando $D \ll n$, a decomposição espectral da matriz de covariância \mathbf{F} em 3.9 é aplicada e o conjunto dimensionalmente reduzido \mathbf{Y} é computado usando a fórmula 3.10. Quando $D \geq n$, a decomposição em valores singulares da matriz centralizada \mathbf{X} é aplicada para gerar \mathbf{V}_d , Σ_d e \mathbf{U}_d , e então \mathbf{Y} é obtido usando a fórmula 3.8.

3.2 Classical Multidimensional Scaling

O *Classical Multidimensional Scaling* é um termo amplo usado para designar uma família de métodos. O *Multidimensional Scaling* ou métrico, ou clássico, foi a primeira abordagem algébrica para esta classe de métodos. Ele foi proposto independentemente por vários autores: [Torgerson \(1958\)](#), [C. \(1966\)](#) e [T. e H. \(1965\)](#). O objetivo principal do método é auxiliar na análise de dados multivariados possibilitando sua visualização em duas ou três dimensões. O *Classical Multidimensional Scaling*, comumente, toma como entrada uma matriz de dissimilaridades calculada entre objetos de um conjunto de dados, utilizando para tal a métrica Euclidiana. As dissimilaridades entre os objetos são então armazenadas por uma matriz de distâncias. A tarefa do *Classical Multidimensional Scaling* é encontrar uma configuração $\mathcal{Y} \subset \mathbb{R}^d$ de forma que a matriz de distâncias Euclidianas se aproxime o máximo possível das dissimilaridades dadas.

Do ponto de vista geométrico, em um espaço Euclidiano, a distância descreve a dissimilaridade de um par de pontos enquanto o produto interno descreve a similaridade, ou seja, eles têm um relacionamento muito próximo. Nesta seção, descreve-se brevemente a relação entre a métrica Euclidiana e a matriz de Gram.

Representando o conjunto de dados $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{R}^D$ em forma de uma matriz $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_n] \in \mathfrak{M}_{D,n}$, onde cada coluna de \mathbf{X} é um ponto de \mathcal{X} . Então a totalidade das colunas da matriz de dados \mathbf{X} constitui o conjunto de dados \mathcal{X} .

3.2.1 Matrizes de Distâncias Euclidianas

Lembrando que a distância Euclidiana entre dois vetores $\mathbf{a} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_D]'$ e $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_D]'$ em um espaço \mathbb{R}^D é definida por

$$d_2(\mathbf{a}, \mathbf{x}) = \|\mathbf{x} - \mathbf{a}\| = \sqrt{\sum_{l=1}^D (\mathbf{x}_l - \mathbf{a}_l)^2}$$

A matriz de distâncias Euclidianas sobre o conjunto de dados $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{R}^D$ é

$$\mathbf{D} \stackrel{def}{=} [\mathbf{D}_{ij}] = [d_2(\mathbf{x}_i, \mathbf{x}_j)]_{i,j=1}^n$$

que define uma métrica sobre \mathcal{X} chamada de métrica Euclidiana. Também é necessário computar a matriz dos quadrados das distâncias Euclidianas definida por

$$\mathbf{S} = [d_2^2(\mathbf{x}_i - \mathbf{x}_j)]_{i,j=1}^n$$

3.2.2 Matriz de Gram sobre o Conjunto de Dados

Lembrando que em um espaço Euclidiano \mathbb{R}^D o produto interno entre dois vetores \mathbf{x} e \mathbf{y} é definido por

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^D x_i y_i,$$

e a matriz de Gram sobre o conjunto de dados é definida por

$$\mathbf{G} = [G_{ij}] = [\langle \mathbf{x}_i, \mathbf{x}_j \rangle]_{i,j=1}^n.$$

A matriz G é semidefinida positiva de dimensão $n \times n$ e posto m , então G pode ser decomposta na forma 3.13

$$\mathbf{G} = \mathbf{X}'\mathbf{X}, \quad (3.13)$$

onde $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_n]$ é uma matriz $m \times n$. Por 3.13, G é a matriz de Gram do conjunto de dados $\mathcal{X} \subset \mathbb{R}^m$.

3.2.3 Relação entre Distância Euclidiana e Matriz de Gram

Para revelar a relação entre a matriz de Gram G e a matriz de distância Euclidiana D de um conjunto de dados \mathcal{X} , tomemos $\mathbf{x}, \mathbf{y} \in \mathbb{R}^D$. Pela Lei dos Cossenos:

$$d_2(\mathbf{x}, \mathbf{y}) = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle - 2\langle \mathbf{x}, \mathbf{y} \rangle}$$

o que nos dá

$$D_{ij} = d_2(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\mathbf{G}_{ii} + \mathbf{G}_{jj} - 2\mathbf{G}_{ij}} \quad (3.14)$$

As entradas da matriz de Gram G são os vetores de produtos internos.

Definição 3.2. Para um dado conjunto de dados \mathcal{X} , temos que a matriz de Gram é dada por:

$$\mathbf{G}^c = [\langle \hat{\mathbf{x}}_i, \hat{\mathbf{x}}_j \rangle]_{i,j=1}^n = \hat{\mathbf{X}}'\hat{\mathbf{X}}.$$

Definição 3.3. Escreva $\mathbf{1} = [1, 1, \dots, 1]' \in \mathbb{R}^n$, $\mathbf{E} = \mathbf{1}\mathbf{1}'$, e seja \mathbf{I} a matriz identidade $n \times n$. Então a matriz $n \times n$ $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{E}$ é chamada matriz de centralização.

Lema 3.3. Seja \mathbf{X} a matriz de dados e \mathbf{G} sua matriz de Gram. Então a matriz de dados centrada de \mathbf{X} é $\mathbf{X}\mathbf{H}$, e a matriz de Gram centrada de \mathbf{X} é $\mathbf{G}^c = \mathbf{H}\mathbf{G}\mathbf{H}$.

Teorema 3.4. Defina \mathbf{S} como a matriz \mathbf{D} de distâncias euclidianas quadradas e a matriz de Gram centrada de um conjunto de dados \mathcal{X} , então temos a seguinte relação:

$$\mathbf{G}^c = -\frac{1}{2}\mathbf{S}^c \quad (3.15)$$

3.2.4 Descrição do Método Classical Multidimensional Scaling

Seja $\mathbf{D} = [d_{ij}]_{i,j=1}^n$ a matriz de distâncias de um conjunto de n exemplos. O método de *Classical Multidimensional Scaling* é um procedimento para encontrar uma configuração $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\} \subset \mathbb{R}^d$, de forma que uma dada matriz de distâncias associada com \mathcal{Y} se aproxime o máximo possível da matriz \mathbf{D} de dissimilaridades.

$$d_Y(\mathbf{y}_i, \mathbf{y}_j) = d_{ij}, \quad \forall i, j. \quad (3.16)$$

O *Classical Multidimensional Scaling* adota a métrica Euclidiana para computar as dissimilaridades que servirão de entrada para o método.

Lema 3.5. Assumindo que a matriz $n \times n$ $\mathbf{D} = [d_{ij}]$ é uma métrica euclidiana e $\mathbf{S} = [(d_{ij})^2]$ é a correspondente matriz de distância quadrada. Tome $\mathbf{G}^c = -\frac{1}{2}\mathbf{S}^c$. Se o posto de \mathbf{G}^c é r , então existe um conjunto de vetores r -dimensional centrados dado que

$$d_2(\mathbf{x}_i, \mathbf{x}_j) = d_{ij}, \quad 1 \leq i, j \leq n. \quad (3.17)$$

Chamamos r no lema 3.5 de dimensão de configuração intrínseca de \mathbf{D} e chamamos \mathcal{X} a configuração exata de \mathbf{D} . Se desejarmos usar a configuração para visualização, e r é muito grande para tal propósito, então procuramos por uma configuração de menor dimensão, por exemplo, uma configuração d -dimensional \mathcal{Y} com $d \ll r$.

A função de perda para o método *Classical Multidimensional Scaling* é dada por

$$\eta(\mathcal{Y}) = \sum_{i,j=1}^n (d_{ij}^2 - d_2^2(\mathbf{y}_i, \mathbf{y}_j)), \quad \text{s.a. } \mathcal{Y} = T(\mathcal{X}),$$

onde T é uma projeção ortogonal de \mathbb{R}^r para um d -subespaço $S^d \subset \mathbb{R}^r$ e \mathcal{X} é a configuração exata de \mathbf{D} . Então o conjunto de pontos \mathcal{Y} é a solução para o seguinte problema de minimização

$$\mathbf{Y} = \arg \min_{Y \in \mathcal{M}_{d,n}} \eta(\mathcal{Y}), \quad \text{s.t. } \mathcal{Y} = T(\mathcal{X}). \quad (3.18)$$

Teorema 3.6. Seja $\mathcal{X} \subset \mathbb{R}^r$ a configuração de \mathbf{D} no lema 3.5 onde \mathcal{H} está centrada e a SVD de \mathbf{X} dada por

$$\mathbf{X} = \mathbf{V}\Sigma\mathbf{U}',$$

onde $\Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r) \in \mathfrak{D}_r$, $\mathbf{U} \in \mathfrak{D}_{n,r}$, e $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_r] \in \mathfrak{D}_r$. Para um dado $d \leq r$, tome $\mathbf{V}_d = [\mathbf{v}_1, \dots, \mathbf{v}_d]$ e $\mathbf{Y} = \mathbf{V}'_d \mathbf{X}$. Então \mathbf{Y} é a solução do problema de minimização em 3.18 com

$$\eta(\mathcal{Y}) = \sum_{i=d+1}^r \sigma_i^2. \quad (3.19)$$

Seja S_d um subespaço de dimensão d de \mathbb{R}^r e $\mathbf{B} \in \mathfrak{D}_{r,d}$ a matriz cujas colunas formam uma base ortonormal sobre S_d . Então $\mathbf{B}\mathbf{B}'$ representa a correspondente projeção ortogonal de \mathbb{R}^r para S_d .

Sendo assim o problema de minimização em 3.18 é encontrar uma matriz $\mathbf{B}^* \in \mathfrak{D}_{r,d}$ dado que o conjunto de vetores $\mathbf{B}^* \mathcal{X}$ minimize $\eta(\mathcal{Y})$:

$$\mathbf{B}^* = \arg \min_{\mathbf{B} \in \mathfrak{D}_{r,d}} \sum_{i,j=1}^n \|(\mathbf{I} - \mathbf{B}\mathbf{B}')(\mathbf{x}_i - \mathbf{x}_j)\|^2, \quad (3.20)$$

temos que a solução para \mathbf{B}^* é dada por:

$$\mathbf{B}^* = \mathbf{U}_d$$

o que nos fornece que $\mathcal{Y} = \mathbf{U}'_d \mathbf{X}$ é a solução de 3.18.

3.2.5 Algoritmo para *Classical Multidimensional Scaling*

Passo 1. Criação da matriz de Gram Centralizada. Para a criação da matriz de Gram centralizada, utilizar a seguinte fórmula: $\mathbf{G}^c = -\frac{1}{2}(\mathbf{D}^2 - \mathbf{D}\mathbf{1}\mathbf{1}' - \mathbf{1}\mathbf{1}'\mathbf{D} + \mathbf{1}\mathbf{1}'\mathbf{D}\mathbf{1}\mathbf{1}')$ onde \mathbf{D}^2 é a matriz dos quadrados de \mathbf{D} calculados para cada uma das entradas da matriz \mathbf{D} , também conhecido como quadrado de Hadamard.

Passo 2. Faça a auto decomposição de \mathbf{G} . Assumindo que o posto de \mathbf{G} é r . Considere a decomposição espectral de \mathbf{G} como $\mathbf{G} = \mathbf{U}\Lambda\mathbf{U}'$ onde $\mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_r]$, e $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r)$, com $\lambda_1 \geq \lambda_2 \dots \geq \lambda_r$.

Passo 2. Faça a decomposição espectral de \mathbf{G} . Passo 3. Encontre a configuração. Seja $\mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_d]$ e $\Sigma_d = \text{diag}(\sqrt{\lambda_1} \dots \sqrt{\lambda_d})$. Então a configuração desejada é $\mathbf{Y} = \Sigma'_d \mathbf{U}$.

Lembrete. Seja $\mathbf{A} = [a_{ij}]$ e $\mathbf{B} = [b_{ij}]$ duas matrizes de mesma dimensão. O produto Hadamard de \mathbf{A} e \mathbf{B} é a matriz $\mathbf{A} \odot \mathbf{B} = [a_{ij}b_{ij}]$. Similarmente, o quadrado de Hadamard de \mathbf{A} é a matriz $\mathbf{A}^2 = \mathbf{A} \odot \mathbf{A}$. Estas operações são chamadas simplesmente de produto e quadrado ponto-a-ponto.

3.3 Métodos de Redução não Lineares

3.3.1 Similaridade e Dissimilaridade

Os métodos de redução não lineares são baseados sobre a hipótese de que o conjunto de dados, de alta dimensionalidade, observado $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{R}^D$ está sobre uma d -variedade de baixa dimensão $M \subset \mathbb{R}^D, d \ll D$. Normalmente as aplicações trabalham sobre conjuntos de dados finitos, considerando que os pontos pertencentes a esses conjuntos são amostrados de uma variedade M . Entretanto, usualmente os dados não são amostrados diretamente de uma variedade existente. Por exemplo, quando representamos um documento de texto em um vetor, o vetor é construído pela contagem de palavras e não amostrados de uma variedade conhecido. Similarmente, em reconhecimento facial, um vetor em um conjunto de dados pode ser ordenado de acordo com os valores de níveis de cinza de seus pixels. Estes conjuntos de dados citados não são diretamente amostrados de variedades existentes, mas são representações digitalizadas de objetos. Assim sendo, as similaridades e dissimilaridades entre objetos de um conjunto de dados são matematicamente descritas por um sistema de vizinhança.

3.3.2 Definição de Vizinhança

Os métodos de redução não lineares, abordados neste trabalho, são baseados na construção de um grafo ponderado a partir de um conjunto de dados fornecido. Para a construção do grafo duas das técnicas mais conhecidas e usadas, para tal tarefa, são K -vizinhos e a corte- ϵ . Seja $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{R}^D$ o conjunto de dados e um parâmetro de tolerância $\epsilon > 0$. A ϵ -vizinhança de um ponto $\mathbf{x}_i \in \mathcal{X}$ é um subconjunto de \mathcal{X} definido por

$$O_i = O_i(\epsilon) = \mathcal{X} \cap B_\epsilon(i) \setminus \{\mathbf{x}_i\},$$

onde $B_\epsilon(i) \subset \mathbb{R}^D$ é uma esfera aberta com centro em \mathbf{x}_i de raio ϵ . Note que, se $\mathbf{x}_j \in O_i$, então $\mathbf{x}_i \in O_j$. Normalmente a corte- ϵ não é utilizada devido à dificuldades práticas para a definição do parâmetro ϵ sobre o conjunto de dados. O valor de ϵ é dependente da dimensão e escala dos dados. Por esta razão, comumente, adota-se a k -vizinhos que é independente da dimensão e escala dos dados e, na maioria dos casos, é adequada para descrever a similaridade entre os objetos pertencentes a um conjunto de dados. Para a construção da k -vizinhos sobre os dados fazemos com que cada um dos pontos tenha K vizinhos. Entretanto, em algumas situações, a técnica k -vizinhos pode não descrever verdadeiramente a estrutura geométrica dos dados. Por exemplo, considere as situações em que determinados pontos no conjunto de dados devam ser isolados, usando-se a k -vizinhos esses pontos terão forçadamente K vizinhos distorcendo a estrutura geométrica dos dados. Por esta razão, em algumas aplicações, recomenda-se o uso de um sistema de vizinhança mais adequado, usando-se técnicas mais eficientes como as que são encontradas em (YANG, 2005). A maioria das aplicações utiliza a distância Euclidiana para a

construção da vizinhança, mas outras distâncias também podem ser adotadas dependendo do propósito da aplicação.

3.3.3 Condições de Consistência

Os métodos de redução não lineares constroem, a partir do conjunto de dados, um grafo de vizinhança, representado como uma matriz de adjacência, para descrever a estrutura geométrica dos dados. Conforme mencionado na seção anterior, em modelos não lineares construímos o grafo de vizinhança utilizando a técnica k -vizinhos ou corte- ϵ . Seja \mathbf{A} a matriz de adjacência do grafo G e o conjunto $N(i) = \{j : \mathbf{A}(i, j) = 1\}$. Então a vizinhança de \mathbf{x}_i sobre o grafo é o conjunto

$$O(i) = \{\mathbf{x}_j, j \in N(i)\}$$

Um grafo, G , bem definido é consistente com a geometria da variedade. Seja U_i a vizinhança de \mathbf{x}_i sobre a variedade, então a vizinhança sobre o grafo O_i deve ser um subconjunto da vizinhança sobre a variedade U_i , de maneira que O_i seja um aberto de U_i dado que cada ponto sobre U_i possa ser aproximadamente representado como uma combinação linear de O_i . Verificar se o grafo é consistente com M é uma tarefa que não pode ser realizada, uma vez que a variedade é desconhecida. Então, para se averiguar a consistência do grafo gerado utilizam-se condições de consistência dadas por:

Consistência de conexão. O grafo G é conectado se e somente se M é conectada.

Consistência de dimensão local. $\dim(\text{span}(O(i))) \geq d$, onde d é a dimensão alvo. A primeira condição certifica-se de que o grafo preserva a conexão da variedade, e a segunda condição garante que o grafo não perca características de dados locais. Por exemplo, se a segunda condição não for satisfeita, então existe uma vizinhança O_i que gera somente um subespaço s -dimensional ($s < d$), então as características locais d -dimensionais (em torno de \mathbf{x}_i) serão perdidas durante o processo de redução. As duas condições acima citadas são chamadas condições de consistência e são de extrema importância para que a solução fornecida pelos métodos de redução sejam computadas eficientemente.

3.4 Isometric Mapping

A proposição de métodos de redução não lineares foi incentivada pela necessidade de criação de técnicas eficazes no tratamento de dados imersos em variedades não lineares. O objetivo do método *Isometric Mapping*, proposto por [Tenenbaum, Silva e Langford \(2000\)](#), é realizar a redução de dimensionalidade de forma a preservar a métrica geodésica sobre o conjunto de dados \mathcal{X} .

O *Isometric Mapping*, assim como os demais métodos baseados em grafos, constrói um grafo de vizinhança G , a partir do cálculo das similaridades tomadas sobre todos os pares de pontos do conjunto de dados \mathcal{X} . Para a computação das medidas de similaridades, o método faz uso de uma estimativa para a métrica geodésica. Como o conjunto de dados é constituído por um número finito de pontos, a medida da verdadeira distância geodésica não pode ser computada diretamente. Dessa forma, o método toma as distâncias calculadas com o uso de um algoritmo de caminhos mínimos sobre grafos, por exemplo, o algoritmo de Dijkstra ([\(DIJKSTRA, 1959\)](#)) ou Floyd Warshall como uma aproximação para as distâncias geodésicas.

A representação dos pontos \mathbf{y}_i , em um espaço dimensionalmente menor, dos pontos \mathbf{x}_i é obtida pela aplicação do *Multidimensional Scaling* sobre a matriz de distâncias geodésicas. Um dos pontos fracos do algoritmo *Isometric Mapping* diz respeito à sua instabilidade topológica. O método pode criar conexões errôneas na vizinhança do grafo G , além do que o método somente tem um desempenho satisfatório se a variedade for convexa, em caso contrário, a representação em baixa dimensão retornada pelo método será depreciada. Apesar desse ponto fraco, o método foi aplicado com sucesso a diversas aplicações tais como sequenciamento de vídeos ([PLESS, 2003](#)), posicionamento de sensores em redes sem fio ([\(WANG et al., 2009\)](#)) e análise de áudio espacial ([ARÉVALO, 2014](#)).

Matematicamente, o *Isometric Mapping* baseia-se na seguinte hipótese. Assumindo que o conjunto de dados observado $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{R}^D$ esteja sobre uma variedade, d -dimensional, Riemanniana $M \subset \mathbb{R}^d$, onde a dimensão $d \ll D$. Denota-se a distância geodésica sobre M por d_M . Então existe um mapa coordenado $f : M \rightarrow \mathbb{R}^d$, $f(\mathbf{x}_i) = \mathbf{y}_i$, o qual preserva a distância geodésica:

$$d_2(f(\mathbf{x}), f(\mathbf{z})) = d_M(\mathbf{x}, \mathbf{z}), \forall \mathbf{x}, \mathbf{z} \in M$$

Particularmente, temos que

$$d_2(\mathbf{y}_i, \mathbf{y}_j) = d_M(\mathbf{x}_i, \mathbf{x}_j), \forall 1 \leq i, j \leq n. \quad (3.21)$$

Como a similaridade sobre o conjunto de dados $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ é a mesma do conjunto de dados \mathcal{X} , o conjunto \mathcal{Y} (*DR*) é a redução de dimensionalidade de \mathcal{X} . Entretanto, porque a variedade M é desconhecida, não podemos explicitamente construir o mapeamento

isométrico f para DR (conjunto reduzido). Ao invés disso, construímos uma aproximação para a métrica geodésica d_M sobre os dados \mathcal{X} . Por 3.21, podemos verificar que a métrica geodésica d_M induz a métrica Euclidiana sobre \mathcal{Y} .

3.4.1 Métrica Geodésica sobre Conjuntos Discretos

A principal etapa do método *Isometric Mapping* é a construção da métrica geodésica sobre os dados observados. Como mencionado, na prática, a medida das distâncias geodésicas entre pontos sobre uma variedade não pode ser computada explicitamente sem a sua formulação. Consequentemente, o *Isometric Mapping* usa a *distância grafo* para substituir a distância geodésica. Aqui, o termo *distância grafo* será utilizado para fazer menção às distâncias calculadas por algoritmos de caminhos mínimos sobre grafos, como o de Dijkstra.

Assumindo que o grafo de vizinhança, foi adequadamente construído, sobre o conjunto de dados \mathcal{X} temos o grafo $G = [\mathcal{X}, E]$, onde $(\mathbf{x}_i, \mathbf{x}_j) \in E$ se e somente se \mathbf{x}_i e \mathbf{x}_j são adjacentes. Para a construção da *métrica grafo* d_G sobre G , definimos a *distância grafo* entre $x, y \in \mathcal{X}$ como segue:

1. se $(\mathbf{x}, \mathbf{y}) \in E$, então $d_G(\mathbf{x}, \mathbf{y}) = d_2(\mathbf{x}, \mathbf{y})$.
2. se $(\mathbf{x}, \mathbf{y}) \notin E$ para um caminho

$$\gamma = (x_0, x_1, \dots, x_{s+1}),$$

que conecta \mathbf{x} e \mathbf{y} com $\mathbf{x} = \mathbf{x}_0$ e $\mathbf{x}_{s+1} = \mathbf{y}$, definimos a distância sobre o caminho por

$$d_\gamma = d_2(\mathbf{x}_0, \mathbf{x}_1) + \dots + d_2(\mathbf{x}_s, \mathbf{x}_{s+1}).$$

Seja Γ o conjunto de todos os caminhos que conectam \mathbf{x} e \mathbf{y} . Então a *distância grafo* entre \mathbf{x} e \mathbf{y} é definida por

$$d_G(\mathbf{x}, \mathbf{y}) = \min_{\gamma \in \Gamma} d_\gamma(\mathbf{x}, \mathbf{y})$$

Quando os pontos são suficientemente densos sobre o domínio de M , a *distância grafo* d_G aproxima de forma satisfatória a distância geodésica d_M . Conforme dito anteriormente, a *distância grafo* é computada com o uso de algoritmos de caminhos mínimos sobre grafos como o de Dijkstra que é eficiente para lidar com grafos com um número grande de vértices.

3.4.2 Kernel para Isometric Mapping

Assumindo que o conjunto de dados \mathcal{X} esteja sobre uma variedade $M \subset \mathbb{R}^D$ e f é um mapa isométrico de M para \mathbb{R}^d dado que $f(x_i) = y_i$. O *kernel Isometric Mapping*, para um dado grafo $G = [\mathcal{X}, E]$, é construído por meio das distâncias computadas entre todos os pares de vértices usando o algoritmo de caminhos mínimos. $D_G = [d_G(i, j)]$ e $d_G(i, j)$ é a *distância grafo* entre \mathbf{x}_i e \mathbf{x}_j . A *distância grafo* d_G , satisfaz as seguintes propriedades:

1. $d_G(i, j) \geq 0$ e $d_G(i, j) = 0 \Leftrightarrow i = j$;
2. $d_G(i, j) = d_G(j, i)$;
3. $d_G(i, j) \leq d_G(i, k) + d_G(k, j)$;

A *métrica grafo* (*distância grafo*) é uma aproximação satisfatória para a métrica geodésica (distância geodésica) quando o grafo é suficientemente denso.

Seja $S = D_G x D_G = [d_G^2(i, j)]$ e S^c a matriz S centrada definida por $S^c = H S H$. Pelo Teorema 3.4 da seção 3.2.3, temos que

$$\mathbf{G}^c = -\frac{1}{2} \mathbf{S}^c.$$

Aplicando o *Principal Components Analysis* à matriz \mathbf{G}^c obtemos o conjunto \mathcal{Y} .

3.4.3 Algoritmo Isometric Mapping

Passo 1. Definição da vizinhança. A vizinhança de um ponto no conjunto de dados $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ pode ser construída utilizando-se das técnicas k -vizinhos ou corte- ϵ . A vizinhança de \mathbf{x}_i é denotada por O_i .

Passo 2. Cálculo da distância grafo. A partir do conjunto de dados $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ e da definição de vizinhança é criado um grafo $G = [\mathcal{X}, E]$. A *distância grafo* é calculada para todos os pontos do conjunto \mathcal{X} utilizando, por exemplo, o algoritmo de Dijkstra.

Passo 3. Construção de kernel de redução. Assumindo que os pontos de \mathcal{X} sejam suficientemente densos. Seja \mathbf{D}_G a distância Euclidiana. Compute a matriz $\mathbf{S}_G = [d_G^2(i, j)]$ e

$$\mathbf{G}^c = -\frac{1}{2} \mathbf{H} \mathbf{S}_G \mathbf{H},$$

onde \mathbf{G}^c é o *kernel* de redução do método.

Passo 4. Decomposição espectral. Realize a decomposição espectral de \mathbf{G}^c dada por

$$\mathbf{G}^c = \mathbf{V} \mathbf{\Lambda} \mathbf{V}',$$

onde $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_d, \dots, \lambda_n)$ e $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n]$. Seja $[\mathbf{y}_1, \dots, \mathbf{y}_n]$ determinados por

$$\mathbf{Y}' = \left[\sqrt{\lambda_1}v_1, \sqrt{\lambda_2}v_2, \dots, \sqrt{\lambda_d}v_d \right]$$

Então temos que \mathbf{Y} é o conjunto dimensionalmente reduzido de \mathbf{X} .

3.5 Locally Linear Embedding

Proposto por [Roweis e Saul \(2000\)](#), o método *Locally Linear Embedding* tem como objetivo a redução de dimensionalidade de modo que a geometria local seja preservada, o que significa que pontos próximos no espaço de alta dimensão permanecerão próximos quando projetados no espaço de baixa dimensão. De forma similar aos demais métodos baseados em grafos, inicialmente é construído um grafo de vizinhança G utilizando a técnica k -vizinhos ou corte- ϵ . Para a redução de dimensionalidade, o método *Locally Linear Embedding* assume que o conjunto de dados $\mathcal{X} = \mathbf{x}_1, \dots, \mathbf{x}_n \subset \mathbb{R}^D$ é amostrado de uma variedade suave e a estrutura de vizinhança é definida sobre o conjunto de dados de modo que a vizinhança esteja sobre ou próxima a uma vizinhança coordenada sobre a variedade. Dessa forma, a geometria local de cada uma destas vizinhanças coordenadas é caracterizada por coeficientes lineares que aproximadamente reconstróem cada ponto a partir de seus vizinhos.

Tome $\mathbf{X}_{D \times n} = [\mathbf{x}_1 \dots \mathbf{x}_n]$ a matriz de dados correspondente ao conjunto $\mathcal{X} = \mathbf{x}_1, \dots, \mathbf{x}_n \subset \mathbb{R}^D$ que é amostrado de uma variedade suave d -dimensional $M \subset \mathbb{R}^D$. Assumindo que a vizinhança de cada ponto de \mathcal{X} esteja definida pelo uso da técnica K -vizinhos ou corte- ϵ , temos um grafo $G = [\mathcal{H}, \mathbf{A}]$ sobre \mathcal{X} , onde \mathbf{A} é a matriz de adjacência. Então o conjunto de índices dos vizinhos de \mathbf{x}_i , denotado por $N(i)$, é dado pela i -ésima linha de \mathbf{A} de modo que

$$N(i) = j, \quad \mathbf{A}_{i,j} \neq 0,$$

e a vizinhança de \mathbf{x}_i , denotada por $O(i)$, é o conjunto

$$O(i) = \{x_j \in \mathcal{X}, \quad j \in N(i)\}$$

O método *Locally Linear Embedding* realiza a reconstrução de cada um dos pontos de \mathcal{X} minimizando a seguinte função quadrática de custo

$$\mathcal{E}(W) = \sum_{i=1}^n \left\| \mathbf{x}_i - \sum_{j \in N(i)} w_{i,j} \mathbf{x}_j \right\|^2 \quad (3.22)$$

Na equação 3.22, $w_{i,j}$ são as entradas de uma matriz \mathbf{W} de dimensões $n \times n$, que conterà os pesos dos vizinhos que serão utilizados para a reconstrução de \mathbf{x}_i . Dessa forma, o peso $w_{i,j}$ sumariza a contribuição de \mathbf{x}_j para a reconstrução de \mathbf{x}_i . Para a computação dos pesos, a função de custo apresentada em 3.22 deve ser minimizada segundo duas restrições. A primeira delas é que cada ponto \mathbf{x}_i é reconstruído somente a partir de seus vizinhos $O(i)$, salientando que $w_{i,j} = 0$ se $\mathbf{x}_j \notin O(i)$. A segunda restrição exige que as linhas da matriz de pesos \mathbf{W} tenha soma igual a um:

$$\sum_{j=1}^n w_{i,j} = 1$$

Esta restrição é importante para a preservação da geometria local dos dados e indica que 1 é um autovalor de \mathbf{W} e possui como autovetor correspondente $\mathbf{1} = [1, 1, \dots, 1]'$.

A soma $\sum_{j \in N(i)} w_{i,j} \mathbf{x}_j$ é chamada de representação localmente linear de \mathbf{x}_i .

Finalmente, para a obtenção do conjunto de dados, dimensionalmente reduzido, $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\} \subset \mathbb{R}^d$, o método *Locally Linear Embedding* minimiza a seguinte função de erro:

$$\mathcal{Y} = \arg \min_{\mathcal{Y} \in \mathbb{R}^d} \sum_{i=1}^n \left\| \mathbf{y}_i - \sum_{j \in N(i)} w_{i,j} \mathbf{y}_j \right\|^2 \quad (3.23)$$

sobre as restrições

$$\mathbf{Y}\mathbf{1} = \mathbf{0} \quad e \quad \mathbf{Y}\mathbf{Y}' = \mathbf{I}_d \quad (3.24)$$

Para a resolução do problema de minimização 3.23 sobre as restrições em 3.24, é construído o seguinte *kernel* de redução

$$\mathbf{K} = (\mathbf{I} - \mathbf{W})' (\mathbf{I} - \mathbf{W}) \quad (3.25)$$

que é semidefinido positivo, possui um autovalor igual a 0 e o correspondente autovetor $\mathbf{1}$.

Seja $\mathbf{u}_1 \dots \mathbf{u}_d$ os autovetores correspondentes aos $2 - (d + 1)$ menores autovalores de \mathbf{K} , respectivamente. Escrevemos $\mathbf{U}_d = [\mathbf{u}_1 \dots \mathbf{u}_d]$.

Então $\mathcal{Y} = \mathbf{U}_d'$ é a matriz dimensionalmente reduzida.

3.5.1 Algoritmo *Locally Linear Embedding*

Passo 1. Definição da vizinhança. A vizinhança de cada ponto do conjunto de dados \mathcal{X} é construída usando a técnica k -vizinhos ou a corte- ϵ . Denotamos o grafo construído $G = [\mathcal{X}, \mathbf{A}]$ e \mathbf{A} a matriz de adjacência. O número de pontos na vizinhança deve ser maior que d .

Construção da matriz de pesos. A matriz de pesos $\mathbf{W} = [w_{i,j}]$ é gerada sobre o grafo $G = [\mathcal{X}, \mathbf{A}]$ do seguinte modo. O valor de $w_{i,j} = 0$ se $\mathbf{A}_{i,j} = 0$, caso contrário devemos computar o valor de $w_{i,j}$ usando a equação 3.23. Após a obtenção dos pesos, eles devem ser normalizados de modo que $\sum_{j \in N(i)} w_{i,j} = 1$.

Construção do *kernel* LLE. Escreva a matriz $\mathbf{L} = \mathbf{I} - \mathbf{W}$, onde \mathbf{I} é a matriz identidade de ordem $n \times n$. O *kernel Locally Linear Embedding* é definido por

$$\mathbf{K} = \mathbf{L}'\mathbf{L} = (\mathbf{I} - \mathbf{W})'(\mathbf{I} - \mathbf{W}).$$

Decomposição espectral do kernel *Locally Linear Embedding*. Faça a decomposição espectral de \mathbf{K} dada por

$$\mathbf{K} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}',$$

onde $\mathbf{\Lambda} = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_{n-1})$ com

$$0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{n-1}$$

Assumindo que o grafo $G = [\mathcal{X}, \mathbf{A}]$ é conectado, então $\lambda_1 > 0$. Seja $\mathbf{u}_1, \dots, \mathbf{u}_d$ as colunas de \mathbf{U} tomadas a partir da $2 - (d + 1)$. A matriz $\mathbf{Y} = [\mathbf{u}_1 \dots \mathbf{u}_d]'$ é a matriz de redução.

3.6 Laplacian Eigenmaps

Proposto por, [Belkin e Niyogi \(2003\)](#), o método *Laplacian Eigenmaps*, similarmente ao *Locally Linear Embedding*, tem como objetivo realizar a redução de dimensionalidade preservando a geometria local. O *kernel* de redução, dado pelo método, é derivado a partir do operador de Laplace-Beltrami sobre a variedade. Dado um conjunto de dados $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ e g uma parametrização da variedade M , tome $h = g^{-1} = [h^1, \dots, h^d]'$ o mapa coordenado sobre M de maneira que para cada $x \in M, y = h(x) \in \mathbb{R}^d$ sejam as d coordenadas para x . De forma análoga aos demais métodos baseados em grafos, inicialmente é construído um grafo ponderado $G = [\mathcal{X}, \mathbf{A}]$ sobre \mathcal{X} . Considere o grafo como sendo não direcionado e O_i como a vizinhança de x_i . Uma vez que a vizinhança seja bem definida, cada ponto $x_j \in O_i$ está próximo de x_i sobre M . Para a obtenção do *kernel* de redução, após a construção do grafo faz-se a escolha dos pesos definidos sobre as arestas utilizando o *kernel* de calor ou a matriz de adjacência \mathbf{A} . Caso o *kernel* de calor seja escolhido, se os vértices i e j estão conectados o peso é definido por $W_{i,j} = \frac{\exp^{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}}{t}$ e caso contrário $W_{i,j}$ é igual a 0. Para setar os pesos utilizando a matriz de adjacência toma-se $W_{i,j} = 1$ se os vértices i e j estão conectados e 0 caso contrário. O uso da matriz de adjacência simplifica a aplicação do método, pois elimina a necessidade de definição do valor para o parâmetro t . Assumindo que o grafo G , inicialmente construído, é conectado computa-se os autovalores e autovetores para o problema de autovetores e autovalores generalizado:

$$Lf = \lambda Df, \quad (3.26)$$

onde D é a matriz diagonal de pesos cujas entradas são definidas por $D_{ii} = \sum_j W_{i,j}$ e \mathbf{L} é a matriz Laplaciana definida por $L = D - W$. O Laplaciano é simétrico, semidefinido positivo e pode ser entendido como um operador sobre funções definidas sobre os vértices de G .

Tomando f_0, \dots, f_{k-1} como sendo as soluções da equação 3.26 ordenadas de acordo com seus autovalores,

$$Lf_0 = \lambda_0 Df_0$$

$$Lf_1 = \lambda_1 Df_1$$

...

$$Lf_{k-1} = \lambda_{k-1} Df_{k-1}$$

$$0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{k-1}$$

Lembrando que o autovetor Lf_0 correspondendo ao autovalor 0 é descartado, usamos os d autovetores para a representação em d -dimensional no espaço Euclidiano.

$$x_i \rightarrow (f_1(i), \dots, f_d(i))$$

O mapa promovido pelo método é realizado minimizando a seguinte função objetivo:

$$\sum_{ij} (\mathbf{y}_i - \mathbf{y}_j)^2 W_{i,j}$$

Assim sendo, o problema de minimização se reduz a encontrar

$$\arg \min \text{tr}(\mathbf{y}_T \mathbf{L} \mathbf{y}), \quad (3.27)$$

sobre as seguintes restrições

$$\mathbf{Y} \mathbf{1} = 0 \quad e \quad \mathbf{Y} \mathbf{Y}' = I.$$

A solução para o problema de minimização 3.27 é dada pela matriz de autovetores correspondentes aos d menores autovalores do problema de autovetores e autovalores generalizado apresentado em 3.26.

Caso o leitor, deseje maiores detalhes sobre o operador de Laplace-Beltrami, sua discretização e o procedimento para minimização de 3.27 consulte as seguintes referências (BELKIN; NIYOGI, 2008) e (BELKIN, 2003).

3.6.1 Algoritmo *Laplacian Eigenmaps*

Passo 1. Definição da vizinhança. A etapa de definição da vizinhança é obtida usando-se a técnica k -vizinhos ou a corte- ϵ .

Passo 2. Criação do grafo ponderado. O algoritmo pode adotar a matriz de adjacência \mathbf{A} como a matriz de pesos, ou pode computá-los utilizando a fórmula seguinte

$$\mathbf{W}_{i,j} = \frac{\exp^{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}}{t}$$

para a construção de \mathbf{W} . Usar a matriz de adjacência \mathbf{A} como matriz de pesos é mais vantajoso pois não há a necessidade de se estimar o valor do parâmetro t .

Passo 3. A construção do kernel. Seja a matriz de pesos \mathbf{W} , denotada por $\mathbf{W} = [w_{i,j}]$, $d_i = \sum_j w_{i,j}$ e $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$. Então $\mathbf{L} = \mathbf{D} - \mathbf{W}$ é o *kernel* do método.

Passo 4. Decomposição espectral do kernel de redução. Resolva o problema de autovalores e autovetores generalizado:

$$\mathbf{L}\mathbf{f} = \lambda\mathbf{D}\mathbf{f}.$$

Seja $\{\lambda_i\}_{i=0}^{n-1}$, com $0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{n-1}$ os autovalores obtidos a partir da equação 3.6.1 e $\mathbf{F} = [\mathbf{f}^1 \dots \mathbf{f}^d]$ a matriz de autovetores harmônicos correspondentes ao conjunto de autovalores $\{\lambda_1, \dots, \lambda_n\}$. Assim sendo, $\mathbf{Y} = \mathbf{F}'$ é a matriz de redução. Se for necessário escalar a matriz de dados reduzida, tome $\mathbf{Y} = (\mathbf{F}\mathbf{D}^1/2)'$.

3.7 Local Tangent Space Alignment

Proposto por, [Zhang e Zha \(2004\)](#), o método de redução *Local Tangent Space Alignment* visa preservar a geometria local aproximando o espaço tangente à variedade que contém os dados usando decomposições em valores singulares. Os espaços tangentes locais são alinhados para fornecer as coordenadas globais dos dados relativas à variedade.

3.7.1 Coordenadas Tangentes e Coordenadas em Variedades

Seja $M \subset \mathbb{R}^D$ uma variedade d -dimensional e $f : U \subset \mathbb{R}^d \rightarrow M$ uma parametrização de M dado que, para cada $\mathbf{x} \in M$,

$$\mathbf{x} = f(\mathbf{y}), \quad \mathbf{y} \in U.$$

Seja $h = f^{-1}$ o mapa coordenado sobre M dado que $\mathbf{y} = h(\mathbf{x})$, $\mathbf{x} \in M$, provê as coordenadas de \mathbf{x} para a variedade. Assumindo que o conjunto de dados $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{R}^D$ esteja sobre a variedade M e $\mathbf{y}_i = h(\mathbf{x}_i)$, $1 \leq i \leq n$. Então o conjunto de coordenadas $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\} \subset \mathbb{R}^d$ é a redução de dimensionalidade de \mathcal{X} .

O método *Local Tangent Space Alignment* encontra a redução de dimensionalidade de \mathcal{X} como segue. É conhecido que para cada $\mathbf{p} \in M$, a geometria local de sua vizinhança $O_{\mathbf{p}}$ pode ser aprendida a partir do plano tangente local $T_{\mathbf{p}M}$. Seja $Q_{\mathbf{p}}$ a projeção ortogonal de $O_{\mathbf{p}}$ sobre o plano tangente $T_{\mathbf{p}M}$. Então temos a seguinte aproximação:

$$Q_{\mathbf{p}}(\mathbf{x} - \mathbf{p}) \approx \mathbf{x} - \mathbf{p}, \quad \mathbf{x} \in O_{\mathbf{p}} \quad (3.28)$$

Seja $\mathcal{U} = \{\mathbf{u}_1, \dots, \mathbf{u}_d\}$ uma base ortonormal sobre $T_{\mathbf{p}M}$. Então, temos

$$Q_{\mathbf{p}}(\mathbf{x} - \mathbf{p}) = \sum_{i=1}^d \tau_i \mathbf{u}_i, \quad \boldsymbol{\tau} = [\tau_1, \dots, \tau_d]' \in \mathbb{R}^d$$

Pela aproximação 3.28, também temos:

$$\mathbf{x} - \mathbf{p} \approx \sum_{i=1}^d \tau_i \mathbf{u}_i, \quad (3.29)$$

o que nos fornece uma representação local de coordenadas tangentes de \mathbf{x} . Por outro lado, cada ponto $\mathbf{x} \in M$ tem a representação de coordenadas da variedade $\mathbf{x} = f(\mathbf{y})$, $\mathbf{y} \in \mathbb{R}^d$. Seja $\mathbf{q} = h(\mathbf{p})$ e $df(\mathbf{q})$ sejam a derivada de f em \mathbf{q} . Aplicando o Teorema de Taylor, temos

$$\mathbf{x} - \mathbf{p} = df(\mathbf{q})(\mathbf{y} - \mathbf{q}) + O(d_2^2(\mathbf{y}, \mathbf{q})) \quad (3.30)$$

$$\mathbf{x} - \mathbf{p} = df(\mathbf{q})(\mathbf{y} - \mathbf{q}) + O(d_2^2(\mathbf{y}, \mathbf{q})) \approx df(\mathbf{q})(\mathbf{y} - \mathbf{q}) \quad (3.31)$$

onde $df(\mathbf{q})$ é uma transformação linear invertível. A equação 3.31 provê uma representação de coordenadas global de \mathbf{x} . As equações 3.29 e 3.31 mostram que a coordenada \mathbf{y} da variedade do vetor $\mathbf{x} \in \mathcal{X}$ pode ser obtida pelo alinhamento das coordenadas locais τ de \mathbf{x} . O método *Local Tangent Space Alignment* realiza a redução de dimensionalidade encontrando as coordenadas locais para o conjunto de dados, e então alinhando-as com as coordenadas da variedade.

3.7.2 Representação de Coordenadas Local

Assumindo que sobre o conjunto de dados $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset M \subset \mathbb{R}^D$ e uma vizinhança bem definida que gere o grafo $G = [\mathcal{X}, \mathbf{A}]$ sobre \mathcal{X} . Denotamos por $O(i)$ a vizinhança de \mathbf{x}_i e por $N(i) = \{i_1, \dots, i_k\}$ o conjunto de índices de $O(i)$. O conjunto $N(i)$ também corresponde às entradas não nulas da i -ésima linha de \mathbf{A} . Seja $k = |N(i)|$. Definimos $\bar{\mathbf{x}} \stackrel{def}{=} \frac{1}{k} \sum_{j \in N(i)} \mathbf{x}_j$, que é o centro geométrico do conjunto de pontos $O(i)$. Entretanto $\bar{\mathbf{x}}$ pode não residir sobre M , mas sim estar muito próximo de M . Por simplicidade, assumimos que $\bar{\mathbf{x}} \in M$. Denote por T_i o espaço tangente de M em $\bar{\mathbf{x}}$ e por H_i o hiperplano tangente $H_i = \bar{\mathbf{x}} + T_i$. Como o sistema de vizinhança sobre \mathcal{X} é bem definido, todos os pontos em $O(i)$ estão muito próximos de suas projeções sobre o hiperplano H_i . Tome $F : O(i) \rightarrow H(i)$ como a projeção ortogonal dado que

$$F(\mathbf{x}_j) = \bar{\mathbf{x}} + \mathbf{p}_j, \quad j \in N(i), \quad \mathbf{p}_j \in T_i \quad (3.32)$$

Dadas as propriedades de aproximação linear de T_i ,

$$\frac{1}{k} \sum_{j \in N(i)} \mathbf{p}_j = \frac{1}{k} \sum_{j \in N(i)} F\mathbf{x}_j - \bar{\mathbf{x}} \approx 0$$

Consequentemente, o conjunto de vetores $\{\mathbf{p}_j\}_{j \in N(i)}$ está centrado e então temos que

$$\mathbf{P}_i = \mathbf{P}_i \mathbf{H},$$

onde \mathbf{H} é a matriz $k \times k$ de centralização e $\mathbf{P}_i = [\mathbf{p}_{i1}, \dots, \mathbf{p}_{ik}]$ onde $i \in N(i)$. Assumindo que $k > d$ e o conjunto de vetores $\{\mathbf{p}_j\}_{j \in N(i)}$ gera T_i . Então podemos construir uma base ortonormal de T_i usando a decomposição de valores singulares da matriz $\mathbf{P}_i = [\mathbf{p}_{i1}, \dots, \mathbf{p}_{ik}]$:

$$\mathbf{P}_i = \mathbf{U} \mathbf{\Sigma} \mathbf{V}', \quad (3.33)$$

onde os vetores colunas de $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_d] \in \mathfrak{D}_{D,d}$ formam uma base ortonormal de T_i . Por 3.33 temos

$$\Theta_i \stackrel{def}{=} \Sigma \mathbf{V}' = [\tau_{s,j}]_{s,j=1}^{d,k}$$

é a matriz de coordenadas local de \mathbf{P}_i e

$$\mathbf{p}_j = \sum_{s=1}^d \tau_{s,j} \mathbf{u}_s, \quad j \in N(i) \quad (3.34)$$

Para computar as coordenadas locais $\tau_{s,j}$, substituímos $\{\mathbf{p}_j\}_{j \in N(i)}$ por sua aproximação $\mathcal{X} \stackrel{def}{=} \{\mathbf{x}_j - \bar{\mathbf{x}}\}_{j \in N(i)}$. Escreva $\mathbf{X}_i = [\mathbf{x}_{i1}, \dots, \mathbf{x}_{ik}]$ e tome a decomposição de valor singular da matriz de dados centrada $\mathbf{X}_i \mathbf{H}$ como

$$\mathbf{X}_i \mathbf{H} = \mathbf{U}_D \Sigma_D \mathbf{V}'_D, \quad (3.35)$$

onde $\mathbf{U}_D = [\mathbf{u}_1 \dots \mathbf{u}_D]$ é uma matriz ortonormal, Σ_D é uma matriz diagonal $D \times D$ consistindo de todos os valores singulares de $\mathbf{X}_i \mathbf{H}$, e $\mathbf{V}_D = [\mathbf{v}_1 \dots \mathbf{v}_D] \in \mathfrak{D}_{k,D}$. Escreva $\mathbf{U}_d = [\mathbf{u}_1 \dots \mathbf{u}_d]$, $\mathbf{V}_d = [\mathbf{v}_1 \dots \mathbf{v}_d]$ e $\Sigma_d = \text{diag}(\sigma_1, \dots, \sigma_d)$. Por 3.32, $\mathbf{U}_d \Sigma_d \mathbf{V}'_d$ é aproximadamente igual a \mathbf{P}_i em 3.33. Por conveniência, deletamos o subscrito d em $\mathbf{U}_d \Sigma_d \mathbf{V}'_d$, identificando com 3.33. A equação 3.35 fornece

$$\mathbf{X}_i \mathbf{H} = \mathbf{U} \Sigma \mathbf{V}' (= \mathbf{U} \Theta_i) \quad (3.36)$$

Agora, discutimos a relação entre a representação de coordenadas local do conjunto \mathbf{H}_i e sua representação com relação à variedade. Aplicando 3.31 ao conjunto de dados \mathbf{H}_i com $\mathbf{p} = \bar{\mathbf{x}}$, temos que:

$$\mathbf{x}_j - \bar{\mathbf{x}} = df(\bar{\mathbf{y}})(\mathbf{y}_j - \bar{\mathbf{y}}), \quad (3.37)$$

onde $\bar{\mathbf{y}} = h(\bar{\mathbf{x}})$. Se a aproximação é suficientemente acurada, temos que $\bar{\mathbf{y}} = \sum_{j \in N(i)} \mathbf{y}_j$. Setando $\mathbf{Y}_i = [\mathbf{y}_{i1} \dots \mathbf{y}_{ik}]$, podemos reescrever 3.37 na forma matricial

$$\mathbf{X}_i \mathbf{H} = df(\bar{\mathbf{y}}) \mathbf{Y}_i \mathbf{H} \quad (3.38)$$

Lembrando que $df(\bar{\mathbf{y}})$ é invertível e $df(\bar{\mathbf{y}})^{-1} = dh(\bar{\mathbf{x}})$. Consequentemente, temos que

$$\mathbf{Y}_i \mathbf{H} = dh(\bar{\mathbf{x}}) \mathbf{X}_i \mathbf{H} = \mathbf{L} \Sigma \mathbf{V}', \quad \mathbf{L} \stackrel{def}{=} dh(\bar{\mathbf{x}}) \mathbf{U}, \quad (3.39)$$

o que nos dá que $\mathbf{L} = \mathbf{Y}_i \mathbf{H} \mathbf{V} \mathbf{\Sigma}^{-1}$ e

$$\mathbf{Y}_i \mathbf{H} (\mathbf{I} - \mathbf{V} \mathbf{V}') = \mathbf{L} \mathbf{\Sigma} \mathbf{V}' - \mathbf{L}' \mathbf{V} \mathbf{V}' = 0 \quad (3.40)$$

3.7.3 Alinhamento Global

Para fazer o alinhamento global para cada aproximação local tangente, primeiramente precisamos definir algumas restrições sobre a matriz \mathbf{Y} de redução de dimensionalidade. É requerido que \mathbf{Y} tenha média zero e $\mathbf{Y} \mathbf{Y}' = \mathbf{I}$. A chave do alinhamento global é converter a relação local 3.40 para uma global. Seja \mathbf{V} a matriz em 3.36. Escrevemos $\mathbf{W}_i = \mathbf{H} (\mathbf{I} - \mathbf{V} \mathbf{V}')$. Por 3.40 $\mathbf{Y}_i \mathbf{W}_i = 0$.

Agora estende-se a matriz $d \times d$, \mathbf{W}_i para uma matriz $n \times n$, \mathbf{W}^i , que é definida por:

$$\mathbf{W}^i(j, k) = \mathbf{W}_i(s, l), \quad \text{se } j = i_s, \quad k = i_l \in N(i) \text{ e } 0 \text{ caso contrário}$$

De maneira similar, a matriz \mathbf{Y}_i é estendida para uma matriz $k \times n$ \mathbf{Y}^i dado que \mathbf{Y}_i é uma submatriz de \mathbf{Y}^i com os índices das colunas de $N(i)$ e as outras colunas de \mathbf{Y}^i iguais a 0. A equação 3.40 imediatamente nos dá que

$$\mathbf{Y}^i \mathbf{W}^i = 0 \quad (3.41)$$

Porque colunas não eliminadas de \mathbf{W}^i devem ter os índices em $N(i)$, temos que

$$\mathbf{Y} \mathbf{W}^i = 0 \quad (3.42)$$

Além disso, por $\mathbf{1}' \mathbf{H} = 0$, também temos que

$$\mathbf{1}' \mathbf{W}^i = 0 \quad (3.43)$$

Então a matriz de dados \mathbf{Y} satisfaz 3.40 para todo $i = 1, \dots, n$. Consequentemente, obtemos o *kernel* para o método *Local Tangent Space Alignment*

$$\mathbf{K} = \sum_{i=1}^n \mathbf{W}^i \quad (3.44)$$

Por 3.42 e 3.44, temos que

$$\mathbf{Y} \mathbf{K} = \sum_{i=1}^n \mathbf{Y} \mathbf{W}^i = 0 \quad (3.45)$$

Similarmente, por 3.43, temos que

$$\frac{1}{n} \mathbf{1}' \mathbf{K} = 0 \quad (3.46)$$

Como \mathbf{K} é simétrico, por 3.45 e 3.46, obtemos $\mathbf{K} \left[\frac{1}{n} \mathbf{1}' \mid \mathbf{Y}' \right] = 0$, o que nos fornece que as colunas da matriz $\left[\frac{1}{n} \mathbf{1}' \mid \mathbf{Y}' \right]$ formam uma base ortonormal do espaço nulo do *kernel* \mathbf{K} .

Finalmente, ressaltamos que o *kernel* \mathbf{K} do método *Local Tangent Space Alignment* é semidefinido positivo.

Considerando possíveis erros de aproximação e desvios dos dados, computamos o conjunto de dados \mathcal{Y} como a solução do seguinte problema de minimização:

$$\mathbf{Y} = \arg \min_{\mathbf{Y} \in \mathfrak{D}_{d,n}} \text{tr}(\mathbf{Y} \mathbf{K} \mathbf{Y}'), \quad \text{s.t. } E(\mathbf{Y}) = 0 \quad (3.47)$$

Então as colunas da matriz \mathbf{Y}' são os autovetores de \mathbf{K} correspondentes ao $2 - (d + 1)$ menores autovalores.

3.7.4 Algoritmo *Local Tangent Space Alignment*

Passo 1. Definição de vizinhança. A vizinhança de cada ponto é construída usando-se uma das técnicas de construção, k -vizinhos ou corte- ϵ . O número de vizinhos deve ser maior que d . O grafo é denotado por $G = [\mathcal{X}, \mathbf{A}]$.

Passo 2. A computação das coordenadas tangentes locais. Seja $\mathbf{X}_i = [\mathbf{x}_{i1} \dots \mathbf{x}_{ik}]$ a matriz cujas colunas são os k vizinhos mais próximos de \mathbf{x}_i . Os dados devem ser centralizados, subtraindo-se a média calculada de cada um dos vetores: $\hat{\mathbf{X}}_i = [\mathbf{x}_{i1} - \bar{\mathbf{x}} \dots \mathbf{x}_{ik} - \bar{\mathbf{x}}]$. As coordenadas locais (em \mathbb{R}^d) de $\hat{\mathbf{X}}_i$ são obtidas pelo *Principal Components Analysis*:

$$\sum_{j=1}^d \sigma_j \mathbf{u}_j (\mathbf{v}_j)'$$

Escreva $\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_d]$ e setamos $\mathbf{G}_i = \left[\frac{1}{k} \mid \mathbf{V}_i \right] \in \mathfrak{D}_{k,d+1}$. Então $\mathbf{W}_i = \mathbf{I} - \mathbf{G}_i \mathbf{G}_i'$.

Passo 3. Construção do *kernel* via alinhamento de coordenadas. O *kernel* \mathbf{K} é a matriz de alinhamento de todas as matrizes locais \mathbf{W}_i . Inicialize \mathbf{K} como a matriz nula. Seja $N(i)$ o conjunto de índices da vizinhança de \mathbf{x}_i e $\mathbf{K}(N(i), N(i))$ a submatriz de \mathbf{K} contendo as linhas e colunas de índices $N(i)$. Então atualize \mathbf{K} setando $\mathbf{K}(N(i), N(i)) = \mathbf{I} - \mathbf{G}_i \mathbf{G}_i'$ para $i = 1, 2, \dots, n$, sucessivamente.

Passo 4. Decomposição espectral do *kernel* de redução. Faça a decomposição espectral do *kernel* \mathbf{K} dada por

$$\mathbf{K} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}',$$

onde $\mathbf{\Lambda} = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_{n-1})$ com

$$0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{n-1}.$$

Assumindo que o grafo $G = [\mathcal{X}, \mathbf{A}]$ é conectado, então temos que $\lambda_1 > 0$. Seja $\mathbf{Y} = [\mathbf{u}_1 \dots \mathbf{u}_d]'$ a matriz de autovetores correspondendo ao $2 - (d + 1)$ menores autovalores de \mathbf{K} . Assim sendo, \mathcal{Y} é a matriz de redução. Este passo é o mesmo que o algoritmo *Locally Linear Embedding*.

3.8 Hessian Locally Linear Embedding

Proposto por, [Donoho e Grimes \(2003\)](#), o *Hessian Locally Linear Embedding* encontra a representação no espaço de baixa dimensionalidade minimizando o funcional Hessiano sobre a variedade. O embasamento conceitual do *Hessian Locally Linear Embedding* é análogo ao do *Laplacian Eigenmaps*, enquanto o *Hessian Locally Linear Embedding* deriva sua solução do espaço nulo do Hessiano o *Laplacian Eigenmaps* deriva do espaço nulo do operador de Laplace-Beltrami. Considerando o modelo utilizado durante o trabalho, temos o conjunto de dados $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{R}^D$ que reside sobre uma variedade M , d -dimensional, $\subset \mathbb{R}^D$. Para cada ponto $\mathbf{x} \in M$, existe uma vizinhança $O_{\mathbf{x}} \subset M$, que é homeomorfismo diferenciável para um conjunto aberto $U \subset \mathbb{R}^d$. Seja $g = [g^1, \dots, g^D]' : U \rightarrow O_{\mathbf{x}}$ como uma parametrização de $O_{\mathbf{x}}$. Então $h = [h^1, \dots, h^d]' = g^{-1} : U \rightarrow O_{\mathbf{x}}$ é um mapa coordenado sobre M . A parametrização de g não é única, e podemos assumir que g é uma isometria de U para $O_{\mathbf{x}}$. Então $h = g^{-1}$ é uma mapa ortonormal coordenado sobre $O_{\mathbf{x}}$. Escrevemos $\mathbf{y} = h(\mathbf{x}) \in U$ e $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ com $\mathbf{y}_i = h(\mathbf{x}_i)$, $1 \leq i \leq n$ é o conjunto dimensionalmente reduzido de \mathcal{X} .

3.8.1 Hessiano sobre Variedades

O método *Hessian Locally Linear Embedding* desempenha a redução de dimensionalidade fazendo com que \mathcal{Y} seja a minimização do funcional Hessiano sobre a variedade M . Seja $f \in C^2(M) : M \rightarrow \mathbb{R}$ uma função suave definida sobre M . Para deixar a ideia mais clara, primeiro definimos o Hessiano de f sobre M com respeito às coordenadas da variedade $h = [h^1, \dots, h^d]$.

Definição 3.4. *Seja $\{e_i\}_{i=1}^d$ as coordenadas canônicas da base de \mathbb{R}^d , $\mathbf{y} \in U \subset \mathbb{R}^d$ e $\mathbf{x} = g(\mathbf{y}) \in M$. Para a função $f \in C^2(M)$, definimos*

$$\frac{\partial}{\partial y^i}(\mathbf{y}) = \lim_{t \rightarrow 0} \frac{f(\mathbf{y} + te_i) - f(\mathbf{y})}{t}$$

Então o Hessiano de f em \mathbf{x} com respeito às coordenadas da variedade $\mathbf{y} = h(\mathbf{x})$ é definida pela matriz

$$\mathbf{H}_{\mathbf{x}}^{iso} = [\mathbf{H}_{\mathbf{x}}^{iso}(f)_{ij}]_{i,j=1}^d, \quad \mathbf{H}_{\mathbf{x}}^{iso}(f)_{ij} \stackrel{def}{=} \frac{\partial^2(f \circ g)(\mathbf{y})}{\partial y_i \partial y_j}$$

e o correspondente funcional Hessiano sobre $C^2(M)$ é definido por

$$\mathcal{H}^{iso}(f) = \int_M \|\mathbf{H}_{\mathbf{x}}^{iso}(f)\|_F^2,$$

onde $\|\cdot\|_F$ é a norma de Frobenius.

Chamamos $\mathbf{H}_{\mathbf{x}}^{iso}(f)$ e $\mathcal{H}^{iso}(f)$ a matriz Hessiana da variedade e o funcional Hessiano da variedade respectivamente como eles são avaliados sobre as coordenadas da variedade. Para cada função coordenada h^i , $h^i \circ g(\mathbf{y}) = y_i$. Consequentemente, para qualquer $\mathbf{x} \in M$, temos o seguinte

$$\mathbf{H}_{\mathbf{x}}^{iso}(h^i) = \mathbf{0}, \quad 1 \leq i \leq d, \quad \mathbf{H}_{\mathbf{x}}^{iso}(e) = \mathbf{0}, \quad (3.48)$$

onde e é a função constante sobre M satisfazendo $e(\mathbf{x}) = 1$. A relação ponto a ponto de 3.48 e 3.49 podem ser estendidas para uma relação global.

$$\mathcal{H}^{iso}(f) = 0, \quad f = e, \quad h^1, \dots, h^d. \quad (3.49)$$

Como descrito no método *Laplacian Eigenmaps*, as equações 3.50 somente são válidas se o conjunto de dados, no modelo ideal, \mathcal{X} reside estritamente sobre M . Em geral, modificamos as equações 3.50 para o seguinte problema de minimização.

$$h = \arg \min_{\langle \mathbf{F}, \mathbf{F} \rangle = \mathbf{I}} \mathcal{H}^{iso}(h), \quad \mathbf{F} = [e, h^1, \dots, h^d], \quad h^i \in C^2(M) \quad (3.50)$$

onde $\langle \mathbf{F}, \mathbf{F} \rangle = \mathbf{I}$ significa que o conjunto de funções $\{e, h^1, \dots, h^d\}$ formam um sistema ortogonal no espaço $C^2(M)$. A solução de 3.51 fornece o conjunto de dados dimensionalmente reduzido \mathcal{Y} por $\mathbf{y}_i = h(\mathbf{x}_j)$.

3.8.2 Hessiano sobre Espaço Tangente

Porque a variedade M subjacente é desconhecida, a parametrização g do funcional Hessiano $\mathcal{H}^{iso}(f)$ não pode ser avaliada diretamente. Para construir o kernel de redução *Hessian Locally Linear Embedding*, necessitamos obter uma representação computável de $\mathcal{H}^{iso}(f)$. Seja $T_{\mathbf{x}}M$ o espaço tangente de M em \mathbf{x} e $L_{\mathbf{x}} = \mathbf{x} + T_{\mathbf{x}}M$ o hiperplano tangente através do vetor \mathbf{x} . Denotamos a projeção ortogonal de $O_{\mathbf{x}}$ para $L_{\mathbf{x}}$ por $P = P_{\mathbf{x}}$. Como $L_{\mathbf{x}}$ está próximo de $O_{\mathbf{x}}$, para cada $\mathbf{p} \in O_{\mathbf{x}}$, $\mathbf{q} = P(\mathbf{p}) \in L_{\mathbf{x}}$ é uma aproximação de \mathbf{p} . Definimos um conjunto aberto $N_{\mathbf{x}} \subset \mathbb{R}^D$ dado que $O_{\mathbf{x}} \cup P(O_{\mathbf{x}}) \subset N_{\mathbf{x}}$. Seja a função $f \in C^2(M)$ suavemente estendida sobre o conjunto aberto $N_{\mathbf{x}} \subset \mathbb{R}^D$.

Agora escolhemos uma base ortonormal de $T_{\mathbf{x}}M$:

$$\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_d\}, \quad \mathbf{b}_i \in \mathbb{R}^D$$

Então cada ponto $\mathbf{t} \in T_{\mathbf{x}}M$ tem uma representação coordenada tangente

$$\mathbf{t} = \sum_{i=1}^d t_i \mathbf{b}_i \quad (3.51)$$

Para a função $f \in C^2(M)$, que foi estendida sobre $N_{\mathbf{x}} \subset \mathbb{R}^D$, podemos computar suas derivadas direcionais na direção de $\mathbf{b}_i \in T_{\mathbf{x}}M$:

$$\frac{\partial f(\mathbf{x})}{\partial t_i} = \lim_{s \rightarrow 0} \frac{f(\mathbf{x} + s\mathbf{b}_i) - f(\mathbf{x})}{s}.$$

Definimos o funcional Hessiano tangente de f em \mathbf{x} como segue.

Definição 3.5. A matriz Hessiana tangente $\mathbf{H}_{\mathbf{x}}^{\text{tan}} = [\mathbf{H}_{\mathbf{x}}^{\text{tan}}(f)_{ij}]$ de $f \in C^2(M)$ em \mathbf{x} (com respeito às coordenadas tangentes \mathbf{t}) é definida por

$$\mathbf{H}_{\mathbf{x}}^{\text{tan}}(f)_{ij} = \frac{\partial^2 t}{\partial t_i \partial t_j}(\mathbf{x}), \quad (3.52)$$

e o correspondente funcional Hessiano em coordenadas tangentes é definido por

$$\mathcal{H}^{\text{tan}}(f) = \int_M \|\mathbf{H}_{\mathbf{x}}^{\text{tan}}\|_F^2, \quad f \in C^2(M). \quad (3.53)$$

Por conveniência, chamaremos apenas de $\mathcal{H}^{\text{tan}}(f)$ o funcional Hessiano tangente.

O Teorema 3.7 nós dá a relação equivalente de $\mathcal{H}^{\text{iso}}(f)$ e $\mathcal{H}^{\text{tan}}(f)$.

Teorema 3.7. O funcional Hessiano da variedade de f é igual ao funcional Hessiano tangente de f :

$$\mathcal{H}^{\text{tan}}(f) = \mathcal{H}^{\text{iso}}(f).$$

Então, o problema de minimização 3.51 pode ser reescrito como:

$$h = \arg \min_{\langle \mathbf{F}, \mathbf{F} \rangle = \mathbf{I}} \mathcal{H}^{\text{tan}}(h), \quad \mathbf{F} = [e, h^1, \dots, h^d]', \quad h^i \in C^2(M) \quad (3.54)$$

Como o hiperplano tangente $L_{\mathbf{x}}$ pode ser "aprendido" da vizinhança de \mathbf{x} , o funcional Hessiano tangente é aproximadamente computável. Pelo Teorema 3.7, nós podemos simplesmente chamar o funcional Hessiano tangente ou o funcional Hessiano da variedade por $\mathcal{H}(f)$. A prova do Teorema 3.7 é bastante técnica e por isso não será apresentada.

3.8.3 Construção do Funcional Hessiano

Nesta seção, é apresentado como construir a forma discreta do funcional Hessiano. A função f , restrita sobre \mathcal{X} , pode ser representada pelo vetor

$$\mathbf{f} = [f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)]'.$$

Assim sendo, o funcional Hessiano \mathcal{H} pode ser representado como a forma quadrática de \mathbf{f} :

$$\mathcal{H}(f) = \mathbf{f}'\mathbf{K}\mathbf{f}, \quad (3.55)$$

onde \mathbf{K} é uma matriz semidefinida positiva $n \times n$, chamada *kernel Hessian Locally Linear Embedding*. O Teorema 3.8 caracteriza o funcional Hessiano.

Teorema 3.8. *O funcional Hessiano $\mathcal{H}(f)$ sobre $C^2(M)$ tem seu espaço nulo de dimensão $d + 1$, que é gerado pela função constante e d funções coordenadas.*

Pelo Teorema 3.8, o conjunto de funções $\{e, h^1, \dots, h^d\}$ é o conjunto das 0-autofunções de \mathbf{K} . Sendo assim, temos $d+1$ autovetores que são dados pelos $d+1$ menores autovalores de \mathbf{K} . Para construir o *kernel* \mathbf{K} , necessitamos de uma estrutura de vizinhança sobre M . Assumindo que um sistema de vizinhança é construído sobre \mathcal{X} . Denotamos por $O_i = \{\mathbf{x}_{i1}, \dots, \mathbf{x}_{ik}\}$ a vizinhança de \mathbf{x}_i e por N_i o conjunto subscrito de O_i . Lembrando que todas as funções quadráticas sobre um espaço vetorial d -dimensional foram um $d_q \stackrel{def}{=} (d+2)(d+1)/2$ dimensional espaço de funções. Para garantir que todas as funções quadráticas discretas sobre O_i não são degeneradas, setamos $k \geq d$. Como usual, assumimos também que o sistema de vizinhança sobre \mathcal{X} é consistente com a estrutura de vizinhança sobre M dado que cada O_i é um subconjunto da vizinhança da variedade $W_i \subset M$ com $O_i = \mathcal{X} \cap W_i$ e $\bigcup_{i=1}^n W_i = M$. Seja $\{\phi_i\}_{i=1}^m$. Seja $\{\phi_i\}_{i=1}^m$ a partição unitária de M com $\phi_i \subset W_i$. Pela definição de integral sobre M , temos

$$\mathcal{H}(f) = \int_M \|\mathbf{H}_{\mathbf{x}}^{tan}(f)\|_F^2 = \sum_{i=1}^m \int_{W_i} \phi_i \|\mathbf{H}_{\mathbf{x}}^{tan}(f)\|_F^2 \quad (3.56)$$

Lembrando que a restrição de f sobre O_i tem a representação local

$$f_i = [f_{\mathbf{x}_{i1}}, \dots, f_{\mathbf{x}_{ik}}].$$

Sendo assim, a forma discreta da integral é a forma quadrática

$$\mathcal{H}(f) | W_i = \int_{W_i} \phi_i \|\mathbf{H}_{\mathbf{x}}^{tan}(f)\|_F^2 = (\mathbf{f}')\mathbf{W}_i\mathbf{f}, \quad (3.57)$$

onde \mathbf{W}_i é uma matriz $k \times k$ semidefinida positiva cujo espaço nulo consiste da função constante e funções coordenadas sobre o espaço tangente $T_{x_i}M$. Para obtermos as coordenadas tangentes para $T_{x_i}M$, tomamos $\bar{\mathbf{x}} = \frac{1}{k} \sum_{j \in N(i)} \mathbf{x}_j$. Definimos a matriz

$$\mathbf{M}^i = [\mathbf{x}_{i1} - \bar{\mathbf{x}}, \dots, \mathbf{x}_{ik} - \bar{\mathbf{x}}], \quad (3.58)$$

e computamos sua SVD:

$$\mathbf{M}^i = \mathbf{U}\mathbf{\Sigma}\mathbf{V}', \quad (3.59)$$

onde $\mathbf{U} \in \mathfrak{D}_{D,D}$, $\mathbf{V} \in \mathfrak{D}_{k,k}$ e $\mathbf{\Sigma}$ é uma matriz diagonal $D \times k$ tendo os valores singulares sobre sua diagonal.

Tome $\mathbf{U}^d = [\mathbf{u}_1, \dots, \mathbf{u}_d]$ como as primeiras d colunas extraídas de \mathbf{U} e $\mathbf{V}_d = [\mathbf{v}_1, \dots, \mathbf{v}_d]$ como as primeiras d colunas extraídas de \mathbf{V} . Como a dimensão de $T_{\mathbf{x}_i}M$ é d , o conjunto $\{\mathbf{u}_1, \dots, \mathbf{u}_d\}$ é uma base ortonormal de $T_{\mathbf{x}_i}M$, e $\mathbf{v}_1, \dots, \mathbf{v}_d$ são as funções coordenadas tangentes discretas sobre O_i . Por 3.60, a matriz de dados local \mathbf{M}^i está centrada de modo que $\langle \mathbf{v}_j, \mathbf{1} \rangle = 0$, $1 \leq j \leq d$. Escreva

$$\mathbf{L}^i = [\mathbf{1}, \mathbf{V}_d^d].$$

As colunas de \mathbf{L}^i forma uma base ortonormal para o espaço de funções lineares sobre O_i . Tome a e b como duas funções sobre $T_{\mathbf{x}_i}M$. Elas são restrições sobre O_i são vetores $\mathbf{a} = [a_1, \dots, a_k]'$ e $\mathbf{b} = [b_1, \dots, b_k]'$, onde $a_j = a(\mathbf{x}_{ij})$ e $b_j = b(\mathbf{x}_{ij})$. Denotamos a forma discreta do produto das funções ab pelo produto Hadamard $\mathbf{a} \odot \mathbf{b}$:

$$\mathbf{a} \odot \mathbf{b} = [a_1 b_1, \dots, a_k b_k]'$$

Então os vetores colunas na matriz

$$\mathbf{Q}_i = [\mathbf{v}_l \odot \mathbf{v}_j]_{1 \leq l \leq j \leq d}$$

forma uma base do espaço de todas as funções quadráticas homogêneas sobre O_i e as colunas da matriz $d_q \times k$

$$\mathbf{B}^i = [\mathbf{L}^i \mathbf{Q}_i]$$

forma uma base do espaço de todas as funções quadráticas sobre O_i . Ortonormalizamos \mathbf{B}^i para

$$\mathbf{B}_{on}^i = [\mathbf{L}^i \mathbf{Q}_{on}^i]$$

de modo que todas as colunas de \mathbf{B}_{on}^i formam uma base ortonormal do espaço das funções quadráticas. Particularmente, as colunas da matriz

$$\mathbf{Q}_{on}^i = [\mathbf{q}_1 \dots \mathbf{q}_r], \quad r = d(d+1)/2,$$

forma uma base ortonormal do espaço de todas as funções quadráticas homogêneas. Definimos

$$\mathbf{H}^i = (\mathbf{Q}_{on}^i)'$$

Então

$$\mathbf{H}^i \mathbf{e} = \mathbf{0},$$

$$\mathbf{H}^i \mathbf{v}_j = \mathbf{0}, \quad j = 1, \dots, d,$$

e

$$\mathbf{H}^i \mathbf{q}_j = \mathbf{e}_j,$$

onde \mathbf{e}_j é o j -ésimo vetor unitário. Afirmamos que o Hessiano local discreto \mathbf{W}_i em 3.60 tem a representação

$$\mathbf{W}_i = (\mathbf{H}^i)' \mathbf{H}^i, \quad (3.60)$$

porque temos

$$\mathbf{1}' \mathbf{W}_i \mathbf{1} = \mathbf{0},$$

$$\mathbf{v}'_j \mathbf{W}_i \mathbf{v}_j = \mathbf{0}, \quad 1 \leq j \leq d,$$

$$\mathbf{q}'_j \mathbf{W}_i \mathbf{q}_l = \delta_{lj}, \quad 1 \leq j \leq l \leq d.$$

3.8.4 Construção do *kernel Hessian Locally Linear Embedding*

Para construir o *kernel Hessian Locally Linear Embedding* em 3.55, aplicamos a fórmula 3.56. Estendemos a fórmula local 3.57 para uma global como segue. Seja \mathbf{W}^i uma matriz $n \times n$ de modo que todas as suas entradas sejam iguais a zeros exceto

$$\mathbf{W}^i(N(i), N(i)) = \mathbf{W}_i,$$

onde $\mathbf{W}^i(N(i), N(i))$ denota a submatriz $k \times k$ de \mathbf{W}^i com ambos os índices das linhas e colunas tomados de $N(i)$. Então a matriz

$$\mathbf{K} = \sum_{i=1}^n \mathbf{W}^i \quad (3.61)$$

é o *kernel Hessian Locally Linear Embedding*.

3.8.5 Algoritmo *Hessian Locally Linear Embedding*

Passo 1. Definição da vizinhança. A construção da vizinhança é feita usando-se uma das técnicas k -vizinhos corte- ϵ . Tome $r = (d+2)(d+1)/2$. O algoritmo utiliza como tamanho de vizinhança um valor $k \geq r$ na computação dos funcionais Hessianos locais.

Passo 2. Criação das coordenadas tangentes locais. Primeiramente as coordenadas do espaço tangente em cada ponto \mathbf{x}_i são estimadas usando o *Principal Components Analysis* sobre a vizinhança. Seja o conjunto de dados local definido por $\mathcal{X}^i = \{\mathbf{p}_{j_1}, \dots, \mathbf{p}_{j_k}\}$. Aplique o algoritmo *Principal Components Analysis* para obter os d principais componentes de \mathcal{X}^i em uma matriz $k \times d$ $\mathbf{V}^i = [\mathbf{v}_1 \dots \mathbf{v}_d]$. Então as colunas de \mathbf{V}^i são as funções coordenadas tangentes sobre \mathcal{X}^i .

Passo 3. Construção do funcional local Hessiano. Seja $\mathbf{1} = [1, \dots, 1]' \in \mathbb{R}^k$ e $\mathbf{Q}^i = [\mathbf{v}_i \mathbf{v}_j]_{1 \leq i \leq j \leq d}$. Definimos

$$\mathbf{V}^a = [\mathbf{1}, \mathbf{V}^i, \mathbf{Q}^i].$$

Aplique o procedimento de Gram-Schmidt sobre \mathbf{V}^a para obter os componentes de \mathbf{V}^a ortonormalizados $[\mathbf{1}, \mathbf{V}^i, \tilde{\mathbf{Q}}^i]$. Assim sendo, o funcional local Hessiano é dado por $\mathbf{W}^i = \tilde{\mathbf{Q}}^i (\tilde{\mathbf{Q}}^i)'$.

Passo 4. Construção do *kernel* de redução. Inicialize o *kernel* \mathbf{K} como uma matriz nula de dimensão $n \times n$. Atualize \mathbf{K} por $\mathbf{K}(N(i), N(i)) = \mathbf{K}(N(i), N(i)) + \mathbf{W}^i$, onde $\mathbf{K}(N(i), N(i))$ denota a submatriz de \mathbf{K} contendo as linhas e colunas com os índices em $N(i)$.

Passo 5. Decomposição espectral do *kernel* \mathbf{K} . Seja $\{Y^0, Y^1, \dots, Y^d\} \subset \mathbb{R}^n$ como os $d+1$ autovetores correspondentes aos $d+1$ menores autovalores de \mathbf{K} . Então o conjunto dimensionalmente reduzido é dado por $\mathbf{Y} = [Y^1; Y^2; \dots; Y^d]'$.

3.9 Diffusion Maps

Proposto por, [Coifman e Lafon \(2006\)](#), o método *Diffusion Maps* baseia-se na ideia de que dada uma família de mapas de difusão, no qual cada um representará o conjunto de dados dentro de um espaço Euclidiano de modo que a distância Euclidiana no espaço seja igual à distância de difusão computada sobre os dados. As várias distâncias de difusão induzidas pelos mapas de difusão descrevem as estruturas geométricas multiescalas sobre os dados, representados por funções características. Adotamos o mesmo modelo de dados como anteriormente, assumindo que o conjunto de dados $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{R}^D$ reside sobre uma variedade d -dimensional $\subset \mathbb{R}^D$, que tem uma parametrização g . Tome $h \approx g^{-1} = [h^1, \dots, h^d]'$ como o mapa característico sobre a variedade. O conjunto de dados dimensionalmente reduzido é obtido como a representação característica de $\mathcal{X} : \mathcal{Y} = h(\mathcal{X})$.

3.9.1 Operador de Difusão sobre a Variedade

O método de redução *Diffusion Maps* é análogo ao método *Laplacian Eigenmaps*. Consequentemente, primeiramente explicamos a relação entre o operador de Laplace-Beltrami \mathcal{L} e o operador de difusão do calor de Neumann \mathcal{A}_t . Lembrando que o operador de Laplace-Beltrami e o operador de Neumann sobre M são relacionados pela fórmula exponencial

$$\mathcal{A}_t = e^{-t\mathcal{L}}. \quad (3.62)$$

Como \mathcal{A}_t tem representação integral

$$\mathcal{A}_t(\mathbf{x}) = \int_M G_t(\mathbf{x}, \mathbf{y}) f(\mathbf{y}), \quad f \in C(M), \quad (3.63)$$

e pode ser estendido sobre um grande espaço de funções, ou seja, pode ser considerado como um operador sobre $L^2(M)$. Seja o produto interno sobre $L^2(M)$ denotado por $\langle \cdot, \cdot \rangle$. A decomposição espectral de \mathcal{L} nos dá

$$\mathcal{L}(f) = \sum_{i=0}^{\infty} \lambda \phi^i \langle \phi^i, f \rangle, \quad (3.64)$$

onde

$$0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_n \rightarrow \infty$$

são autovalores de \mathcal{L} e $\phi^i, : 0 \leq i < \infty$, são as correspondentes autofunções.

O método *Diffusion Maps* usa a decomposição espectral de \mathcal{A}_t para encontrar o conjunto de dados dimensionalmente reduzido. Na prática é usado o seguinte modelo de maximização. Para um certo t ,

$$h = \arg \max \langle f, \mathcal{A}_t(f) \rangle, : s.t. \mathbf{F}'\mathbf{F} = \mathbf{I}, \quad (3.65)$$

cujo resultado é o mesmo obtido no método *Laplacian Eigenmaps*, então setamos $\mathbf{F} = [e, h^1, \dots, h^d]$.

3.9.2 Normalização de *kernel* de Difusão

A construção do *kernel* para o método *Diffusion Maps* é muito similar ao do método *Laplacian Eigenmaps*. O *kernel* do método *Diffusion Maps* é derivado do operador de Neumann o qual pode ser aproximado pela Gaussiano. Conseqüentemente, o método pode ser construído normalizando o Gaussiano. Construimos um *kernel* simétrico, substituindo a necessidade de se ter as linhas com soma igual a 1, por uma condição mais fraca: é requerido então que 1 seja o maior autovalor do *kernel*. Construimos dois diferentes tipos de *kernels* de difusão: um tipo Grafo-Laplaciano e o outro do tipo Laplace-Beltrami. Seja A_t o Gaussiano não normalizado:

$$A_t(\mathbf{x}, \mathbf{y}) = e^{-\frac{\|\mathbf{x}-\mathbf{y}\|}{t}},$$

que gera um funcional bilinear $\langle g, \mathcal{A}(f) \rangle$ sobre $L^2(M) \times L^2(M)$ tendo a seguinte forma

$$\langle g, \mathcal{A}(f) \rangle = \int_M \int_M A_t(\mathbf{x}, \mathbf{y}) f(\mathbf{x}) g(\mathbf{y}).$$

Escreva

$$S(\mathbf{x}) = \int_M A_t(\mathbf{x}, \mathbf{y}) d\mathbf{y}$$

e normalize $A_t(\mathbf{x}, \mathbf{y})$ por

$$\tilde{A}_t(\mathbf{x}, \mathbf{y}) = \frac{A_t(\mathbf{x}, \mathbf{y})}{S(\mathbf{x})}$$

Então $\tilde{A}_t(\mathbf{x}, \mathbf{y})$ tem a propriedade de ter suas linhas somando 1:

$$\int_M \tilde{A}_t(\mathbf{x}, \mathbf{y}) d\mathbf{y} = 1.$$

Mas $\tilde{A}_t(\mathbf{x}, \mathbf{y})$ não é simétrico. O *kernel* \tilde{A}_t gera o operador

$$\tilde{A}_t = \int_M \tilde{A}_t(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) d\mathbf{y}.$$

Para construir um *kernel* simétrico, setamos

$$\hat{A}_t(\mathbf{x}, \mathbf{y}) = \frac{A_t(\mathbf{x}, \mathbf{y})}{\sqrt{S(\mathbf{x})S(\mathbf{y})}},$$

o qual gera o operador

$$\hat{\mathcal{A}}_t(f) = \int_M \hat{A}_t(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) d\mathbf{y}.$$

Podemos ver que os operadores $\tilde{\mathcal{A}}_t$ e $\hat{\mathcal{A}}_t$ têm o mesmo conjunto de autovalores. Além disso, f é um autovetor de $\tilde{\mathcal{A}}_t$ correspondente ao autovalor λ se e somente se $\frac{f}{\sqrt{S}}$ é um autovetor de $\hat{\mathcal{A}}_t$ correspondente ao mesmo autovalor.

Um outro tipo de normalização para A_t emprega o *kernel* simétrico

$$A_t^{sim}(\mathbf{x}, \mathbf{y}) = \frac{A_t(\mathbf{x}, \mathbf{y})}{S(\mathbf{x})S(\mathbf{y})}.$$

Tome

$$S^{sim}(\mathbf{x}, \mathbf{y}) = \int_M A_t^{sim}(\mathbf{x}, \mathbf{y}) d\mathbf{y},$$

que normaliza A_t^{sim} para o *kernel Diffusion Maps*

$$\bar{A}_t = \frac{A_t^{sim}(\mathbf{x}, \mathbf{y})}{\sqrt{S^{sim}(\mathbf{x})S^{sim}(\mathbf{y})}}, \quad (3.66)$$

correspondendo ao operador

$$\bar{\mathcal{A}}_t = \int_M \bar{A}_t(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) d\mathbf{y}. \quad (3.67)$$

Porque o infinitesimal de $\bar{\mathcal{A}}_t$ é \mathcal{L} :

$$s - \lim_{t \rightarrow 0} \frac{\bar{\mathcal{A}}_t - I}{t} = \mathcal{L},$$

chamamos \bar{A}_t o *kernel* de difusão do tipo Laplace-Beltrami. As discretizações dos *kernels* \hat{A}_t e \bar{A}_t não serão apresentadas.

3.9.3 Algoritmo *Diffusion Maps*

Passo 1. Definição da vizinhança. De forma análoga, aos demais métodos anteriores, a vizinhança é construída usando-se uma das técnicas k -vizinhos ou corte- ϵ .

Criação da matriz de pesos. Assumimos que o grafo $[\mathcal{X}, \mathbf{A}]$ é definido sobre o conjunto de dados \mathcal{X} , e que determina um sistema de vizinhança sobre \mathcal{X} . Considere O_i a vizinhança de \mathbf{x}_i e $N(i)$ o correspondente conjunto de índices correspondendo à vizinhança de \mathbf{x}_i , que é a i -ésima linha de \mathbf{A} . A matriz de pesos $\mathbf{W} = [w_{ij}]_{ij=1}^n$ por

$$w_{ij} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{t}\right), \quad j \in N(i) \cup \{i\}, \quad \text{e } 0 \text{ caso contrário}$$

onde $t > 0$ é um parâmetro. Para criar uma matriz de pesos sem parâmetro, nós escolhemos

$$t = \frac{1}{\sum_{i=1}^n |N(i)|} \sum_{i=1}^n \sum_{j \in N(i)} \|\mathbf{x}_i - \mathbf{x}_j\|^2.$$

Passo 3. Construção do *kernel* de difusão. Nós normalizamos a matriz de pesos para construir o *kernel* de difusão. Setamos $\mathbf{v}_i = \sqrt{\mathbf{W}_i \mathbf{1}}$, onde \mathbf{W}^i é a i -ésima linha de \mathbf{W} , e definimos

$$k_{ij} = \frac{w_{ij}}{v_i v_j}.$$

Passo 4. Decomposição do *kernel* de redução. Seja $\mathbf{v}^0, \mathbf{v}^1, \mathbf{v}^2, \dots, \mathbf{v}^d$ como os autovetores de \mathbf{K} correspondentes aos $(d+1)$ maiores autovalores $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d > 0$. Escale \mathbf{v}_i para

$$\tilde{\mathbf{v}}_i = \left[\frac{v_{1i}}{v_{10}}, \dots, \frac{v_{ni}}{v_{n0}} \right], \quad 1 \leq i \leq d.$$

Então o conjunto de dados reduzido $\mathbf{Y} = [\tilde{\mathbf{v}}_1, \dots, \tilde{\mathbf{v}}_d]'$.

3.10 Maximum Variance Unfolding

Proposto por, [Weinberger, Packer e Saul \(2005\)](#), o método de redução *Maximum Variance Unfolding* também é conhecido como *Semidefinite Embedding* (representação semidefinida). O método utiliza técnicas de programação semidefinida ([WEINBERGER; SAUL, 2004](#)). O *Maximum Variance Unfolding* mantém a máxima variância no processamento de redução de dimensionalidade por meio da preservação de similaridades locais representadas pela matriz de Gram local.

3.10.1 Descrição do Método *Maximum Variance Unfolding*

Assumimos que o conjunto de dados $\mathcal{X} \subset \mathbb{R}^D$ está sobre uma variedade convexa d -dimensional. Seja h um mapa coordenado sobre M de modo que para cada $\mathbf{p} \in M$, $\mathbf{q} = h(\mathbf{p}) \in \mathbb{R}^d$ é a representação em coordenadas de \mathbf{p} . Então $g = h^{-1} : U \subset \mathbb{R}^d \rightarrow M$ é uma parametrização de M . Sem perda de generalidade, assumimos que $dh \stackrel{def}{=} dh_{\mathbf{p}}$ é isométrica, então temos que $dh \in \mathfrak{D}_{d,D}$ com $(dh)^T dh = I$ em cada ponto $\mathbf{p} \in M$. Denotamos a imagem de \mathcal{X} sobre o mapa h por $\mathcal{Y} = h(\mathcal{X})$. Então \mathcal{Y} é o conjunto de dados dimensionalmente reduzido que desejamos encontrar. Lembrando que assumimos que \mathcal{Y} possui média zero:

$$\sum_{j=1}^n y_{i,j} = 0, \quad i = 1, \dots, d. \quad (3.68)$$

Definimos a matriz de Gram de \mathcal{Y} por $\mathbf{K} = \mathbf{Y}'\mathbf{Y}$. Uma vez que \mathbf{K} é criada, \mathbf{Y} pode ser obtida a partir da decomposição espectral de \mathbf{K} . Para a construção do *kernel* \mathbf{K} , o *Maximum Variance Unfolding* usa o seguinte modelo. Seja P_i a vizinhança do ponto $\mathbf{x}_i \in \mathcal{X}$ e Q_i sua imagem $Q_i = h(P_i) \subset \mathbb{R}^d$. Como a vizinhança P_i esta muito próxima do hiperplano tangente de M em \mathbf{x}_i , a geometria local de P_i assemelha-se à de Q_i . Mais precisamente, temos que

$$\|\mathbf{x}_j - \mathbf{x}_k\| \approx \|\mathbf{y}_j - \mathbf{y}_k\|, \quad j, k \in N(i), \quad (3.69)$$

onde $N(i)$ é o conjunto de índices de pontos em P_i . O método *Maximum Variance Unfolding* formula \mathcal{Y} como a solução do seguinte problema de minimização:

$$\mathcal{Y} = \arg \min_{\mathcal{Y} \subset \mathbb{R}^d} \sum_{i=1}^n \sum_{j \sim k} (\|\mathbf{x}_k - \mathbf{x}_j\|^2 - \|\mathbf{y}_k - \mathbf{y}_j\|^2)^2, \quad (3.70)$$

onde $k \sim j$ significa que \mathbf{x}_k e \mathbf{x}_j estão na mesma vizinhança, por exemplo, em P_i . Consideramos também que \mathbf{x}_i na relação \sim para que $k \sim i$ e $j \sim i$. O problema de minimização 3.70 não é convexo. Consequentemente, é difícil para ser resolvido diretamente. O método *Maximum Variance Unfolding* então converte-o em um problema de programação semidefinida

como segue. Assumimos que \mathbf{K} é a matriz de Gram de \mathbf{Y} . Da relação existente entre a matriz de Gram centrada e a métrica Euclidiana, temos que:

$$\|\mathbf{y}_j - \mathbf{y}_k\|^2 = \mathbf{K}_{k,k} + \mathbf{K}_{j,j} - 2\mathbf{K}_{k,j},$$

sujeita às restrições

$$\mathbf{K}_{k,k} + \mathbf{K}_{j,j} - 2\mathbf{K}_{k,j} = S_{k,j}, \quad j, k \in N(i), j > k, 1 \leq i \leq n, \quad (3.71)$$

onde $S_{k,j} = \|\mathbf{y}_j - \mathbf{y}_k\|^2$. Porque \mathcal{Y} está centrada, a equação 3.68 nos fornece a seguinte restrição sobre \mathbf{K} :

$$\sum_{i,j=1}^n \mathbf{K}_{i,j} = 0. \quad (3.72)$$

Finalmente, é requerido que o conjunto de dados \mathcal{Y} tenha variância máxima:

$$\mathcal{Y} = \arg \max_{\mathcal{Y} \subset \mathbb{R}^d} \sum_{i,j=1}^n \|\mathbf{y}_i - \mathbf{y}_j\|^2 = \arg \max_{\mathcal{Y} \subset \mathbb{R}^d} \|\mathbf{D}_Y\|_F^2, \quad (3.73)$$

onde \mathbf{D}_Y é a métrica Euclidiana sobre \mathcal{Y} . Pela maximização 3.73, os pontos de \mathcal{Y} são dispostos o mais distante uns dos outros como possível, sujeito às restrições 3.71 e 3.72. Lembrando que \mathbf{K} é a matriz de Gram centrada de \mathcal{Y} , temos

$$\frac{1}{2n} \|\mathbf{D}_Y\|_F^2 = \text{tr}(\mathbf{G}_Y^c) = \text{tr}(\mathbf{K}).$$

Assim sendo, o problema de maximização 3.73 é finalmente convertido para

$$\mathbf{K} = \arg \max_{\mathbf{K} \in \mathfrak{S}\mathfrak{B}_n} \text{tr}(\mathbf{K}) \quad (3.74)$$

com as restrições 3.71 e 3.72. Combinando-as, obtemos o problema de maximização para o *kernel* \mathbf{K} :

$$\arg \max \text{tr}(\mathbf{K}), \quad \mathbf{K} \in \mathfrak{S}, \quad (3.75)$$

$$s.a. \quad \sum_{i,j=1}^n \mathbf{K}_{i,j} = 0, \quad (3.76)$$

$$\mathbf{K}_{k,k} + \mathbf{K}_{j,j} - 2\mathbf{K}_{k,j} = S_{k,j}, \quad j, k \in N(i), 1 \leq i \leq n, \quad (3.77)$$

$$\mathbf{K} \succeq 0, \quad (3.78)$$

onde $\mathbf{K} \succeq 0$ é uma notação abreviada para dizer que \mathbf{K} é semidefinido positivo.

Como o problema de maximização para \mathbf{K} em 3.75 envolve somente a matriz dos quadrados das distâncias Euclidianas \mathbf{S} , o algoritmo aceita a matriz de distâncias locais \mathbf{D} como entrada. Consequentemente, *Maximum Variance Unfolding* pode ser aplicado em aplicações onde somente as relações de similaridades entre os objetos esteja disponível. O modelo de dados usado pelo método *Maximum Variance Unfolding* define um tipo especial de grafo sobre o conjunto de dados. As restrições 3.71 indicam que as distâncias e ângulos locais entre todos os pontos na vizinhança de \mathbf{x} sejam preservadas. Assim sendo, os pontos em uma mesma vizinhança devem estar mutualmente conectados. Chamamos este tipo de grafo de um grafo localmente completo. De modo geral, para um grafo construído com a técnica k -vizinhos ou corte- ϵ , os nós não estão mutualmente conectados. Um grafo localmente completo pode ser criado pelo refinamento de um grafo usual $G = [\mathcal{X}, \mathbf{E}]$. Seja $G_r = [\mathcal{X}, E_r]$ denota o grafo localmente completo refinado de G . Então $(k, j) \in E_r$ se e somente se $(k, j) \in E$ ou $(k, i), (j, i) \in E$ para algum i .

3.10.2 Algoritmo *Maximum Variance Unfolding*

Passo 1. Definição de vizinhança. Esta etapa é necessária somente se como entrada para o método for fornecido o conjunto de dados \mathbf{X} . Esta etapa é a mesma realizada para o método de redução *Isometric Mapping*. Denotamos o grafo $G = [\mathcal{X}, \mathbf{A}]$. O método *Maximum Variance Unfolding* utiliza a matriz de adjacência \mathbf{A} para determinar os índices das restrições.

Computar a matriz de Gram para o conjunto de dados. Podemos computar a matriz de Gram centralizada \mathbf{G}^c ou a matriz dos quadrados das distâncias Euclidianas \mathbf{S} para construir as restrições para o problema de programação semidefinida.

Geração das restrições para o problema de programação semidefinida. Quando $\mathbf{A}(k, j) = 1$, definimos a restrição

$$\mathbf{K}_{k,k} + \mathbf{K}_{j,j} - 2\mathbf{K}_{k,j} = \mathbf{G}_{k,k} + \mathbf{G}_{j,j} - 2\mathbf{k}, \mathbf{j}$$

ou

$$\mathbf{K}_{k,k} + \mathbf{K}_{j,j} - 2\mathbf{K}_{k,j} = s_{k,j}$$

para o problema de programação semidefinida, dependendo de qual matriz de entrada \mathbf{G} ou \mathbf{S} é disponibilizada. A condição de centralização

$$\sum_{j=1}^n \mathbf{K}_{k,j} = 0$$

deve ser adicionada ao conjunto de restrições.

Passo 4. Construção do *kernel* usando Programação SemiDefinida. Aplicar um algoritmo de programação semidefinida para a criação do *kernel* \mathbf{K} .

Passo 5. Decomposição espectral do *kernel* do *Maximum Variance Unfolding*. Faça a decomposição espectral da matriz \mathbf{K} e compute os d principais autovetores de \mathbf{K} de forma idêntica ao que é feito para o método *Isometric Mapping*.

3.11 Métricas para avaliação da qualidade de redução de dimensionalidade

Após a realização da redução de dimensionalidade é necessário o uso de métricas para a avaliação da qualidade da redução obtida. Com esta finalidade, neste trabalho foram utilizadas quatro métricas. A primeira delas, é o erro de generalização de um classificador 1-NN, que foi treinado sobre a representação dos dados no espaço de baixa dimensionalidade. As demais métricas que foram utilizadas são a *trustworthiness* (ou confiabilidade / fidedignidade), a *continuity* (ou continuidade) e a *QNX* (critério de qualidade independente de escala). Todas as métricas descritas, com exceção do erro de generalização do classificador 1-NN, são obtidas por meio da matriz de *coranking*.

A matriz de *coranking* foi proposta por Lee e Verleysen (2009) e é por meio dela que é obtida as métricas, nomeadas anteriormente, que nos permite inferir a qualidade das reduções obtidas pelas diversas técnicas de redução abordadas. A definição da matriz de *coranking* é dada pela equação 3.11.

Seja δ_{ij} a distância de \mathcal{X}_i para \mathcal{X}_j no espaço original. Analogamente, d_{ij} a distância de \mathcal{Y}_i a \mathcal{Y}_j no espaço reduzido. A partir destas distâncias é possível computar-se as classificações dos vizinhos para cada ponto. A classificação \mathcal{X}_j com relação a \mathcal{X}_i no espaço original é dada por

$$\rho_{ij} = |\{k \mid \delta_{ik} < \delta_{ij} \text{ ou } (\delta_{ik} = \delta_{ij} \text{ e } 1 \leq k < j \leq N)\}|,$$

onde $|A|$ é a cardinalidade do conjunto A , resultado das classificações $1, \dots, N - 1$. Analogamente, a classificação de \mathcal{Y}_i com relação a \mathcal{Y}_j no espaço reduzido é dada por

$$r_{ij} = |\{k \mid d_{ik} < d_{ij} \text{ ou } (r_{ik} = r_{ij} \text{ e } 1 \leq k < j \leq N)\}|.$$

As métricas para avaliação da qualidade, da redução obtida, baseiam-se em medir o quanto as classificações dos pontos são preservadas no espaço de baixa dimensão. Desta forma, os relacionamentos locais são avaliados sem a interferência de questões como, por exemplo, o dimensionamento dos dados.

A matriz de *coranking* Q é definida por

$$Q_{kl} = |\{(i, j) \mid \rho_{ij} = k \text{ e } r_{ij} = l\}|.$$

e é dividida em quatro blocos conforme ilustra a figura 1.

Os erros ocorridos durante o processo de redução são identificados como entradas fora da diagonal principal.

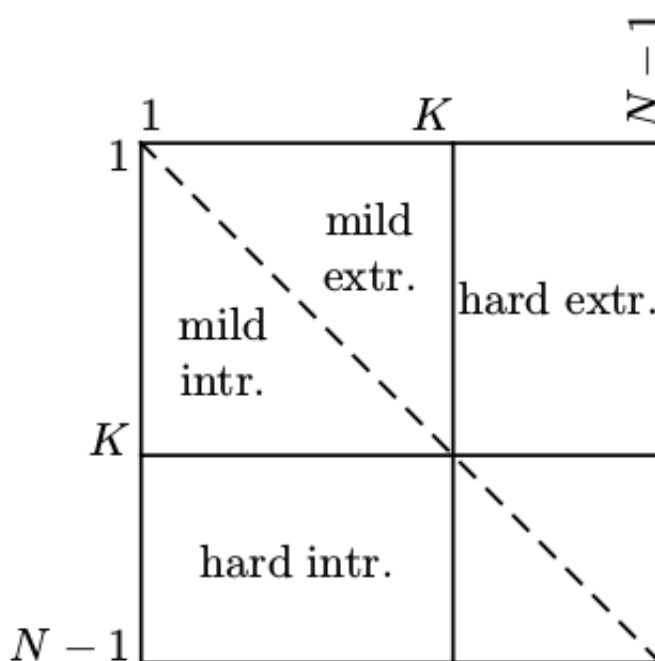


Figura 1 – Matriz de Coranking

Estes erros são denominados de k -intrusões e k -extrusões. Uma k -intrusão ocorre quando a classificação de um vértice i com relação a um vértice vizinho j no espaço de alta dimensão é maior que no espaço de dimensão reduzida. Já uma k -extrusão, ocorre quando a classificação de um vértice i com relação a um vértice vizinho j no espaço de dimensão reduzida é maior que sua classificação no espaço original.

Os autores [Lee e Verleysen \(2009\)](#) ainda fazem uma subdivisão entre as k -intrusões e k -extrusões, classificando-as em k -intrusões/ k -extrusões suaves e difíceis (*hard*). As k -intrusões e k -extrusões suaves são aquelas nas quais a classificação de um vértice i com relação a um vértice vizinho j é alterada, porém seu valor é menor ou igual ao valor do tamanho da vizinhança, ou seja, o movimento ocorre dentro da K -vizinhança previamente estabelecida. Dentro deste cenário, temos as K -intrusões suaves que são eventos para os quais $r_{ij} < \rho_{ij} \leq K$, as K -extrusões suaves onde $\rho_{ij} < r_{ij} \leq K$. Já as K -intrusões difíceis são definidas por $r_{ij} \leq K < \rho_{ij}$, enquanto as K -extrusões difíceis por $\rho_{ij} \leq K < r_{ij}$.

Das definições anteriores são extraídas as métricas de confiabilidade (*trustworthiness*), continuidade (*continuity*) e métrica independente de escala (*QNX*). A confiabilidade mede as k -intrusões difíceis, a continuidade as k -extrusões difíceis e a métrica independente de escala mede as k -intrusões e k -extrusões suaves.

A métrica independente de escala (*QNX*), ([LEE; VERLEYSEN, 2010](#)), é definida do seguinte modo:

$$Q_{NX}(K) = \frac{1}{KN} \sum_{k=1}^K \sum_{l=1}^K Q_{kl}.$$

Note que as métricas de qualidade são dependentes do valor de K que é o tamanho da vizinhança previamente estabelecido. Com o objetivo de definir uma medida que fosse independente do valor de K e fornecesse um único valor escalar que quantificasse a qualidade da redução, os autores Lee e Verleysen propuseram uma métrica local e global definida a partir da $QNX(K)$. Para a definição de quais valores seriam considerados locais e quais globais, (LEE; VERLEYSSEN, 2009) e (LEE; VERLEYSSEN, 2010), foi proposto o seguinte ponto para particionamento:

$$\arg \max_K LCMC(k) = \arg \max_K QNX(K) - \frac{K}{N-1}$$

A partir do ponto de separação é obtida uma medida de qualidade definida sobre a média dos valores de $QNX(K)$. Então tem-se Q_{local} definido do seguinte modo:

$$Q_{local} = \frac{1}{K_{max}} \sum_{k=1}^{K_{max}} QNX(K),$$

E de modo análogo, temos para Q_{global} :

$$Q_{global} = \frac{1}{N - K_{max}} \sum_{k=K_{max}}^{N-1} QNX(K),$$

A métrica de qualidade independente de escala ($QNX(K)$) foi baseada na métrica local definida como Meta Critério de Continuidade Local (*Local Continuity Meta Criterio*) que é apresentada na seguinte referência (CHEN; BUJA, 2006).

Utilizando a definição 3.11 da matriz de *coranking*, as métricas de confiabilidade e continuidade são definidas. Para a confiabilidade temos a seguinte definição

$$T(K) = 1 - \frac{2}{G_K} \sum_{(k,l) \in LL_k} (k - K)q_{kl},$$

De modo análogo, para a medida de continuidade temos

$$C(K) = 1 - \frac{2}{G_K} \sum_{(k,l) \in UR_k} (l - K)q_{kl},$$

onde o fator de normalização G_K é definido como

$$G_K = \begin{cases} NK(2N - 3K - 1), & \text{se } K < \frac{N}{2} \\ N(N - K)(N - K - 1), & \text{se } K \geq \frac{N}{2} \end{cases}$$

É importante notar que as métricas de confiabilidade, continuidade e a medida independente de escala são todas dependentes do tamanho da vizinhança estabelecido e são definidas no intervalo de 0 a 1, onde o valor 1 significa um mapeamento perfeito, ou seja, a topologia local definida para cada um dos vértices e seus respectivos k -vizinhos é preservada nos espaços de alta e baixa dimensão.

4 Experimentos

Os experimentos realizados, durante a execução deste trabalho, tiveram como objetivos o estudo e a análise comparativa das principais técnicas de redução de dimensionalidade lineares e não lineares; avaliar-se-á o poder de redução de cada uma das técnicas por meio das métricas descritas no capítulo . Na seção 4.1 é apresentado o procedimento para seleção dos conjuntos de dados naturais, a seção 4.2 descreve o processo de geração de dados artificiais e a seção 4.3 apresenta a seleção de parâmetros utilizados para os experimentos.

4.1 Seleção de conjuntos de dados naturais

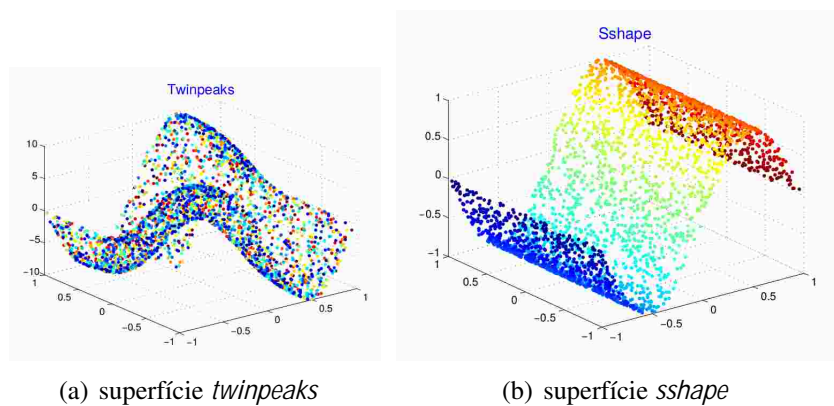
Para a realização dos experimentos foram selecionados um total de dez conjuntos de dados naturais. Os conjuntos de dados estão listados na tabela 4.1 e foram obtidos do repositório online *OpenML* (VANSCHOREN et al., 2014). Na tabela 4.1, com os conjuntos de dados selecionados, além do nome do conjunto é exibido o número de instâncias, o número de características e o número de classes. Estes conjuntos foram selecionados utilizando como critérios de seleção o número elevado de características, o fato de não apresentarem valores de atributos ausentes e por possuírem todos os atributos numéricos. Por limitações computacionais foram selecionados 3000 exemplos para a aplicação da técnica de redução *Diffusion Maps* e 5000 exemplos para a aplicação das demais técnicas sobre os conjuntos de dados com um número de exemplos superior a 5000. A seleção dos exemplos foi realizada de forma aleatória e estratificada utilizando-se os módulos de divisão de conjuntos de dados presentes na biblioteca *Scikit Learn* (PEDREGOSA et al., 2011).

Nome	Nº. de instâncias	Nº. de características	Nº. de classes
Eating	941	6373	7
Gas drift	13910	128	6
Hill valley	1212	100	2
Isolet	7767	617	26
Mfeat factors	2000	216	10
Mfeat zernike	2000	47	10
Mnist	70000	784	10
Nomao	34465	118	2
Ozone level 8-hr	2534	72	2
USPS	9298	256	10

Tabela 1 – Descrição dos conjuntos de dados naturais

4.2 Geração de conjuntos de dados artificiais

Para a avaliação de desempenho, das técnicas de redução, sobre conjuntos de dados artificiais serão utilizadas um total de sete superfícies, sendo que cinco delas foram geradas com o software R e o pacote *Ider* (HINO, 2017), e são nomeadas, respectivamente, de *cylinder*, *moebius*, *sinusoidal*, *sshape*, *swiss roll*. As superfícies *twinpeaks* e *broken swiss* foram gerados com o *toolbox* desenvolvido em Matlab (MAATEN, 2007). A superfície *sinusoidal* possui dimensionalidade intrínseca igual a 1 e as demais iguais a 2. Todas as superfícies são representações de variedades mergulhadas em um espaço euclidiano tridimensional. As figuras 2(a) e 2(b) são exemplos de superfícies artificiais em um espaço tridimensional.



4.3 Seleção de parâmetros

Para a realização dos experimentos, o parâmetro que define o tamanho da vizinhança, definido por K , variou de 3 (três) a 50 (cinquenta). Para a aplicação do método de redução Mapas de Difusão utilizou-se o valor de 1 (um) para a variância do *kernel* gaussiano.

Para a definição do valor de dimensionalidade para redução aplicou-se a técnica Análise de Componentes Principais (PCA) com manutenção da variância igual a 95,0%, e ao número de componentes mínimos necessários, retornados pela técnica somou-se várias unidades resultando no valor de dimensionalidade de redução a ser utilizado para cada um dos conjuntos de dados naturais. Para os artificiais o valor da dimensionalidade alvo é igual a 2 (dois). Na tabela abaixo, é exibido o número de componentes fornecido pelo PCA, o valor da dimensionalidade para redução e o valor da variância retida.

Note que para os conjuntos de dados *USPS*, *Mnist* e *Isolet* o valor da dimensionalidade para redução utilizada não é igual ou superior ao número mínimo de componentes retornados pela técnica PCA, pois por se tratar de conjuntos com um número grande de instâncias e de dimensões, por limitações computacionais o valor da dimensionalidade usada nos experimentos foi reduzida.

Nome	Número de componentes	D para redução	Variância
Eating	1	25	100,0%
Gas drift	9	25	100,0%
Hill valley	1	25	99,9%
Isolet	170	37	80,0%
Mfeat factors	12	25	98,6%
Mfeat zernike	13	25	100,0%
Mnist	149	49	82,4%
Nomao	22	35	99,1%
Ozone level 8-hr	4	25	99,9%
USPS	33	25	92,1%

Tabela 2 – Dimensionalidade para redução aplicada aos conjuntos naturais

5 Resultados e discussão para conjuntos de dados artificiais

Neste capítulo são exibidos os resultados dos experimentos obtidos sobre os conjuntos de dados artificiais. São mostradas as tabelas para as métricas de confiabilidade, continuidade, erro de generalização e medida independente de escala para cada um dos métodos abordados. Somente a melhor execução é reportada.

5.1 Medidas de confiabilidade

Superfície	HLLE	ISOMAP	LE	LLE	LTSA	DM	MVU	MDS	PCA
<i>cylinder</i>	0,889	0,899	0,910	0,878	0,889	0,893	0,897	0,894	0,894
<i>moebius</i>	0,993	0,998	0,949	0,985	0,993	0,999	0,986	0,999	0,999
<i>sinusoidal</i>	1,000	1,000	0,996	1,000	1,000	1,000	0,998	1,000	1,000
<i>sshape</i>	1,000	1,000	0,966	1,000	1,000	0,957	0,999	0,957	0,957
<i>swiss roll</i>	1,000	1,000	0,833	1,000	1,000	0,885	0,875	0,884	0,884
<i>twinpeaks</i>	0,989	0,994	0,970	0,991	0,989	0,971	0,988	0,983	0,983
<i>broken swiss</i>	0,497	0,989	0,971	0,992	0,768	0,978	0,989	0,979	0,977

Tabela 3 – Medidas de confiabilidade para superfícies artificiais

5.2 Medidas de continuidade

Superfície	HLLE	ISOMAP	LE	LLE	LTSA	DM	MVU	MDS	PCA
<i>cylinder</i>	0,999	0,999	0,991	0,999	0,999	0,999	0,999	0,999	0,999
<i>moebius</i>	1,000	0,999	0,987	0,999	1,000	1,000	0,999	1,000	1,000
<i>sinusoidal</i>	1,000	1,000	0,988	1,000	1,000	1,000	0,999	1,000	1,000
<i>sshape</i>	1,000	1,000	0,991	1,000	1,000	0,999	1,000	0,999	0,999
<i>swiss roll</i>	1,000	1,000	0,977	1,000	1,000	0,999	0,995	0,999	0,999
<i>twinpeaks</i>	0,996	0,999	0,976	0,998	0,996	0,994	0,998	0,999	0,999
<i>broken swiss</i>	0,552	0,969	0,979	0,995	0,896	0,999	0,996	0,999	0,985

Tabela 4 – Medidas de continuidade para superfícies artificiais

5.3 Erro de generalização

Superfície	ORIGINAL	HLL	ISOMAP	LE	LLE	LTSA	DM	MVU	MDS	PCA
cylinder	40,02%	56,85%	54,35%	51,85%	60,22%	57,18%	55,58%	54,95%	54,78%	55,05%
moebius	80,93%	81,00%	81,00%	82,25%	81,88%	81,02%	80,22%	81,78%	80,85%	80,20%
sinusoidal	79,72%	80,00%	79,80%	79,42%	80,08%	79,65%	80,07%	80,38%	80,45%	80,45%
sshape	2,03%	3,40%	2,25%	21,63%	3,67%	35,83%	39,72%	2,58%	40,53%	41,03%
swiss roll	4,05%	5,00%	3,80%	37,55%	5,33%	4,95%	32,20%	34,98%	32,12%	31,50%
twinpeaks	0,73%	0,23%	0,75%	4,30%	0,73%	0,23%	0,18%	1,05%	0,32%	0,33%
broken swiss	18,88%	86,00%	21,38%	30,75%	18,97%	61,10%	29,55%	20,60%	28,68%	29,35%

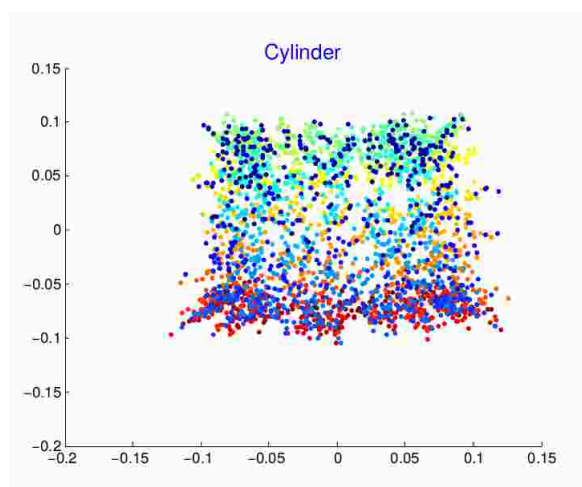
Tabela 5 – Erro de generalização para superfícies artificiais

5.4 Medida independente de escala

Superfície	HLLE	ISOMAP	LE	LLE	LTSA	DM	MVU	MDS	PCA
<i>cylinder</i>	0,543	0,556	0,412	0,513	0,543	0,549	0,543	0,527	0,527
<i>moebius</i>	0,813	0,720	0,340	0,789	0,813	0,828	0,682	0,830	0,830
<i>sinusoidal</i>	0,997	1,000	0,454	0,999	0,996	1,000	0,892	1,000	1,000
<i>sshape</i>	0,758	0,957	0,463	0,751	0,758	0,576	0,922	0,625	0,625
<i>swiss roll</i>	0,685	0,951	0,239	0,759	0,685	0,462	0,321	0,452	0,452
<i>twipeaks</i>	0,613	0,710	0,312	0,622	0,613	0,386	0,626	0,672	0,672
<i>broken swiss</i>	0,033	0,686	0,378	0,430	0,152	0,648	0,478	0,606	0,579

Tabela 6 – Medida independente de escala para superfícies artificiais

Dos resultados exibidos nas tabelas 5.1 e 5.2 podem ser feitas algumas observações. A primeira delas é com relação ao fraco desempenho dos métodos de redução não lineares sobre a superfície *cylinder*. Este fato pode ser justificado devido à dificuldade encontrada pelos métodos não lineares de "mergulharem" superfícies naturalmente circulares em um espaço euclidiano, o que pode ser visto na seguinte referência (LEE; VERLEYSEN, 2005). A figura 2 mostra, como exemplo, a superfície *cylinder*, reduzida para duas dimensões, com o método *Laplacian Eigenmaps*.

Figura 2 – Superfície *cylinder* reduzida com método *Laplacian Eigenmaps*

A segunda é com relação ao desempenho insatisfatório de métodos não lineares como o *Hessian Locally Linear Embedding* e o *Local Tangent Space Alignment* quando aplicados a superfícies ou variedades desconectadas como o *broken swiss*. Este fato pode ser justificado pelo fato de que estes métodos são construídos sobre a hipótese de que as variedades são suaves, ou seja, que não possuem descontinuidades ou componentes desconectados, são variedades diferenciáveis. A figura 3 mostra, como exemplo, a superfície *broken swiss*, reduzida para duas dimensões com o método *Hessian Locally Linear Embedding*.

A terceira observação diz respeito aos altos valores de confiabilidade obtidos pela aplicação de métodos não lineares como *Isometric Mapping*, *Maximum Variance Unfolding*, *Lo-*

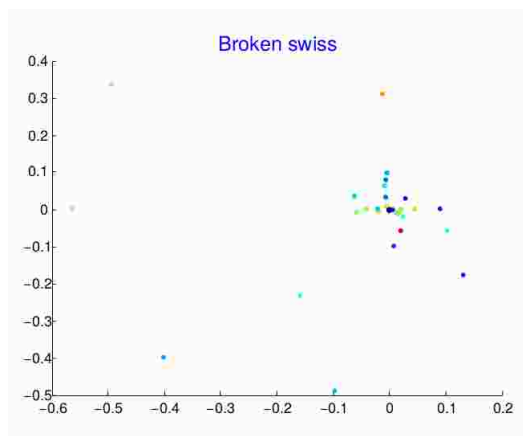


Figura 3 – Superfície *broken swiss* reduzida com método *Hessian Locally Linear Embedding*

cally Linear Embedding, *Laplacian Eigenmaps*, *Hessian Locally Linear Embedding* e *Local Tangent Space Alignment* quando aplicados a superfícies isométricas ao espaço euclidiano como é o caso do *swiss roll*, o mesmo não ocorre quando métodos lineares são utilizados. A figura 4 mostra, como exemplo, a superfície *swiss roll*, reduzida para duas dimensões, com o método *Isometric Mapping*.

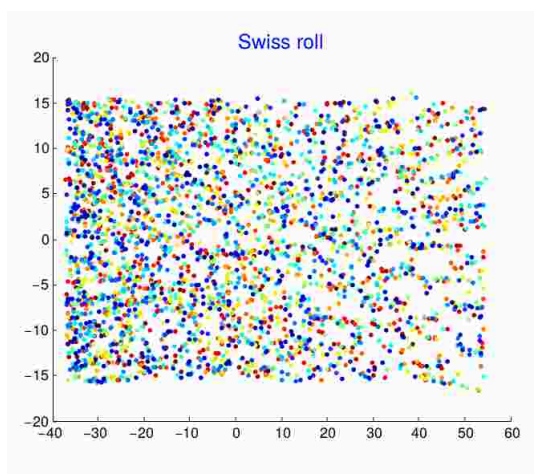


Figura 4 – Superfície *swiss roll* reduzida com método *Isometric Mapping*

Os resultados, exibidos na tabela 5.3, demonstram que o erro de generalização computado sobre as superfícies artificiais reduzidas para um espaço bidimensional foi superior ao apresentado sobre as superfícies no espaço original. Os menores erros de generalização foram obtidos com as reduções realizadas pelos métodos *Isometric Mapping* e *Diffusion Maps*. Os menores valores, para o erro de generalização, são exibidos em negrito.

Diante dos resultados obtidos pela aplicação da métrica independente de escala, apresentados na tabela 5.4, podemos observar que o método *Isometric Mapping* teve o melhor desempenho quando aplicado às superfícies artificiais, notadamente para a superfície *broken swiss*. O método *Isometric Mapping* apresentou uma alta capacidade de preservação da vizinhança nos espaços original e reduzido.

6 Resultados e discussão para conjuntos de dados naturais

Com o objetivo de demonstrar a diferença estatística entre os diversos métodos de redução de dimensionalidade abordados foi realizado o teste não paramétrico de Friedman, utilizando como medida a métrica de confiabilidade. Para a realização do teste foi utilizado os resultados das métricas de confiabilidade para as reduções de 2 (duas) a 25 (vinte e cinco) dimensões para todos os conjuntos de dados. O teste de Friedman foi computado usando-se o software de código aberto, KEEL (*Knowledge Extraction based on Evolutionary Learning*) (ALCALÁ-FDEZ et al., 2009), (ALCALÁ-FDEZ et al., 2011), que é uma ferramenta Java utilizada para diversas tarefas de descoberta de conhecimento em bases de dados; além de possuir um módulo com vários testes estatísticos não paramétricos e procedimentos de correção para análise de comparação múltipla.

Na tabela 6 é mostrado a classificação dos diversos métodos retornada pelo teste de Friedman.

Método de redução	Classificação
PCA	2,2063
MDS	2,3229
ISOMAP	2,9833
LLE	5,65
HLLE	5,7917
MVU	5,8167
LE	6,35
MD	6,7583
LTSA	7,1208

Tabela 7 – Teste de Friedman para conjuntos naturais

O teste de Friedman faz uso da distribuição de probabilidade Qui-Quadrado e forneceu como resposta o valor de 965,3475 para a estatística computada com 8 graus de liberdade.

O p-valor retornado pelo teste, com nível de significância de 5%, foi de 3,2049E-10.

O valor tabelado da distribuição Qui-Quadrado com 8 graus de liberdade e 5% de nível de significância é de 15,507. Como o valor, obtido pelo teste de Friedman, é superior ao valor tabelado, a hipótese nula é rejeitada. O teste considera como hipótese nula a igualdade de desempenho dos diversos métodos de redução; uma vez que a hipótese nula foi rejeitada conclui-se que estatisticamente existe diferenças entre as respostas médias fornecidas pelos diversos métodos de redução.

Uma vez que a hipótese nula foi rejeitada, a fim de detectar entre quais métodos po-

demostre estabelecer uma diferença estatística de seus desempenhos, realizou-se a avaliação *post hoc*, com correção de p-valores realizada de acordo com o procedimento de Nemenyi. O procedimento de Nemenyi rejeita todas as hipóteses que possuem p-valores não corrigidos $p \leq 0,001389$.

A tabela 6 mostra todas as comparações pareadas entre os diversos métodos e seus respectivos p-valores não corrigidos e corrigidos pelo procedimento de Nemenyi. As hipóteses rejeitadas são exibidas em negrito, ou seja, são aqueles métodos nos quais pode-se afirmar estatisticamente a existência de diferença entre seus desempenhos.

i	hipótese	p não ajustado	$p_{Nemenyi}$
1	LTSA x PCA	0	0
2	LTSA x MDS	0	0
3	MD x PCA	0	0
4	MD x MDS	0	0
5	LE x PCA	0	0
6	ISOMAP x LTSA	0	0
7	LE x MDS	0	0
8	ISOMAP x MD	0	0
9	MVU x PCA	0	0
10	HLLE x PCA	0	0
11	MVU x MDS	0	0
12	HLLE x MDS	0	0
13	LLE x PCA	0	0
14	ISOMAP x LE	0	0
15	LLE x MDS	0	0
16	ISOMAP x MVU	0	0
17	HLLE x ISOMAP	0	0
18	ISOMAP x LLE	0	0
19	LLE x LTSA	0	0
20	HLLE x LTSA	0	0,000004
21	LTSA x MVU	0	0,000007
22	LLE x MD	0,000009	0,000334
23	HLLE x MD	0,00011	0,003972
24	MD x MVU	0,000165	0,005956
25	ISOMAP x PCA	0,001881	0,067732
26	LE x LTSA	0,002047	0,07369
27	LE x LLE	0,00511	0,183969
28	ISOMAP x MDS	0,00825	0,296997
29	HLLE x LE	0,025527	0,918971
30	LE x MVU	0,032897	1,184306
31	LE x MD	0,102399	3,686361
32	LTSA x MD	0,147059	5,294107
33	LLE x MVU	0,504985	18,179463
34	HLLE x LLE	0,570941	20,553864
35	MDS x PCA	0,640738	23,066582
36	HLLE x MVU	0,920344	33,132396

Tabela 8 – p-valores corrigidos

Da classificação obtida pelo teste de Friedman, exibida na tabela 6, podemos observar que o desempenho de métodos lineares como *Principal Components Analysis* e *Classical Multidimensional Scaling* sobre conjuntos de dados naturais é superior ao desempenho apresentado por métodos não lineares. A partir das comparações pareadas, exibidas na tabela 6, podemos observar que para *ISOMAP x PCA*, *ISOMAP x MDS* e *PCA x MDS* não existe diferença estatística significativa entre seus desempenhos. Este fato também é observado na tabela 6. Do ponto de vista teórica, o fato do teste de Friedman não ter detectado diferenças estatísticas significativas quanto aos pares citados, pode ser justificado pela semelhança existente entre os métodos.

A redução de dimensionalidade obtida com o *Principal Components Analysis* é resultado da decomposição espectral da matriz de covariância dos dados, enquanto que para o *Classical Multidimensional Scaling* o resultado é obtido pela decomposição espectral da matriz de Gram. O método *Isometric Mapping* utiliza o mesmo *kernel* utilizado pelo método *Classical Multidimensional Scaling* com a diferença de que a matriz de distâncias euclidianas computadas pelo *Classical Multidimensional Scaling* no *Isometric Mapping* é substituída por uma estimativa das distâncias geodésicas computadas por meio de um algoritmo de caminhos mínimos sobre grafos como o Dijkstra. Os métodos *Principal Components Analysis*, *Classical Multidimensional Scaling* e *Isometric Mapping* obtêm as coordenadas no espaço dimensionalmente reduzido por meio dos autovetores computados associados aos maiores autovalores de suas respectivas matrizes *kernel*.

Na tabela 6 podemos ver para cada um dos conjuntos de dados e para cada método o menor valor de D (dimensão de redução) alcançado que manteve 95,0% de variância e os maiores valores para a métrica de confiabilidade.

Conjunto de dados	HLLE	ISOMAP	LE	LLE	LTSA	MD	MVU	MDS	PCA
Hill valley	2	2	3	2	3	2	2	2	2
Eating	2	2	8	3	4	6	3	2	2
Ozone	6	5	5	6	7	6	8	4	4
Factors	7	5	15	7	10	14	19	5	5
Zernike	7	6	14	14	9	14	14	6	6
USPS	8	7	11	10	15	7	16	8	8
Mnist	9	9	40	18	22	41	26	10	10
Isolet	11	8	13	20	20	18	7	8	8
Gas	17	9	8	11	19	9	7	11	11
Nomao	19	14	7	17	3	17	15	22	22

Tabela 9 – Menor dimensão de redução

A partir da análise da tabela 6 podemos observar que os métodos que obtiveram os menores valores para a dimensão de redução foram os métodos lineares *Principal Components Analysis*, *Classical Multidimensional Scaling*, seguido do método não linear *Isometric Mapping*.

7 Conclusão

A partir dos resultados dos experimentos exibidos no apêndice e da análise estatística realizada no capítulo 6, podemos destacar os principais pontos fortes e fracos dos métodos de redução abordados. Os métodos de redução descritos, neste trabalho, são classificados como métodos espectrais podendo ser esparsos ou densos dependendo da sua matriz núcleo (*kernel*). Na categoria de métodos espectrais densos estão inclusos o *Isometric Mapping*, o *Maximum Variance Unfolding* e o *Diffusion Maps*. O *Isometric Mapping* é um método de redução cujo objetivo é construir um mapa isométrico de modo que as distâncias definidas sobre o conjunto de dados seja preservada. Desse modo, a distância euclidiana sobre o conjunto de dados dimensionalmente reduzido é igual à distância geodésica computada sobre o conjunto de dados original. Comumente apresenta resultados satisfatórios desde que a variedade sobre a qual é aplicado seja convexa, ou seja, que não apresente cavidades. Um dos principais pontos fracos apresentado pelo *Isometric Mapping* é o problema conhecido na literatura como *short-circuiting*. *short-circuiting* ocorre quando a variedade é dobrada e suas dobras tornam-se mais próximas do que a distância existente definida pela estrutura de vizinhança, que frequentemente é construída usando-se um grafo k -NN, onde k é o número de vizinhos. Com o intuito de solucionar os problemas decorrentes da existência de *short-circuiting* foram propostas modificações na estrutura original do método que podem ser encontradas nas seguintes referências (SAXENA; GUPTA; MUKERJEE, 2004) e (CHOI; CHOI, 2007). Problema semelhante também é apresentado pelo método *Maximum Variance Unfolding*.

Com relação aos pontos fracos apresentados pelo método *Diffusion Maps* podemos citar seu fraco desempenho causado pela função *kernel* utilizada. Em particular o uso do *kernel* gaussiano com um valor pequeno para a variância pode tornar o algoritmo susceptível à maldição da dimensionalidade intrínseca. Seu desempenho torna-se inversamente proporcional à sua dimensionalidade intrínseca, ou seja, quanto maior for sua dimensionalidade intrínseca menor será sua performance. Uma vantagem que pode ser relatada quanto ao *Diffusion Maps* é que o mesmo soluciona o problema de *short-circuiting* apresentado pelo *Isometric Mapping* e *Maximum Variance Unfolding* por realizar a integração de todos os caminhos construídos através do grafo definido a partir dos dados.

Quanto aos métodos espectrais esparsos sua principal fraqueza refere-se à definição de suas funções objetivos. Como exemplo, a solução ótima para a função de custo do *Locally Linear Embedding* é a solução trivial na qual as coordenadas de todos os pontos de seu espaço reduzido são zeros. Esta solução não é selecionada porque o *Locally Linear Embedding* possui uma restrição definida sobre sua função objetivo de modo que $YY' = 1$, onde Y é o conjunto obtido após a redução. Problema semelhante ocorre com os métodos *Hessian Locally Linear Embedding*, *Laplacian Eigenmaps* e *Local Tangent Space Alignment* que tendem a construir

soluções nas quais a maioria dos pontos são mergulhados para a origem ou colapsados para um único ponto.

Uma importante fraqueza comum a métodos espectrais esparsos é com relação à resolução de um autoproblema, no qual a solução são os menores autovalores. Pelo fato dos autovalores procurados serem muito pequenos os algoritmos utilizados para seus cálculos muitas vezes são incapazes de diferenciar entre os menores autovalores e o autovalor trivial (zero) fazendo com que os métodos retornem soluções sub-ótimas.

Em conjuntos de dados reais um outro problema comumente encontrado é chamado de *folding* que é causado por um valor de k muito alto com relação à amostragem de densidade da variedade. O problema de *folding* causa a violação da hipótese de linearidade local fazendo com que os métodos retornem soluções errôneas devido à distorções causadas na estrutura da variedade. O problema é comum a conjuntos de dados reais porque nestes a densidade dos dados não é uniforme. Para conjuntos de dados com dimensionalidade intrínseca baixa, uma solução para o problema de *folding* é o uso de seleção de vizinhança adaptativa.

Continuando a descrever os pontos fracos dos métodos espectrais esparsos, podemos citar sua sensibilidade à presença de ruído ou outliers nos dados. Em métodos de redução locais, outliers são conectados aos k vizinhos mais próximos, quando na verdade eles estão distantes. Assim sendo, outliers degradam o desempenho dos métodos. Uma possível solução para este problema é o uso da técnica de construção de vizinhança corte- ϵ na qual os pontos são conectados a todos os demais pontos dentro de uma esfera de raio ϵ . Uma segunda alternativa para a solução da presença de outliers nos dados é seu pré-processamento.

Como já dito anteriormente, os métodos de redução esparsos são construídos sobre a hipótese de linearidade local de modo que a variedade seja suave ou diferenciável, não apresentando descontinuidades; porém em conjuntos de dados reais a presença de descontinuidades causadas por ruídos ou componentes desconectados é comum e pode depreciar o desempenho dos métodos.

Por fim, as principais fraquezas ou limitações dos métodos espectrais são: a restrição simples sobre sua solução de modo que frequentemente se obtêm soluções sub-ótimas, problemas e instabilidades numéricas provocadas com a resolução do problema de autovalores e autovetores, a maldição da dimensionalidade intrínseca que é a necessidade de um número muito grande de pontos para a descrição adequada da variedade, quando esta possui alta dimensionalidade e o alto custo computacional para a aplicação de métodos como o *Hessian Locally Linear Embedding* que calcula uma aproximação para a matriz hessiana e o *Isometric Mapping* que aproxima as distâncias geodésicas utilizando o algoritmo de caminhos mínimos sobre grafos.

Neste trabalho foram abordados métodos de redução de dimensionalidade convexas e espectrais. Para trabalhos futuros, o estudo poderá se estender a métodos de redução não

convexos e a extensões que sejam capazes de fazer o mapeamento de novos exemplos uma vez computado o *kernel* de redução. O problema é conhecido na literatura como *out-of-sample* (BENGIO et al., 2003).

Apêndices

Neste apêndice são exibidos os resultados dos experimentos obtidos sobre os conjuntos de dados naturais. São mostradas as tabelas para as métricas de confiabilidade, continuidade, erro de generalização e medida independente de escala para cada um dos métodos de redução, abordados durante a realização deste trabalho. Somente a melhor execução é reportada.

40	7,61%									
41	6,87%									
42	7,42%									
43	7,37%									
44	7,85%									
45	7,61%									
46	7,99%									
47	8,35%									
48	7,84%									
49	7,45%									

Tabela 12 – Erro de generalização para *Hessian Locally Linear Embedding*

.1.4 Medida independente de escala

Na tabela .1.4 é exibido os resultados da métrica independente de escala para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
2	0,098	0,174	0,011	0,177	0,338	0,010	0,212	0,835	0,946	0,187
3	0,168	0,157	0,011	0,190	0,355	0,010	0,231	0,323	0,940	0,340
4	0,262	0,216	0,014	0,488	0,390	0,010	0,296	0,329	0,937	0,456
5	0,381	0,220	0,011	0,501	0,493	0,010	0,345	0,322	0,935	0,580
6	0,446	0,308	0,011	0,630	0,552	0,010	0,429	0,364	0,923	0,545
7	0,460	0,384	0,011	0,687	0,635	0,010	0,501	0,355	0,856	0,514
8	0,502	0,427	0,012	0,709	0,657	0,010	0,518	0,307	0,840	0,513
9	0,525	0,426	0,011	0,729	0,677	0,010	0,529	0,223	0,830	0,416
10	0,532	0,433	0,011	0,722	0,702	0,010	0,601	0,199	0,657	0,353
11	0,558	0,480	0,011	0,716	0,714	0,146	0,645	0,143	0,637	0,330
12	0,577	0,493	0,011	0,700	0,724	0,360	0,673	0,145	0,653	0,368
13	0,591	0,492	0,017	0,692	0,714	0,524	0,691	0,138	0,654	0,347
14	0,605	0,513	0,066	0,669	0,693	0,561	0,694	0,124	0,650	0,324
15	0,621	0,535	0,163	0,655	0,680	0,517	0,700	0,115	0,651	0,305
16	0,627	0,548	0,256	0,638	0,640	0,630	0,706	0,107	0,646	0,248
17	0,640	0,557	0,280	0,638	0,638	0,682	0,710	0,086	0,643	0,240
18	0,650	0,568	0,331	0,634	0,640	0,712	0,722	0,113	0,641	0,263
19	0,659	0,570	0,368	0,625	0,622	0,714	0,722	0,105	0,639	0,257
20	0,667	0,571	0,259	0,624	0,626	0,604	0,720	0,099	0,636	0,219
21	0,670	0,584	0,296	0,617	0,583	0,587	0,714	0,137	0,635	0,208
22	0,679	0,587	0,303	0,612	0,587	0,586	0,713	0,109	0,633	0,211
23	0,681	0,607	0,318	0,601	0,576	0,585	0,713	0,098	0,632	0,195
24	0,688	0,607	0,342	0,605	0,569	0,584	0,713	0,107	0,630	0,216
25	0,692	0,616	0,343	0,608	0,554	0,545	0,710	0,089	0,628	0,220
26	0,673	0,610	0,355							
27	0,697	0,610	0,359							
28	0,700	0,624	0,378							
29	0,707	0,614	0,377							
30	0,710	0,617	0,384							
31	0,709	0,616	0,400							
32	0,710	0,612	0,376							
33	0,708	0,618	0,380							
34	0,710	0,619	0,378							
35	0,710	0,610	0,380							

36	0,710	0,611								
37	0,706	0,610								
38	0,706									
39	0,705									
40	0,703									
41	0,699									
42	0,697									
43	0,695									
44	0,695									
45	0,691									
46	0,689									
47	0,687									
48	0,685									
49	0,682									

Tabela 13 – Medida independente de escala para *Hessian Locally Linear Embedding*

.2 Método de redução *Isometric Mapping*

Nesta seção é exibido os resultados das métricas computadas sobre os conjuntos de dados naturais usando o método de redução *Isometric Mapping*.

.2.1 Medidas de confiabilidade

Na tabela .2.1 é exibido os resultados da métrica confiabilidade para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
2	0,752	0,879	0,866	0,875	0,849	0,855	0,832	0,993	0,995	0,954
3	0,860	0,946	0,920	0,950	0,948	0,919	0,931	0,995	0,995	0,984
4	0,925	0,970	0,945	0,987	0,976	0,933	0,971	0,996	0,995	0,989
5	0,962	0,983	0,955	0,993	0,987	0,963	0,985	0,996	0,995	0,996
6	0,980	0,987	0,962	0,995	0,994	0,982	0,990	0,997	0,995	0,997
7	0,987	0,989	0,970	0,996	0,995	0,985	0,994	0,997	0,995	0,997
8	0,989	0,991	0,972	0,996	0,996	0,989	0,996	0,997	0,995	0,998
9	0,993	0,992	0,975	0,996	0,996	0,991	0,997	0,997	0,995	0,998
10	0,994	0,993	0,976	0,997	0,997	0,994	0,998	0,998	0,996	0,998
11	0,995	0,993	0,977	0,997	0,997	0,996	0,998	0,998	0,996	0,998
12	0,995	0,994	0,979	0,997	0,998	0,996	0,998	0,999	0,996	0,998
13	0,996	0,994	0,979	0,997	0,998	0,997	0,998	0,999	0,996	0,998
14	0,996	0,994	0,981	0,998	0,998	0,997	0,998	0,999	0,996	0,998
15	0,997	0,994	0,982	0,998	0,998	0,997	0,998	0,999	0,996	0,998
16	0,997	0,995	0,983	0,998	0,998	0,998	0,999	0,999	0,996	0,998
17	0,997	0,995	0,984	0,998	0,998	0,998	0,999	0,999	0,996	0,998
18	0,998	0,995	0,985	0,998	0,998	0,998	0,999	0,999	0,996	0,998
19	0,998	0,995	0,985	0,998	0,998	0,998	0,999	0,999	0,996	0,998
20	0,998	0,996	0,986	0,998	0,998	0,998	0,999	0,999	0,996	0,998
21	0,998	0,996	0,986	0,998	0,998	0,998	0,999	0,999	0,996	0,998
22	0,998	0,996	0,987	0,998	0,998	0,998	0,999	0,999	0,996	0,998

23	0,998	0,996	0,987	0,998	0,998	0,998	0,999	0,999	0,996	0,998
24	0,998	0,996	0,988	0,998	0,998	0,998	0,999	0,999	0,996	0,998
25	0,998	0,996	0,988	0,998	0,998	0,998	0,999	0,999	0,996	0,998
26	0,999	0,996	0,989							
27	0,999	0,996	0,989							
28	0,999	0,996	0,989							
29	0,999	0,996	0,989							
30	0,999	0,996	0,990							
31	0,999	0,996	0,990							
32	0,999	0,996	0,990							
33	0,999	0,996	0,991							
34	0,999	0,996	0,991							
35	0,999	0,996	0,991							
36	0,999	0,996								
37	0,999	0,996								
38	0,999									
39	0,999									
40	0,999									
41	0,999									
42	0,999									
43	0,999									
44	0,999									
45	0,999									
46	0,999									
47	0,999									
48	0,999									
49	0,999									

Tabela 14 – Confiabilidade para *Isometric Mapping*

.2.2 Medida de continuidade

Na tabela .2.2 é exibido os resultados da métrica continuidade para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
2	0,952	0,971	0,980	0,977	0,973	0,980	0,975	0,997	0,996	0,990
3	0,979	0,985	0,986	0,990	0,990	0,993	0,990	0,998	0,996	0,996
4	0,989	0,990	0,989	0,995	0,994	0,995	0,995	0,998	0,996	0,996
5	0,993	0,992	0,992	0,996	0,995	0,997	0,996	0,998	0,996	0,998
6	0,995	0,993	0,993	0,997	0,997	0,998	0,997	0,998	0,996	0,998
7	0,996	0,994	0,994	0,997	0,997	0,999	0,998	0,998	0,996	0,998
8	0,997	0,995	0,994	0,997	0,997	0,999	0,998	0,998	0,996	0,998
9	0,997	0,995	0,995	0,997	0,998	0,999	0,998	0,998	0,996	0,998
10	0,998	0,995	0,995	0,997	0,998	0,999	0,998	0,999	0,997	0,998
11	0,998	0,995	0,995	0,997	0,998	0,999	0,999	0,999	0,997	0,998
12	0,998	0,996	0,996	0,997	0,998	0,999	0,999	0,999	0,997	0,998
13	0,998	0,996	0,996	0,998	0,998	0,999	0,999	0,999	0,997	0,998
14	0,998	0,996	0,996	0,998	0,998	0,999	0,999	0,999	0,997	0,998
15	0,998	0,996	0,996	0,998	0,998	0,999	0,999	0,999	0,997	0,998
16	0,998	0,996	0,996	0,998	0,998	0,999	0,999	0,999	0,997	0,998
17	0,999	0,996	0,996	0,998	0,998	0,999	0,999	0,999	0,997	0,998
18	0,999	0,996	0,996	0,998	0,998	0,999	0,999	0,999	0,997	0,998

19	0,999	0,996	0,996	0,998	0,998	0,999	0,999	0,999	0,997	0,998
20	0,999	0,997	0,996	0,998	0,998	0,999	0,999	0,999	0,997	0,998
21	0,999	0,997	0,996	0,998	0,998	0,999	0,999	0,999	0,997	0,998
22	0,999	0,997	0,997	0,998	0,999	0,999	0,999	0,999	0,997	0,998
23	0,999	0,997	0,997	0,998	0,999	0,999	0,999	0,999	0,997	0,998
24	0,999	0,997	0,997	0,998	0,998	0,999	0,999	0,999	0,997	0,998
25	0,999	0,997	0,997	0,998	0,998	0,999	0,999	0,999	0,997	0,998
26	0,999	0,997	0,997							
27	0,999	0,997	0,997							
28	0,999	0,997	0,997							
29	0,999	0,997	0,997							
30	0,999	0,997	0,997							
31	0,999	0,997	0,997							
32	0,999	0,997	0,997							
33	0,999	0,997	0,997							
34	0,999	0,997	0,997							
35	0,999	0,997	0,997							
36	0,999	0,997								
37	0,999	0,997								
38	0,999									
39	0,999									
40	0,999									
41	0,999									
42	0,999									
43	0,999									
44	0,999									
45	0,999									
46	0,999									
47	0,999									
48	0,999									
49	0,999									

Tabela 15 – Continuidade para *Isometric Mapping*

.2.3 Erro de generalização

Na tabela .2.3 é exibido os resultados da métrica erro de generalização para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
	6.78%	14.36%	12.08%	4.70%	19.68%	53.60%	3.27%	79.68%	42.63%	9.23%
2	61,62%	59,56%	20,81%	40,15%	49,85%	60,85%	43,79%	83,49%	52,84%	10,87%
3	53,33%	44,64%	17,77%	25,13%	37,58%	58,18%	25,68%	81,38%	52,55%	11,38%
4	36,34%	35,02%	15,68%	14,48%	29,65%	57,55%	12,50%	83,28%	53,58%	10,00%
5	25,48%	27,56%	15,53%	11,78%	27,68%	56,47%	8,31%	80,32%	53,50%	10,34%
6	17,54%	24,98%	14,88%	9,68%	24,75%	56,55%	8,05%	81,75%	53,74%	10,41%
7	14,21%	23,61%	14,33%	9,73%	22,50%	56,31%	6,71%	80,32%	53,00%	10,32%
8	12,72%	20,65%	14,21%	9,00%	22,45%	55,63%	6,25%	81,22%	54,61%	10,61%
9	10,02%	20,27%	14,59%	9,35%	21,50%	55,60%	5,82%	80,26%	53,42%	10,04%
10	10,24%	20,50%	15,44%	8,68%	21,13%	55,88%	5,68%	81,06%	49,51%	10,41%
11	9,65%	19,78%	15,03%	8,80%	21,18%	55,13%	5,85%	81,22%	49,01%	10,37%
12	9,23%	19,74%	14,55%	8,20%	21,45%	54,16%	5,43%	80,48%	49,88%	10,85%
13	8,64%	19,27%	14,61%	8,23%	20,85%	54,87%	4,84%	79,89%	49,59%	10,75%
14	8,58%	18,91%	14,40%	9,13%	21,18%	54,74%	4,82%	80,58%	50,12%	10,39%

15	8,61%	19,82%	14,83%	8,55%	21,48%	54,54%	5,12%	80,42%	49,63%	10,55%
16	8,15%	19,40%	14,62%	8,83%	21,50%	55,14%	5,18%	81,01%	50,41%	10,57%
17	8,11%	19,49%	13,92%	8,90%	20,85%	54,83%	5,14%	80,90%	50,04%	9,94%
18	7,85%	19,06%	14,42%	8,98%	20,28%	54,84%	5,48%	80,26%	50,78%	10,02%
19	8,16%	19,43%	14,40%	8,55%	20,90%	55,05%	5,16%	80,37%	50,12%	10,65%
20	8,47%	18,69%	13,62%	8,08%	21,08%	55,42%	5,76%	80,32%	48,89%	10,08%
21	8,05%	19,14%	14,00%	9,73%	21,25%	55,96%	5,14%	80,05%	50,41%	9,94%
22	7,82%	18,46%	14,02%	8,45%	20,90%	55,22%	5,30%	80,42%	49,96%	10,55%
23	7,42%	18,41%	13,91%	9,10%	21,08%	55,51%	5,32%	80,37%	51,56%	9,80%
24	7,42%	19,10%	13,54%	8,65%	20,85%	55,24%	4,95%	80,32%	48,81%	10,67%
25	7,86%	18,53%	14,04%	8,78%	20,88%	55,06%	5,42%	79,47%	50,12%	10,10%
26	7,60%	18,45%	14,02%							
27	8,19%	17,94%	13,84%							
28	7,59%	18,49%	13,79%							
29	7,75%	18,74%	13,54%							
30	7,69%	19,15%	13,80%							
31	7,41%	18,25%	13,43%							
32	7,82%	18,37%	13,13%							
33	7,53%	18,76%	13,27%							
34	7,53%	18,17%	13,06%							
35	7,61%	17,39%	13,63%							
36	7,63%	17,57%								
37	7,67%	17,23%								
38	7,81%									
39	7,78%									
40	7,82%									
41	7,78%									
42	7,09%									
43	7,92%									
44	7,91%									
45	7,66%									
46	7,66%									
47	7,68%									
48	7,94%									
49	7,70%									

Tabela 16 – Erro de generalização para *Isometric Mapping*

.2.4 Medida independente de escala

Na tabela .2.4 é exibido os resultados da métrica independente de escala para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
2	0,128	0,198	0,313	0,360	0,310	0,405	0,235	0,907	0,947	0,457
3	0,247	0,310	0,392	0,520	0,480	0,489	0,376	0,909	0,946	0,613
4	0,355	0,379	0,434	0,653	0,560	0,542	0,487	0,910	0,947	0,633
5	0,444	0,438	0,464	0,698	0,613	0,609	0,545	0,909	0,948	0,698
6	0,504	0,459	0,497	0,716	0,671	0,667	0,567	0,910	0,948	0,741
7	0,541	0,480	0,531	0,723	0,686	0,683	0,594	0,911	0,948	0,742
8	0,560	0,496	0,540	0,733	0,700	0,698	0,611	0,911	0,948	0,745
9	0,585	0,511	0,556	0,733	0,709	0,717	0,629	0,910	0,948	0,746
10	0,601	0,519	0,561	0,735	0,714	0,742	0,664	0,932	0,964	0,788
11	0,614	0,526	0,568	0,736	0,724	0,755	0,671	0,932	0,964	0,789

12	0,627	0,533	0,581	0,739	0,732	0,763	0,678	0,934	0,964	0,786
13	0,639	0,540	0,589	0,744	0,737	0,778	0,684	0,933	0,964	0,786
14	0,646	0,545	0,594	0,766	0,767	0,780	0,692	0,932	0,964	0,786
15	0,653	0,550	0,596	0,769	0,769	0,784	0,694	0,931	0,964	0,784
16	0,661	0,556	0,603	0,769	0,773	0,794	0,698	0,932	0,964	0,784
17	0,666	0,563	0,608	0,768	0,773	0,795	0,700	0,932	0,964	0,783
18	0,670	0,566	0,610	0,769	0,773	0,796	0,720	0,931	0,964	0,783
19	0,674	0,569	0,609	0,771	0,775	0,795	0,723	0,930	0,964	0,784
20	0,678	0,588	0,619	0,772	0,776	0,796	0,721	0,930	0,964	0,783
21	0,681	0,590	0,621	0,771	0,777	0,796	0,724	0,929	0,964	0,784
22	0,684	0,593	0,624	0,772	0,777	0,796	0,727	0,929	0,964	0,783
23	0,689	0,594	0,624	0,772	0,777	0,796	0,727	0,929	0,964	0,784
24	0,691	0,597	0,627	0,773	0,778	0,796	0,730	0,929	0,964	0,784
25	0,694	0,598	0,629	0,775	0,777	0,797	0,731	0,928	0,964	0,784
26	0,710	0,600	0,632							
27	0,711	0,601	0,635							
28	0,713	0,602	0,639							
29	0,715	0,603	0,643							
30	0,717	0,604	0,648							
31	0,719	0,607	0,648							
32	0,721	0,608	0,650							
33	0,722	0,609	0,652							
34	0,723	0,610	0,653							
35	0,724	0,610	0,652							
36	0,726	0,612								
37	0,726	0,612								
38	0,726									
39	0,728									
40	0,729									
41	0,730									
42	0,730									
43	0,731									
44	0,731									
45	0,732									
46	0,733									
47	0,734									
48	0,734									
49	0,735									

Tabela 17 – Medida independente de escala para *Isometric Mapping*

.3 Método de redução *Laplacian Eigenmaps*

Nesta seção é exibido os resultados das métricas computadas sobre os conjuntos de dados naturais usando o método de redução *Laplacian Eigenmaps*.

.3.1 Medida de confiabilidade

Na tabela .3.1 é exibido os resultados da métrica confiabilidade para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
2	0,772	0,885	0,825	0,868	0,849	0,748	0,836	0,978	0,987	0,850
3	0,868	0,928	0,899	0,921	0,914	0,827	0,892	0,982	0,991	0,955
4	0,912	0,941	0,926	0,945	0,951	0,920	0,927	0,985	0,992	0,959
5	0,925	0,950	0,941	0,964	0,955	0,941	0,934	0,987	0,993	0,963
6	0,944	0,952	0,947	0,973	0,961	0,952	0,942	0,988	0,994	0,967
7	0,948	0,958	0,951	0,972	0,968	0,955	0,945	0,988	0,994	0,970
8	0,951	0,959	0,947	0,975	0,970	0,963	0,948	0,989	0,994	0,971
9	0,954	0,962	0,951	0,977	0,973	0,969	0,951	0,989	0,994	0,975
10	0,954	0,966	0,955	0,978	0,976	0,971	0,958	0,990	0,996	0,976
11	0,955	0,967	0,957	0,978	0,978	0,974	0,960	0,990	0,996	0,980
12	0,955	0,969	0,956	0,979	0,978	0,975	0,964	0,990	0,996	0,980
13	0,956	0,971	0,957	0,979	0,978	0,976	0,965	0,990	0,996	0,984
14	0,959	0,971	0,958	0,979	0,982	0,977	0,967	0,990	0,996	0,984
15	0,960	0,971	0,959	0,980	0,983	0,978	0,968	0,991	0,996	0,986
16	0,961	0,971	0,959	0,981	0,985	0,979	0,969	0,991	0,996	0,986
17	0,962	0,971	0,961	0,981	0,986	0,978	0,971	0,991	0,996	0,987
18	0,964	0,972	0,961	0,982	0,986	0,979	0,973	0,992	0,996	0,989
19	0,964	0,973	0,952	0,983	0,986	0,981	0,973	0,992	0,996	0,989
20	0,964	0,974	0,952	0,984	0,986	0,983	0,975	0,992	0,996	0,989
21	0,964	0,975	0,952	0,986	0,986	0,983	0,975	0,992	0,996	0,990
22	0,965	0,976	0,952	0,986	0,986	0,983	0,977	0,992	0,996	0,990
23	0,964	0,976	0,953	0,987	0,986	0,983	0,979	0,993	0,996	0,991
24	0,965	0,976	0,954	0,987	0,987	0,984	0,979	0,993	0,996	0,992
25	0,968	0,976	0,955	0,987	0,987	0,985	0,979	0,993	0,996	0,992
26	0,972	0,977	0,955							
27	0,973	0,978	0,956							
28	0,974	0,977	0,956							
29	0,975	0,978	0,956							
30	0,976	0,979	0,956							
31	0,976	0,979	0,955							
32	0,976	0,980	0,955							
33	0,977	0,981	0,955							
34	0,977	0,981	0,955							
35	0,977	0,982	0,955							
36	0,977	0,982								
37	0,978	0,982								
38	0,980									
39	0,980									
40	0,980									
41	0,981									
42	0,982									
43	0,982									
44	0,983									
45	0,983									
46	0,983									
47	0,983									
48	0,983									
49	0,983									

Tabela 18 – Confiabilidade para *Laplacian Eigenmaps*

.3.2 Medida de continuidade

Na tabela .3.2 é exibido os resultados da métrica continuidade para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
2	0,922	0,844	0,921	0,919	0,932	0,959	0,923	0,976	0,957	0,944
3	0,933	0,893	0,931	0,951	0,950	0,963	0,938	0,984	0,972	0,969
4	0,943	0,899	0,936	0,959	0,960	0,969	0,940	0,988	0,979	0,970
5	0,951	0,900	0,944	0,963	0,961	0,973	0,956	0,990	0,983	0,974
6	0,954	0,908	0,953	0,966	0,965	0,974	0,959	0,993	0,984	0,981
7	0,961	0,907	0,954	0,967	0,969	0,974	0,960	0,993	0,987	0,981
8	0,963	0,907	0,972	0,970	0,970	0,979	0,963	0,994	0,988	0,982
9	0,965	0,920	0,971	0,972	0,973	0,981	0,968	0,994	0,990	0,983
10	0,965	0,923	0,971	0,973	0,974	0,982	0,972	0,996	0,997	0,989
11	0,965	0,925	0,972	0,973	0,975	0,988	0,973	0,996	0,997	0,989
12	0,966	0,925	0,975	0,973	0,974	0,988	0,977	0,996	0,997	0,989
13	0,965	0,934	0,977	0,975	0,974	0,989	0,977	0,996	0,997	0,990
14	0,967	0,933	0,977	0,980	0,980	0,990	0,977	0,996	0,997	0,991
15	0,968	0,934	0,977	0,981	0,981	0,990	0,977	0,996	0,997	0,992
16	0,969	0,937	0,979	0,981	0,983	0,990	0,978	0,996	0,997	0,992
17	0,968	0,937	0,981	0,983	0,984	0,990	0,979	0,996	0,997	0,992
18	0,968	0,937	0,980	0,984	0,984	0,991	0,982	0,996	0,997	0,993
19	0,968	0,937	0,982	0,984	0,985	0,992	0,982	0,996	0,997	0,993
20	0,968	0,954	0,983	0,984	0,985	0,993	0,982	0,996	0,997	0,993
21	0,968	0,955	0,983	0,986	0,985	0,994	0,982	0,996	0,997	0,993
22	0,968	0,955	0,983	0,986	0,985	0,994	0,983	0,996	0,997	0,994
23	0,969	0,955	0,983	0,987	0,985	0,994	0,983	0,996	0,997	0,994
24	0,969	0,955	0,983	0,987	0,985	0,994	0,983	0,996	0,997	0,994
25	0,970	0,955	0,984	0,987	0,985	0,994	0,983	0,997	0,997	0,994
26	0,976	0,956	0,984							
27	0,976	0,957	0,983							
28	0,977	0,957	0,984							
29	0,976	0,958	0,984							
30	0,977	0,958	0,985							
31	0,977	0,957	0,985							
32	0,976	0,957	0,985							
33	0,977	0,958	0,985							
34	0,977	0,958	0,986							
35	0,977	0,958	0,986							
36	0,977	0,957								
37	0,978	0,957								
38	0,979									
39	0,979									
40	0,980									
41	0,980									
42	0,981									
43	0,981									
44	0,981									
45	0,982									
46	0,982									
47	0,982									
48	0,982									
49	0,982									

Tabela 19 – Continuidade para Laplacian Eigenmaps

46	0,451									
47	0,454									
48	0,455									
49	0,456									

Tabela 21 – Medida independente de escala para *Laplacian Eigenmaps*

.4 Método de redução *Locally Linear Embedding*

Nesta seção é exibido os resultados das métricas computadas sobre os conjuntos de dados naturais usando o método de redução *Locally Linear Embedding*.

.4.1 Medida de confiabilidade

Na tabela .4.1 é exibido os resultados da métrica confiabilidade para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
2	0,688	0,924	0,922	0,731	0,712	0,822	0,764	0,979	0,996	0,761
3	0,804	0,941	0,928	0,871	0,777	0,894	0,845	0,987	0,996	0,866
4	0,840	0,954	0,934	0,939	0,875	0,926	0,876	0,989	0,996	0,917
5	0,874	0,967	0,937	0,968	0,911	0,944	0,912	0,987	0,996	0,957
6	0,899	0,973	0,938	0,978	0,935	0,952	0,935	0,986	0,996	0,974
7	0,912	0,974	0,940	0,981	0,949	0,958	0,942	0,987	0,996	0,977
8	0,930	0,978	0,939	0,982	0,965	0,962	0,948	0,987	0,996	0,979
9	0,937	0,981	0,940	0,986	0,972	0,966	0,954	0,986	0,996	0,981
10	0,949	0,983	0,941	0,988	0,977	0,967	0,968	0,986	0,996	0,987
11	0,954	0,984	0,940	0,989	0,981	0,981	0,971	0,988	0,996	0,986
12	0,957	0,984	0,950	0,989	0,984	0,982	0,973	0,989	0,995	0,986
13	0,961	0,984	0,958	0,989	0,986	0,983	0,975	0,990	0,994	0,986
14	0,965	0,984	0,966	0,993	0,991	0,984	0,975	0,990	0,994	0,986
15	0,966	0,986	0,968	0,992	0,992	0,986	0,978	0,991	0,993	0,986
16	0,968	0,986	0,970	0,992	0,993	0,986	0,978	0,992	0,993	0,985
17	0,970	0,986	0,974	0,993	0,994	0,986	0,980	0,992	0,993	0,986
18	0,973	0,986	0,975	0,993	0,994	0,988	0,985	0,993	0,992	0,985
19	0,974	0,986	0,981	0,993	0,994	0,989	0,986	0,993	0,992	0,985
20	0,977	0,990	0,981	0,993	0,995	0,991	0,986	0,993	0,991	0,986
21	0,978	0,991	0,979	0,993	0,995	0,991	0,986	0,994	0,990	0,986
22	0,978	0,991	0,980	0,994	0,995	0,990	0,987	0,994	0,990	0,986
23	0,978	0,991	0,983	0,994	0,995	0,990	0,988	0,994	0,990	0,987
24	0,979	0,992	0,984	0,994	0,995	0,990	0,988	0,994	0,990	0,987
25	0,980	0,992	0,984	0,994	0,995	0,990	0,989	0,995	0,989	0,986
26	0,976	0,993	0,985							
27	0,977	0,993	0,985							
28	0,978	0,993	0,986							
29	0,979	0,993	0,987							
30	0,979	0,994	0,987							
31	0,980	0,994	0,987							
32	0,981	0,994	0,987							
33	0,980	0,994	0,987							
34	0,981	0,994	0,987							

35	0,981	0,994	0,988							
36	0,981	0,994								
37	0,981	0,994								
38	0,980									
39	0,981									
40	0,981									
41	0,981									
42	0,981									
43	0,981									
44	0,981									
45	0,980									
46	0,981									
47	0,982									
48	0,982									
49	0,982									

Tabela 22 – Confiabilidade para *Locally Linear Embedding*

.4.2 Medida de continuidade

Na tabela .4.2 é exibido os resultados da métrica continuidade para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
2	0,886	0,969	0,972	0,915	0,904	0,946	0,918	0,984	0,997	0,885
3	0,930	0,975	0,974	0,950	0,932	0,960	0,946	0,991	0,997	0,938
4	0,935	0,977	0,975	0,967	0,950	0,969	0,948	0,992	0,997	0,947
5	0,941	0,984	0,976	0,977	0,963	0,973	0,952	0,991	0,997	0,961
6	0,952	0,985	0,977	0,981	0,972	0,974	0,963	0,992	0,998	0,954
7	0,951	0,984	0,977	0,981	0,971	0,974	0,956	0,991	0,998	0,955
8	0,956	0,986	0,977	0,982	0,975	0,977	0,954	0,989	0,998	0,956
9	0,955	0,990	0,978	0,980	0,978	0,978	0,955	0,988	0,998	0,958
10	0,963	0,991	0,978	0,981	0,978	0,977	0,960	0,986	0,996	0,972
11	0,964	0,991	0,975	0,983	0,979	0,985	0,959	0,987	0,995	0,971
12	0,967	0,991	0,979	0,982	0,979	0,986	0,959	0,986	0,993	0,972
13	0,967	0,990	0,983	0,981	0,981	0,986	0,960	0,987	0,991	0,970
14	0,967	0,990	0,985	0,989	0,988	0,986	0,961	0,988	0,989	0,968
15	0,967	0,991	0,984	0,989	0,988	0,987	0,962	0,987	0,987	0,966
16	0,966	0,991	0,984	0,989	0,988	0,988	0,962	0,989	0,985	0,964
17	0,966	0,990	0,986	0,989	0,988	0,988	0,963	0,990	0,986	0,964
18	0,968	0,990	0,987	0,989	0,989	0,988	0,970	0,990	0,985	0,963
19	0,968	0,991	0,992	0,988	0,989	0,988	0,972	0,991	0,984	0,962
20	0,969	0,992	0,990	0,989	0,989	0,992	0,973	0,991	0,984	0,962
21	0,969	0,993	0,991	0,989	0,989	0,991	0,974	0,992	0,983	0,961
22	0,969	0,993	0,991	0,990	0,989	0,991	0,975	0,992	0,983	0,960
23	0,966	0,994	0,992	0,990	0,989	0,990	0,976	0,992	0,983	0,961
24	0,964	0,994	0,993	0,990	0,989	0,990	0,977	0,992	0,982	0,961
25	0,964	0,994	0,993	0,990	0,989	0,989	0,976	0,993	0,982	0,960
26	0,957	0,994	0,993							
27	0,957	0,994	0,994							
28	0,958	0,994	0,994							
29	0,959	0,994	0,994							
30	0,958	0,994	0,994							
31	0,959	0,994	0,994							

32	0,959	0,994	0,994							
33	0,959	0,993	0,994							
34	0,959	0,993	0,994							
35	0,958	0,993	0,994							
36	0,958	0,993								
37	0,957	0,993								
38	0,957									
39	0,957									
40	0,957									
41	0,957									
42	0,956									
43	0,956									
44	0,955									
45	0,955									
46	0,956									
47	0,956									
48	0,956									
49	0,956									

Tabela 23 – Continuidade para *Locally Linear Embedding*

.4.3 Erro de generalização

Na tabela .4.3 é exibido os resultados da métrica erro de generalização para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
	6.78%	14.36%	12.08%	4.70%	19.68%	53.60%	3.27%	79.68%	42.63%	9.23%
2	70,02%	45,09%	22,65%	60,83%	74,68%	66,66%	52,01%	83,54%	50,74%	11,18%
3	51,00%	36,65%	21,96%	40,05%	64,35%	61,37%	37,67%	83,23%	50,58%	10,45%
4	44,46%	33,63%	22,07%	26,45%	53,40%	59,18%	29,25%	84,39%	50,16%	11,16%
5	37,87%	30,48%	22,32%	21,05%	47,13%	57,59%	21,54%	82,43%	51,07%	10,73%
6	32,61%	28,31%	22,57%	16,00%	40,33%	57,34%	17,58%	83,02%	51,56%	9,59%
7	30,78%	29,27%	23,36%	14,28%	37,98%	57,06%	17,02%	82,75%	51,44%	9,98%
8	27,54%	25,16%	22,64%	13,40%	33,35%	55,66%	15,01%	84,02%	50,41%	9,49%
9	26,45%	26,09%	22,55%	14,33%	30,45%	56,55%	14,00%	83,02%	49,18%	10,22%
10	23,47%	25,42%	23,24%	11,25%	27,88%	56,15%	12,86%	80,58%	48,31%	9,17%
11	21,96%	24,83%	22,33%	9,77%	25,83%	54,84%	12,34%	81,43%	46,54%	9,09%
12	21,58%	24,82%	21,11%	10,20%	25,33%	55,86%	11,56%	83,23%	45,76%	8,76%
13	19,62%	25,37%	19,34%	10,23%	25,88%	55,45%	11,14%	81,90%	41,32%	9,45%
14	18,34%	24,29%	17,95%	7,30%	23,15%	55,20%	11,35%	82,33%	42,26%	8,76%
15	17,89%	22,93%	18,44%	7,08%	22,73%	56,10%	10,64%	82,75%	40,70%	8,76%
16	17,75%	23,36%	17,85%	5,43%	21,98%	55,64%	10,05%	82,49%	41,56%	9,31%
17	17,28%	23,04%	16,78%	6,60%	21,25%	56,15%	9,63%	83,33%	38,02%	8,62%
18	16,38%	22,50%	16,47%	6,23%	21,98%	56,07%	10,06%	82,80%	38,85%	9,27%
19	15,62%	23,23%	15,37%	6,80%	22,25%	56,24%	9,70%	82,33%	36,38%	9,63%
20	14,87%	21,25%	15,78%	6,25%	21,45%	54,78%	8,63%	81,38%	35,19%	9,37%
21	14,28%	20,13%	15,61%	6,88%	21,45%	54,10%	9,23%	81,01%	31,32%	8,88%
22	13,67%	20,95%	15,30%	7,25%	21,18%	54,15%	9,35%	82,06%	31,11%	9,25%
23	14,00%	20,07%	15,60%	6,45%	20,65%	54,73%	8,74%	82,28%	30,25%	8,74%
24	13,55%	20,10%	14,62%	6,98%	20,30%	54,61%	8,37%	82,49%	29,88%	9,59%
25	13,19%	19,78%	14,91%	6,78%	21,38%	53,78%	8,53%	82,91%	30,91%	8,97%
26	14,23%	19,55%	15,10%							
27	14,29%	19,85%	14,65%							

28	13,85%	19,42%	13,72%							
29	13,91%	19,70%	14,64%							
30	13,43%	18,76%	14,45%							
31	13,31%	18,49%	14,78%							
32	12,56%	18,20%	14,13%							
33	12,34%	19,05%	14,20%							
34	12,32%	18,84%	14,36%							
35	12,29%	19,05%	13,78%							
36	11,81%	19,01%								
37	11,81%	19,31%								
38	12,66%									
39	11,89%									
40	12,31%									
41	11,62%									
42	11,22%									
43	10,81%									
44	10,97%									
45	10,77%									
46	11,64%									
47	10,98%									
48	11,00%									
49	10,88%									

Tabela 24 – Erro de generalização para *Locally Linear Embedding*

.4.4 Medida independente de escala

Na tabela .4.4 é exibido os resultados da métrica independente de escala para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
2	0,131	0,234	0,337	0,185	0,153	0,290	0,129	0,588	0,936	0,158
3	0,212	0,266	0,368	0,313	0,220	0,397	0,206	0,617	0,936	0,237
4	0,235	0,279	0,370	0,380	0,283	0,455	0,236	0,627	0,934	0,287
5	0,253	0,320	0,368	0,436	0,328	0,493	0,267	0,568	0,933	0,335
6	0,270	0,331	0,370	0,493	0,369	0,497	0,286	0,561	0,932	0,330
7	0,279	0,332	0,369	0,500	0,384	0,525	0,290	0,541	0,930	0,334
8	0,305	0,352	0,375	0,493	0,415	0,543	0,297	0,527	0,929	0,339
9	0,311	0,379	0,373	0,504	0,436	0,554	0,308	0,508	0,928	0,349
10	0,322	0,396	0,375	0,517	0,448	0,553	0,322	0,487	0,876	0,424
11	0,327	0,399	0,369	0,530	0,461	0,597	0,326	0,512	0,841	0,416
12	0,330	0,401	0,412	0,529	0,470	0,607	0,327	0,517	0,795	0,415
13	0,333	0,388	0,433	0,528	0,480	0,618	0,328	0,533	0,743	0,414
14	0,335	0,393	0,453	0,557	0,505	0,623	0,331	0,548	0,709	0,408
15	0,337	0,402	0,449	0,559	0,508	0,622	0,332	0,547	0,670	0,411
16	0,337	0,402	0,434	0,558	0,513	0,624	0,328	0,573	0,657	0,404
17	0,338	0,403	0,453	0,555	0,520	0,622	0,334	0,589	0,656	0,406
18	0,344	0,403	0,458	0,557	0,527	0,631	0,365	0,589	0,649	0,401
19	0,349	0,408	0,494	0,552	0,527	0,632	0,369	0,595	0,631	0,401
20	0,357	0,433	0,483	0,560	0,529	0,651	0,373	0,607	0,621	0,396
21	0,359	0,443	0,476	0,557	0,530	0,644	0,373	0,616	0,582	0,394
22	0,360	0,450	0,481	0,561	0,534	0,638	0,373	0,618	0,576	0,391
23	0,356	0,462	0,491	0,559	0,536	0,630	0,377	0,608	0,581	0,394
24	0,354	0,464	0,505	0,559	0,535	0,619	0,375	0,621	0,572	0,393

25	0,352	0,466	0,507	0,558	0,533	0,613	0,378	0,631	0,566	0,391
26	0,330	0,470	0,519							
27	0,332	0,471	0,526							
28	0,333	0,481	0,525							
29	0,335	0,483	0,530							
30	0,335	0,484	0,532							
31	0,337	0,487	0,534							
32	0,338	0,489	0,532							
33	0,337	0,487	0,539							
34	0,336	0,487	0,543							
35	0,336	0,487	0,548							
36	0,334	0,489								
37	0,333	0,490								
38	0,332									
39	0,331									
40	0,331									
41	0,331									
42	0,331									
43	0,330									
44	0,329									
45	0,328									
46	0,329									
47	0,330									
48	0,330									
49	0,329									

Tabela 25 – Medida independente de escala para *Locally Linear Embedding*

.5 Método de redução *Local Tangent Space Alignment*

Nesta seção é exibido os resultados das métricas computadas sobre os conjuntos de dados naturais usando o método de redução *Local Tangent Space Alignment*.

.5.1 Medida de confiabilidade

Na tabela .5.2 é exibido os resultados da métrica confiabilidade para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
2	0,525	0,614	0,861	0,416	0,682	0,447	0,557	0,344	0,550	0,548
3	0,525	0,614	0,909	0,416	0,761	0,448	0,557	0,344	0,995	0,548
4	0,525	0,615	0,922	0,416	0,865	0,395	0,557	0,975	0,995	0,548
5	0,651	0,614	0,582	0,416	0,914	0,447	0,557	0,985	0,995	0,836
6	0,782	0,614	0,559	0,768	0,929	0,661	0,557	0,960	0,995	0,978
7	0,857	0,614	0,542	0,884	0,972	0,361	0,557	0,910	0,995	0,986
8	0,919	0,606	0,659	0,948	0,981	0,513	0,557	0,819	0,995	0,987
9	0,953	0,614	0,684	0,985	0,991	0,448	0,557	0,824	0,995	0,992
10	0,972	0,614	0,745	0,992	0,994	0,744	0,892	0,858	0,995	0,998
11	0,971	0,614	0,714	0,995	0,996	0,552	0,902	0,835	0,996	0,997
12	0,973	0,615	0,692	0,997	0,997	0,384	0,916	0,824	0,995	0,995

13	0,978	0,615	0,676	0,997	0,998	0,712	0,923	0,819	0,995	0,994
14	0,980	0,615	0,686	0,997	0,999	0,724	0,956	0,818	0,995	0,992
15	0,980	0,615	0,660	0,998	0,999	0,880	0,985	0,817	0,994	0,991
16	0,981	0,615	0,648	0,997	0,998	0,962	0,988	0,769	0,995	0,991
17	0,985	0,615	0,663	0,997	0,997	0,835	0,990	0,771	0,995	0,990
18	0,985	0,615	0,667	0,998	0,997	0,964	0,998	0,736	0,995	0,986
19	0,986	0,615	0,733	0,998	0,997	0,992	0,998	0,741	0,994	0,983
20	0,986	0,990	0,743	0,997	0,997	0,972	0,998	0,722	0,994	0,979
21	0,988	0,991	0,723	0,997	0,997	0,987	0,999	0,691	0,993	0,976
22	0,991	0,991	0,711	0,997	0,996	0,981	0,999	0,671	0,992	0,972
23	0,989	0,991	0,711	0,997	0,996	0,981	0,999	0,664	0,991	0,972
24	0,987	0,991	0,717	0,997	0,995	0,982	0,999	0,726	0,990	0,973
25	0,988	0,991	0,715	0,997	0,993	0,981	0,999	0,718	0,989	0,970
26	0,997	0,992	0,714							
27	0,997	0,992	0,709							
28	0,997	0,993	0,719							
29	0,997	0,994	0,741							
30	0,997	0,994	0,730							
31	0,998	0,994	0,737							
32	0,998	0,994	0,750							
33	0,998	0,994	0,757							
34	0,998	0,995	0,770							
35	0,998	0,995	0,768							
36	0,998	0,995								
37	0,998	0,995								
38	0,998									
39	0,998									
40	0,998									
41	0,999									
42	0,999									
43	0,999									
44	0,999									
45	0,999									
46	0,999									
47	0,998									
48	0,998									
49	0,998									

Tabela 26 – Confiabilidade para *Local Tangent Space Alignment*

.5.2 Medida de continuidade

Na tabela .5.2 é exibido os resultados da métrica continuidade para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
2	0,506	0,501	0,946	0,501	0,500	0,507	0,502	0,506	0,494	0,508
3	0,506	0,501	0,952	0,501	0,937	0,507	0,502	0,506	0,997	0,508
4	0,506	0,501	0,952	0,501	0,962	0,508	0,502	0,990	0,997	0,508
5	0,905	0,501	0,860	0,501	0,968	0,507	0,502	0,981	0,997	0,954
6	0,954	0,501	0,820	0,934	0,979	0,508	0,502	0,965	0,997	0,991
7	0,962	0,501	0,800	0,977	0,991	0,508	0,502	0,921	0,997	0,995
8	0,981	0,552	0,829	0,990	0,994	0,511	0,502	0,802	0,997	0,995
9	0,982	0,501	0,828	0,994	0,996	0,508	0,502	0,803	0,997	0,996

10	0,988	0,501	0,946	0,996	0,997	0,610	0,980	0,821	0,994	0,998
11	0,989	0,501	0,840	0,997	0,998	0,506	0,980	0,806	0,994	0,996
12	0,989	0,501	0,812	0,997	0,998	0,515	0,982	0,791	0,993	0,995
13	0,990	0,501	0,793	0,997	0,998	0,523	0,984	0,780	0,993	0,992
14	0,990	0,501	0,768	0,997	0,999	0,845	0,991	0,778	0,991	0,990
15	0,991	0,501	0,747	0,997	0,999	0,907	0,995	0,776	0,990	0,988
16	0,991	0,501	0,738	0,997	0,998	0,979	0,995	0,715	0,991	0,988
17	0,992	0,501	0,731	0,997	0,998	0,957	0,995	0,712	0,988	0,986
18	0,992	0,501	0,725	0,997	0,997	0,990	0,998	0,673	0,988	0,980
19	0,992	0,501	0,869	0,997	0,997	0,992	0,999	0,673	0,986	0,974
20	0,992	0,992	0,872	0,997	0,997	0,993	0,999	0,655	0,985	0,966
21	0,993	0,993	0,850	0,997	0,997	0,987	0,999	0,638	0,982	0,960
22	0,994	0,993	0,842	0,996	0,996	0,977	0,999	0,626	0,980	0,954
23	0,993	0,993	0,830	0,996	0,995	0,975	0,998	0,628	0,979	0,952
24	0,992	0,993	0,821	0,996	0,993	0,974	0,999	0,667	0,977	0,955
25	0,992	0,993	0,813	0,996	0,991	0,970	0,999	0,661	0,975	0,949
26	0,998	0,993	0,804							
27	0,998	0,993	0,795							
28	0,998	0,994	0,788							
29	0,998	0,994	0,788							
30	0,998	0,994	0,790							
31	0,998	0,994	0,790							
32	0,998	0,994	0,799							
33	0,998	0,994	0,810							
34	0,998	0,994	0,813							
35	0,998	0,994	0,815							
36	0,998	0,994								
37	0,998	0,994								
38	0,998									
39	0,998									
40	0,998									
41	0,998									
42	0,998									
43	0,998									
44	0,998									
45	0,998									
46	0,998									
47	0,998									
48	0,998									
49	0,998									

Tabela 27 – Continuidade para *Local Tangent Space Alignment*

.5.3 Erro de generalização

Na tabela .5.3 é exibido os resultados da métrica erro de generalização para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
	6.78%	14.36%	12.08%	4.70%	19.68%	53.60%	3.27%	79.68%	42.63%	9.23%
2	90,07%	96,14%	23,05%	90,05%	90,00%	80,01%	89,73%	86,03%	50,04%	6,31%
3	90,28%	96,11%	18,62%	90,00%	62,15%	81,40%	89,44%	85,66%	49,88%	15,05%
4	90,01%	96,17%	17,93%	90,00%	47,48%	81,17%	89,75%	83,44%	51,98%	15,03%
5	66,58%	96,18%	36,22%	89,98%	42,13%	81,59%	90,55%	83,28%	48,85%	11,26%

6	52,02%	96,13%	38,68%	66,70%	41,35%	85,55%	90,18%	81,69%	50,74%	11,54%
7	41,62%	96,12%	38,46%	51,95%	35,25%	82,36%	88,08%	80,37%	50,16%	10,85%
8	35,92%	95,90%	29,63%	33,70%	33,63%	80,74%	90,18%	79,79%	52,22%	9,76%
9	28,53%	96,18%	25,86%	21,53%	28,10%	81,10%	88,83%	80,26%	53,00%	10,69%
10	20,71%	96,15%	22,09%	18,33%	25,90%	79,98%	39,70%	80,32%	44,36%	9,82%
11	21,85%	96,17%	25,73%	13,60%	25,60%	81,67%	37,38%	79,37%	44,57%	8,74%
12	21,54%	96,14%	26,63%	11,78%	24,43%	82,56%	34,38%	79,26%	39,92%	9,15%
13	19,22%	96,18%	27,19%	9,90%	22,33%	79,31%	33,16%	78,04%	41,23%	9,25%
14	18,70%	96,15%	26,94%	8,65%	20,70%	79,46%	20,83%	79,10%	41,19%	9,84%
15	17,67%	96,17%	28,76%	8,13%	20,20%	75,30%	13,02%	79,63%	42,22%	9,21%
16	17,79%	96,14%	30,15%	8,68%	20,28%	61,45%	12,47%	79,89%	39,47%	8,92%
17	16,01%	96,17%	29,08%	7,48%	20,15%	61,99%	10,75%	79,95%	40,00%	9,25%
18	16,89%	96,14%	30,32%	7,25%	21,00%	54,36%	6,35%	79,10%	39,05%	9,01%
19	15,36%	96,18%	20,16%	5,05%	20,60%	56,77%	5,59%	81,43%	38,93%	8,95%
20	16,47%	23,28%	16,86%	5,43%	20,03%	53,87%	5,15%	77,78%	40,29%	8,92%
21	15,21%	22,95%	19,57%	5,40%	19,83%	56,52%	5,17%	79,26%	40,00%	8,60%
22	11,99%	22,74%	19,22%	5,63%	19,90%	56,32%	5,49%	77,83%	32,26%	8,52%
23	15,04%	22,59%	21,42%	4,95%	19,13%	57,10%	5,49%	78,68%	27,70%	8,78%
24	17,23%	22,83%	22,08%	4,45%	19,05%	58,05%	5,23%	80,21%	28,15%	8,68%
25	15,51%	21,32%	24,02%	4,88%	19,45%	57,69%	5,12%	78,41%	26,26%	8,99%
26	9,18%	20,27%	23,88%							
27	8,98%	20,34%	25,16%							
28	8,69%	18,14%	24,83%							
29	8,09%	18,25%	25,60%							
30	8,47%	17,93%	26,74%							
31	8,53%	17,61%	25,73%							
32	8,33%	17,87%	26,84%							
33	7,70%	17,87%	26,67%							
34	7,45%	18,42%	26,95%							
35	7,75%	17,22%	27,15%							
36	7,33%	17,52%								
37	7,04%	17,66%								
38	7,85%									
39	7,49%									
40	7,16%									
41	6,72%									
42	6,61%									
43	6,67%									
44	6,55%									
45	6,89%									
46	6,72%									
47	6,36%									
48	6,66%									
49	6,74%									

Tabela 28 – Erro de generalização para *Local Tangent Space Alignment*

.5.4 Medida independente de escala

Na tabela .5.4 é exibido os resultados da métrica independente de escala para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
---	-------	--------	-------	---------	---------	-----	------	--------	------	-------

2	0,011	0,010	0,317	0,024	0,028	0,011	0,010	0,056	0,041	0,017
3	0,011	0,010	0,333	0,024	0,208	0,011	0,010	0,056	0,930	0,017
4	0,011	0,010	0,375	0,024	0,304	0,011	0,010	0,823	0,934	0,017
5	0,073	0,010	0,071	0,024	0,336	0,011	0,010	0,469	0,936	0,208
6	0,164	0,010	0,054	0,173	0,376	0,011	0,010	0,319	0,934	0,496
7	0,208	0,010	0,045	0,338	0,510	0,011	0,010	0,204	0,932	0,591
8	0,317	0,012	0,106	0,489	0,552	0,011	0,010	0,146	0,930	0,587
9	0,325	0,010	0,125	0,623	0,629	0,011	0,010	0,149	0,929	0,593
10	0,410	0,010	0,210	0,682	0,655	0,017	0,268	0,175	0,863	0,668
11	0,403	0,010	0,148	0,723	0,684	0,010	0,273	0,156	0,867	0,612
12	0,406	0,010	0,129	0,739	0,699	0,012	0,292	0,154	0,860	0,562
13	0,423	0,010	0,125	0,746	0,709	0,015	0,330	0,153	0,865	0,524
14	0,438	0,010	0,133	0,726	0,733	0,119	0,424	0,154	0,858	0,499
15	0,448	0,010	0,126	0,723	0,716	0,230	0,503	0,155	0,853	0,489
16	0,451	0,010	0,123	0,703	0,691	0,476	0,512	0,130	0,854	0,489
17	0,473	0,010	0,137	0,704	0,657	0,212	0,532	0,129	0,852	0,475
18	0,474	0,010	0,140	0,705	0,643	0,573	0,679	0,113	0,849	0,432
19	0,477	0,010	0,176	0,698	0,645	0,672	0,683	0,114	0,839	0,406
20	0,476	0,456	0,198	0,687	0,641	0,627	0,687	0,107	0,820	0,378
21	0,488	0,467	0,178	0,683	0,631	0,599	0,690	0,099	0,791	0,359
22	0,508	0,468	0,169	0,675	0,618	0,537	0,695	0,093	0,758	0,340
23	0,493	0,465	0,170	0,665	0,602	0,533	0,696	0,095	0,726	0,341
24	0,486	0,466	0,173	0,663	0,583	0,536	0,704	0,139	0,691	0,349
25	0,490	0,474	0,180	0,660	0,556	0,523	0,705	0,136	0,662	0,337
26	0,603	0,488	0,185							
27	0,606	0,496	0,182							
28	0,618	0,514	0,186							
29	0,627	0,523	0,203							
30	0,636	0,526	0,201							
31	0,640	0,528	0,207							
32	0,641	0,530	0,216							
33	0,645	0,532	0,226							
34	0,655	0,537	0,232							
35	0,656	0,541	0,236							
36	0,658	0,548								
37	0,661	0,553								
38	0,660									
39	0,666									
40	0,672									
41	0,680									
42	0,684									
43	0,688									
44	0,689									
45	0,689									
46	0,690									
47	0,691									
48	0,693									
49	0,693									

Tabela 29 – Medida independente de escala para *Local Tangent Space Alignment*

37	0,320	0,423								
38	0,322									
39	0,323									
40	0,333									
41	0,336									
42	0,337									
43	0,340									
44	0,343									
45	0,346									
46	0,343									
47	0,345									
48	0,347									
49	0,333									

Tabela 33 – Medida independente de escala para *Diffusion Maps*

.7 Método de redução *Maximum Variance Unfolding*

Nesta seção é exibido os resultados das métricas computadas sobre os conjuntos de dados naturais usando o método de redução *Maximum Variance Unfolding*.

.7.1 Medida de confiabilidade

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
2	0,817	0,674	0,868	0,935	0,768	0,881	0,717	0,986	0,995	0,756
3	0,893	0,710	0,923	0,961	0,940	0,918	0,745	0,988	0,995	0,847
4	0,948	0,781	0,944	0,972	0,971	0,958	0,790	0,989	0,996	0,874
5	0,956	0,805	0,957	0,983	0,974	0,975	0,817	0,990	0,995	0,911
6	0,970	0,824	0,962	0,987	0,979	0,977	0,840	0,990	0,996	0,947
7	0,975	0,832	0,964	0,987	0,982	0,980	0,839	0,991	0,995	0,948
8	0,978	0,833	0,961	0,988	0,986	0,983	0,865	0,991	0,995	0,955
9	0,980	0,843	0,962	0,988	0,987	0,984	0,887	0,992	0,996	0,963
10	0,981	0,854	0,964	0,988	0,987	0,984	0,920	0,978	0,995	0,978
11	0,982	0,681	0,967	0,988	0,987	0,986	0,927	0,978	0,995	0,983
12	0,982	0,694	0,920	0,989	0,975	0,986	0,941	0,977	0,995	0,984
13	0,984	0,692	0,969	0,990	0,989	0,988	0,945	0,978	0,995	0,988
14	0,985	0,669	0,969	0,989	0,991	0,988	0,949	0,978	0,995	0,988
15	0,985	0,684	0,971	0,990	0,991	0,989	0,955	0,978	0,995	0,989
16	0,985	0,669	0,972	0,990	0,992	0,990	0,965	0,977	0,995	0,990
17	0,986	0,684	0,972	0,990	0,992	0,990	0,970	0,976	0,995	0,991
18	0,987	0,672	0,973	0,990	0,993	0,990	0,931	0,979	0,995	0,993
19	0,987	0,699	0,966	0,990	0,993	0,990	0,943	0,979	0,995	0,993
20	0,987	0,726	0,967	0,990	0,993	0,983	0,938	0,980	0,995	0,993
21	0,988	0,712	0,967	0,992	0,993	0,983	0,944	0,979	0,995	0,994
22	0,988	0,699	0,968	0,992	0,993	0,983	0,946	0,980	0,995	0,994
23	0,989	0,702	0,968	0,993	0,993	0,983	0,958	0,980	0,995	0,995
24	0,989	0,710	0,969	0,993	0,993	0,983	0,956	0,980	0,995	0,995
25	0,989	0,694	0,969	0,993	0,993	0,983	0,961	0,978	0,995	0,995
26	0,992	0,698	0,970							
27	0,992	0,695	0,970							
28	0,993	0,708	0,971							
29	0,993	0,707	0,972							
30	0,993	0,703	0,972							

31	0,994	0,718	0,972							
32	0,994	0,722	0,973							
33	0,994	0,722	0,974							
34	0,994	0,741	0,974							
35	0,995	0,742	0,974							
36	0,995	0,763								
37	0,995	0,777								
38	0,995									
39	0,996									
40	0,996									
41	0,996									
42	0,996									
43	0,996									
44	0,996									
45	0,996									
46	0,996									
47	0,996									
48	0,996									
49	0,996									

Tabela 34 – Confiabilidade para *Maximum Variance Unfolding*

7.2 Medida de continuidade

Na tabela 7.2 é exibido os resultados da métrica continuidade para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
2	0,957	0,702	0,975	0,975	0,816	0,991	0,857	0,993	0,997	0,737
3	0,966	0,766	0,981	0,983	0,977	0,992	0,892	0,993	0,997	0,792
4	0,977	0,854	0,983	0,985	0,984	0,994	0,924	0,994	0,997	0,799
5	0,981	0,878	0,987	0,988	0,984	0,995	0,943	0,994	0,997	0,846
6	0,983	0,896	0,988	0,989	0,987	0,995	0,955	0,994	0,997	0,908
7	0,985	0,909	0,988	0,989	0,989	0,995	0,959	0,994	0,997	0,910
8	0,986	0,913	0,988	0,990	0,990	0,995	0,972	0,994	0,997	0,920
9	0,987	0,923	0,988	0,991	0,990	0,995	0,980	0,994	0,997	0,935
10	0,988	0,937	0,989	0,991	0,990	0,996	0,991	0,538	0,762	0,977
11	0,989	0,871	0,990	0,991	0,991	0,995	0,993	0,538	0,762	0,981
12	0,988	0,890	0,849	0,991	0,974	0,995	0,995	0,538	0,762	0,948
13	0,989	0,904	0,991	0,991	0,991	0,995	0,996	0,538	0,762	0,984
14	0,990	0,908	0,991	0,990	0,992	0,996	0,997	0,538	0,762	0,985
15	0,990	0,921	0,992	0,990	0,992	0,996	0,998	0,538	0,762	0,986
16	0,991	0,925	0,992	0,990	0,993	0,996	0,998	0,538	0,762	0,987
17	0,991	0,938	0,992	0,992	0,993	0,996	0,999	0,538	0,762	0,987
18	0,991	0,944	0,992	0,992	0,993	0,996	0,993	0,538	0,762	0,988
19	0,992	0,955	0,990	0,992	0,993	0,996	0,994	0,538	0,762	0,988
20	0,991	0,973	0,991	0,992	0,994	0,667	0,993	0,538	0,762	0,988
21	0,992	0,976	0,991	0,993	0,994	0,667	0,994	0,538	0,762	0,989
22	0,992	0,976	0,991	0,993	0,994	0,667	0,994	0,538	0,762	0,989
23	0,992	0,979	0,991	0,993	0,994	0,667	0,995	0,538	0,762	0,990
24	0,992	0,982	0,992	0,994	0,994	0,667	0,995	0,538	0,762	0,990
25	0,993	0,980	0,992	0,994	0,994	0,667	0,995	0,538	0,762	0,990
26	0,992	0,983	0,992							
27	0,993	0,985	0,992							

28	0,993	0,986	0,992							
29	0,993	0,988	0,992							
30	0,993	0,988	0,993							
31	0,994	0,990	0,993							
32	0,993	0,990	0,993							
33	0,993	0,991	0,993							
34	0,994	0,992	0,993							
35	0,994	0,993	0,993							
36	0,994	0,994								
37	0,994	0,995								
38	0,994									
39	0,994									
40	0,995									
41	0,995									
42	0,995									
43	0,995									
44	0,995									
45	0,995									
46	0,995									
47	0,995									
48	0,995									
49	0,995									

Tabela 35 – Continuidade para *Maximum Variance Unfolding*

7.3 Erro de generalização

Na tabela 7.3 é exibido os resultados da métrica erro de generalização para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
	6.78%	14.36%	12.08%	4.70%	19.68%	53.60%	3.27%	79.68%	42.63%	9.23%
2	39,37%	86,85%	22,48%	23,50%	51,68%	61,77%	56,95%	81,80%	47,78%	11,42%
3	29,64%	80,36%	20,28%	18,85%	28,65%	60,66%	49,25%	82,70%	47,90%	11,70%
4	22,70%	75,05%	18,37%	15,58%	25,28%	60,78%	42,82%	84,76%	49,26%	11,64%
5	20,58%	68,59%	17,58%	12,93%	25,73%	59,86%	38,38%	84,55%	46,30%	10,97%
6	16,51%	64,38%	17,25%	11,65%	24,65%	59,27%	36,69%	83,70%	47,57%	11,16%
7	15,23%	62,86%	17,72%	12,20%	24,38%	59,16%	37,31%	86,08%	47,94%	11,52%
8	13,40%	60,06%	17,73%	11,45%	23,65%	59,39%	33,59%	84,18%	49,05%	11,40%
9	11,96%	59,16%	16,91%	11,83%	22,88%	59,06%	31,00%	85,50%	47,74%	10,53%
10	11,85%	59,00%	16,82%	12,08%	22,38%	59,17%	26,04%	0,00%	0,00%	4,14%
11	11,66%	72,81%	17,40%	12,20%	22,95%	59,12%	25,56%	0,00%	0,00%	4,71%
12	11,50%	71,96%	20,37%	12,05%	21,98%	57,74%	22,26%	0,00%	0,00%	0,00%
13	11,12%	74,74%	15,89%	11,90%	21,90%	57,81%	20,10%	0,00%	0,00%	4,89%
14	11,28%	74,78%	15,86%	12,35%	23,30%	59,54%	17,65%	0,00%	0,00%	5,13%
15	10,99%	77,34%	16,24%	12,85%	23,40%	57,44%	17,91%	0,00%	0,00%	5,40%
16	10,94%	77,69%	15,44%	13,50%	24,13%	57,90%	15,80%	0,00%	0,00%	5,07%
17	10,98%	75,94%	15,90%	13,78%	23,88%	57,45%	13,60%	0,00%	0,00%	5,27%
18	10,99%	76,22%	15,77%	13,83%	23,30%	57,86%	12,83%	0,00%	0,00%	5,48%
19	10,43%	73,12%	15,72%	13,28%	23,53%	57,79%	10,14%	0,00%	0,00%	5,42%
20	10,52%	67,98%	15,90%	12,78%	23,60%	0,00%	11,96%	0,00%	0,00%	5,27%
21	11,06%	69,09%	15,52%	12,53%	24,65%	0,00%	10,78%	0,00%	0,00%	5,42%
22	11,50%	68,39%	15,76%	11,43%	23,38%	0,00%	9,86%	0,00%	0,00%	5,46%
23	11,04%	70,47%	15,70%	11,45%	23,80%	0,00%	7,79%	0,00%	0,00%	6,17%

24	10,98%	68,05%	15,97%	10,55%	23,93%	0,00%	7,74%	0,00%	0,00%	5,46%
25	11,05%	71,96%	14,82%	11,33%	23,88%	0,00%	6,81%	0,00%	0,00%	6,00%
26	12,25%	69,54%	15,41%							
27	11,60%	69,30%	15,53%							
28	11,59%	67,01%	15,37%							
29	10,84%	67,72%	14,98%							
30	10,79%	67,69%	14,73%							
31	10,54%	65,80%	14,68%							
32	10,45%	64,17%	14,58%							
33	10,02%	63,14%	15,28%							
34	10,82%	61,31%	15,02%							
35	10,43%	58,48%	15,03%							
36	9,91%	56,49%								
37	9,89%	53,28%								
38	9,76%									
39	10,10%									
40	9,86%									
41	9,76%									
42	9,62%									
43	9,60%									
44	10,01%									
45	9,89%									
46	9,56%									
47	9,22%									
48	9,24%									
49	8,81%									

Tabela 36 – Erro de generalização para *Maximum Variance Unfolding*

.7.4 Medida independente de escala

Na tabela .7.4 é exibido os resultados da métrica independente de escala para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
2	0,199	0,031	0,269	0,433	0,144	0,423	0,068	0,796	0,909	0,110
3	0,277	0,037	0,374	0,486	0,434	0,439	0,083	0,800	0,911	0,141
4	0,336	0,044	0,420	0,518	0,505	0,513	0,102	0,807	0,911	0,158
5	0,364	0,047	0,486	0,555	0,509	0,579	0,120	0,814	0,912	0,204
6	0,384	0,049	0,506	0,572	0,543	0,580	0,117	0,814	0,910	0,277
7	0,394	0,048	0,512	0,571	0,558	0,591	0,135	0,809	0,912	0,284
8	0,408	0,045	0,505	0,587	0,575	0,609	0,165	0,814	0,913	0,310
9	0,415	0,049	0,507	0,596	0,577	0,623	0,207	0,817	0,913	0,339
10	0,423	0,048	0,514	0,595	0,578	0,635	0,308	0,169	0,501	0,450
11	0,426	0,027	0,522	0,593	0,583	0,660	0,324	0,169	0,501	0,479
12	0,422	0,029	0,214	0,596	0,461	0,663	0,386	0,169	0,501	0,509
13	0,426	0,030	0,544	0,599	0,592	0,665	0,397	0,169	0,501	0,518
14	0,435	0,029	0,546	0,597	0,610	0,669	0,435	0,169	0,501	0,536
15	0,439	0,030	0,557	0,603	0,617	0,668	0,456	0,169	0,501	0,546
16	0,443	0,031	0,564	0,603	0,629	0,671	0,504	0,169	0,501	0,557
17	0,448	0,039	0,565	0,617	0,632	0,675	0,539	0,169	0,501	0,569
18	0,452	0,040	0,566	0,620	0,636	0,675	0,505	0,169	0,501	0,582
19	0,458	0,050	0,563	0,623	0,638	0,678	0,538	0,169	0,501	0,586
20	0,456	0,057	0,569	0,625	0,645	0,297	0,536	0,169	0,501	0,592

21	0,455	0,064	0,575	0,630	0,646	0,297	0,530	0,169	0,501	0,605
22	0,460	0,062	0,582	0,644	0,647	0,296	0,529	0,169	0,501	0,608
23	0,467	0,068	0,582	0,646	0,646	0,298	0,542	0,169	0,501	0,622
24	0,476	0,073	0,583	0,652	0,647	0,297	0,546	0,168	0,501	0,630
25	0,482	0,068	0,586	0,658	0,649	0,298	0,551	0,169	0,501	0,637
26	0,500	0,076	0,588							
27	0,508	0,082	0,591							
28	0,511	0,090	0,595							
29	0,520	0,095	0,597							
30	0,524	0,098	0,596							
31	0,529	0,107	0,595							
32	0,530	0,110	0,600							
33	0,537	0,113	0,605							
34	0,542	0,134	0,607							
35	0,549	0,146	0,608							
36	0,555	0,157								
37	0,561	0,179								
38	0,565									
39	0,570									
40	0,576									
41	0,580									
42	0,581									
43	0,585									
44	0,586									
45	0,590									
46	0,590									
47	0,594									
48	0,597									
49	0,598									

Tabela 37 – Medida independente de escala para *Maximum Variance Unfolding*

.8 Método de redução *Classical Multidimensional Scaling*

Nesta seção é exibido os resultados das métricas computadas sobre os conjuntos de dados naturais usando o método de redução *Classical Multidimensional Scaling*.

.8.1 Medida de confiabilidade

Na tabela .8.1 é exibido os resultados da métrica confiabilidade para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
2	0,735	0,847	0,871	0,848	0,843	0,81	0,815	0,994	0,997	0,956
3	0,831	0,91	0,897	0,937	0,927	0,900	0,884	1,000	0,998	0,986
4	0,902	0,952	0,914	0,981	0,965	0,933	0,933	1,000	0,999	0,994
5	0,936	0,975	0,916	0,993	0,981	0,937	0,971	1,000	0,999	0,999
6	0,965	0,983	0,921	0,996	0,992	0,952	0,983	1,000	0,999	0,999
7	0,975	0,988	0,925	0,998	0,996	0,974	0,989	1,000	0,999	0,999
8	0,983	0,991	0,928	0,999	0,998	0,980	0,994	1,000	1,000	1,000

9	0,989	0,993	0,933	0,999	0,999	0,986	0,996	1,000	1,000	1,000
10	0,992	0,994	0,936	0,999	0,999	0,990	0,998	1,000	1,000	1,000
11	0,993	0,996	0,940	0,999	1,000	0,995	0,998	1,000	1,000	1,000
12	0,995	0,997	0,943	1,000	1,000	0,997	0,999	1,000	1,000	1,000
13	0,996	0,997	0,941	1,000	1,000	0,998	0,999	1,000	1,000	1,000
14	0,997	0,998	0,943	1,000	1,000	0,999	0,999	1,000	1,000	1,000
15	0,997	0,998	0,943	1,000	1,000	0,998	1,000	1,000	1,000	1,000
16	0,998	0,998	0,944	1,000	1,000	1,000	1,000	1,000	1,000	1,000
17	0,998	0,999	0,946	1,000	1,000	1,000	1,000	1,000	1,000	1,000
18	0,998	0,999	0,965	1,000	1,000	1,000	1,000	1,000	1,000	1,000
19	0,998	0,999	0,967	1,000	1,000	1,000	1,000	1,000	1,000	1,000
20	0,999	0,999	0,968	1,000	1,000	1,000	1,000	1,000	1,000	1,000
21	0,999	0,999	0,969	1,000	1,000	1,000	1,000	1,000	1,000	1,000
22	0,999	0,999	0,972	1,000	1,000	1,000	1,000	1,000	1,000	1,000
23	0,999	0,999	0,975	1,000	1,000	1,000	1,000	1,000	1,000	1,000
24	0,999	0,999	0,976	1,000	1,000	1,000	1,000	1,000	1,000	1,000
25	0,999	0,999	0,976	1,000	1,000	1,000	1,000	1,000	1,000	1,000
26	0,999	0,999	0,977							
27	0,999	0,999	0,977							
28	0,999	0,999	0,977							
29	0,999	0,999	0,979							
30	0,999	0,999	0,981							
31	1,000	0,999	0,981							
32	1,000	1,000	0,981							
33	1,000	1,000	0,982							
34	1,000	1,000	0,982							
35	1,000	1,000	0,983							
36	1,000	1,000								
37	1,000	1,000								
38	1,000									
39	1,000									
40	1,000									
41	1,000									
42	1,000									
43	1,000									
44	1,000									
45	1,000									
46	1,000									
47	1,000									
48	1,000									
49	1,000									

Tabela 38 – Confiabilidade para *Classical Multidimensional Scaling*

.8.2 Medida de continuidade

Na tabela .8.2 é exibido os resultados da métrica continuidade para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
2	0,937	0,960	0,945	0,975	0,963	0,947	0,966	0,998	0,998	0,991
3	0,971	0,982	0,954	0,990	0,986	0,959	0,983	1,000	0,999	0,997
4	0,984	0,990	0,957	0,996	0,993	0,964	0,992	1,000	0,999	0,999
5	0,990	0,993	0,961	0,998	0,996	0,983	0,995	1,000	0,999	0,999

6	0,993	0,995	0,964	0,999	0,998	0,990	0,997	1,000	1,000	1,000
7	0,995	0,996	0,969	0,999	0,999	0,992	0,998	1,000	1,000	1,000
8	0,996	0,997	0,970	0,999	0,999	0,993	0,999	1,000	1,000	1,000
9	0,997	0,997	0,970	0,999	0,999	0,994	0,999	1,000	1,000	1,000
10	0,998	0,998	0,971	1,000	1,000	0,996	0,999	1,000	1,000	1,000
11	0,998	0,998	0,973	1,000	1,000	0,997	0,999	1,000	1,000	1,000
12	0,998	0,998	0,974	1,000	1,000	0,997	1,000	1,000	1,000	1,000
13	0,999	0,999	0,979	1,000	1,000	0,999	1,000	1,000	1,000	1,000
14	0,999	0,999	0,979	1,000	1,000	0,999	1,000	1,000	1,000	1,000
15	0,999	0,999	0,979	1,000	1,000	1,000	1,000	1,000	1,000	1,000
16	0,999	0,999	0,980	1,000	1,000	1,000	1,000	1,000	1,000	1,000
17	0,999	0,999	0,981	1,000	1,000	1,000	1,000	1,000	1,000	1,000
18	0,999	0,999	0,983	1,000	1,000	1,000	0,999	1,000	1,000	1,000
19	0,999	0,999	0,983	1,000	1,000	1,000	0,999	1,000	1,000	1,000
20	0,999	0,999	0,984	1,000	1,000	1,000	0,999	1,000	1,000	1,000
21	0,999	0,999	0,985	1,000	1,000	1,000	0,999	1,000	1,000	1,000
22	0,999	0,999	0,987	1,000	1,000	1,000	0,999	1,000	1,000	1,000
23	1,000	0,999	0,988	1,000	1,000	1,000	0,999	1,000	1,000	1,000
24	1,000	0,999	0,988	1,000	1,000	1,000	0,999	1,000	1,000	1,000
25	1,000	1,000	0,988	1,000	1,000	1,000	0,999	1,000	1,000	1,000
26	1,000	1,000	0,988							
27	1,000	1,000	0,988							
28	1,000	1,000	0,987							
29	1,000	1,000	0,988							
30	1,000	1,000	0,989							
31	1,000	1,000	0,989							
32	1,000	1,000	0,989							
33	1,000	1,000	0,990							
34	1,000	1,000	0,990							
35	1,000	1,000	0,990							
36	1,000	1,000								
37	1,000	1,000								
38	1,000									
39	1,000									
40	1,000									
41	1,000									
42	1,000									
43	1,000									
44	1,000									
45	1,000									
46	1,000									
47	1,000									
48	1,000									
49	1,000									

Tabela 39 – Continuidade para *Classical Multidimensional Scaling*

.8.3 Erro de generalização

Na tabela .8.3 é exibido os resultados da métrica erro de generalização para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
	6.78%	14.36%	12.08%	4.70%	19.68%	53.60%	3.27%	79.68%	42.63%	9.23%

2	63,32%	74,61%	21,05%	60,45%	56,05%	59,95%	52,98%	81,59%	50,53%	11,34%
3	57,47%	65,60%	19,65%	39,95%	47,95%	49,11%	44,82%	79,37%	47,65%	10,63%
4	43,06%	50,36%	18,97%	27,53%	39,73%	44,74%	29,51%	80,16%	43,21%	10,77%
5	33,23%	39,81%	19,66%	17,23%	35,35%	44,87%	21,10%	78,41%	39,75%	9,41%
6	22,99%	34,28%	19,22%	15,83%	31,13%	47,01%	15,33%	80,00%	36,87%	9,86%
7	19,87%	29,69%	19,07%	13,55%	27,35%	47,83%	12,01%	80,79%	36,38%	9,43%
8	15,68%	28,14%	17,93%	11,55%	24,18%	48,34%	9,70%	80,79%	37,53%	9,55%
9	14,47%	28,05%	16,93%	9,28%	22,53%	47,61%	8,67%	81,38%	38,27%	8,95%
10	12,98%	27,40%	16,97%	9,55%	22,45%	49,56%	8,01%	80,63%	32,43%	9,45%
11	12,24%	24,78%	16,29%	8,95%	21,73%	50,21%	6,50%	79,58%	32,47%	8,97%
12	10,73%	24,55%	15,83%	8,13%	20,10%	50,28%	5,82%	79,31%	32,02%	9,23%
13	9,22%	23,35%	14,99%	7,70%	19,80%	51,77%	5,16%	80,63%	32,51%	8,97%
14	9,01%	22,77%	15,63%	7,10%	20,48%	51,99%	4,93%	79,42%	33,05%	9,09%
15	9,31%	21,51%	13,86%	6,03%	19,75%	53,77%	4,83%	80,26%	34,61%	9,29%
16	8,39%	20,82%	13,98%	6,35%	20,25%	53,36%	5,16%	78,04%	34,20%	8,86%
17	7,91%	19,46%	13,89%	6,55%	19,73%	53,87%	4,74%	81,16%	34,53%	8,95%
18	7,56%	20,05%	14,78%	6,05%	19,83%	53,31%	1,35%	80,42%	33,09%	9,21%
19	7,08%	18,22%	14,32%	5,68%	19,28%	54,68%	1,41%	80,63%	32,67%	8,97%
20	6,98%	18,94%	13,82%	6,03%	20,60%	53,42%	1,28%	79,79%	33,09%	8,62%
21	6,83%	17,57%	14,30%	6,03%	19,13%	53,13%	1,55%	79,10%	33,50%	9,37%
22	6,88%	16,76%	13,94%	5,33%	19,55%	53,46%	1,48%	79,37%	34,81%	8,80%
23	6,18%	17,54%	13,61%	5,33%	19,53%	53,09%	1,15%	79,63%	34,07%	8,34%
24	6,89%	17,10%	12,62%	5,88%	19,40%	53,97%	1,33%	79,79%	34,49%	8,92%
25	6,48%	17,10%	13,45%	4,63%	20,00%	53,51%	1,35%	79,74%	35,06%	9,57%
26	5,94%	16,32%	13,39%							
27	6,24%	15,32%	13,23%							
28	5,59%	15,36%	12,74%							
29	5,84%	15,21%	12,72%							
30	6,09%	15,43%	12,47%							
31	6,03%	16,05%	12,80%							
32	5,51%	15,26%	13,04%							
33	5,70%	15,28%	12,49%							
34	5,67%	15,01%	12,64%							
35	6,00%	14,80%	12,11%							
36	5,26%	14,96%								
37	5,53%	15,59%								
38	5,48%									
39	5,34%									
40	5,54%									
41	6,11%									
42	5,67%									
43	5,79%									
44	5,69%									
45	5,76%									
46	5,29%									
47	5,55%									
48	5,57%									
49	5,79%									

Tabela 40 – Erro de generalização para *Classical Multidimensional Scaling*

.8.4 Medida independente de escala

Na tabela .8.4 é exibido os resultados da métrica independente de escala para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
2	0,117	0,136	0,260	0,302	0,269	0,300	0,170	0,935	0,977	0,471
3	0,197	0,242	0,287	0,471	0,420	0,396	0,291	1,000	0,980	0,663
4	0,278	0,345	0,306	0,633	0,524	0,450	0,395	1,000	0,982	0,766
5	0,346	0,412	0,316	0,717	0,593	0,513	0,478	1,000	0,983	0,878
6	0,424	0,455	0,324	0,780	0,675	0,580	0,532	1,000	0,984	0,909
7	0,462	0,485	0,336	0,814	0,736	0,666	0,581	1,000	0,984	0,924
8	0,500	0,522	0,346	0,847	0,784	0,696	0,629	1,000	0,985	0,942
9	0,531	0,547	0,355	0,867	0,828	0,733	0,657	1,000	0,986	0,953
10	0,557	0,569	0,366	0,892	0,860	0,789	0,697	1,000	0,986	0,960
11	0,573	0,594	0,374	0,910	0,888	0,831	0,728	1,000	0,986	0,965
12	0,597	0,611	0,381	0,922	0,905	0,871	0,751	1,000	0,986	0,969
13	0,619	0,632	0,392	0,930	0,922	0,906	0,774	1,000	0,986	0,973
14	0,638	0,649	0,397	0,940	0,940	0,932	0,790	1,000	0,987	0,976
15	0,649	0,663	0,406	0,947	0,951	0,965	0,806	1,000	0,987	0,980
16	0,664	0,675	0,415	0,952	0,969	0,994	0,822	1,000	0,987	0,983
17	0,674	0,690	0,419	0,958	0,979	0,994	0,835	1,000	0,987	0,985
18	0,685	0,704	0,448	0,966	0,984	0,995	0,808	1,000	0,987	0,986
19	0,697	0,712	0,454	0,970	0,989	0,995	0,816	1,000	0,988	0,987
20	0,706	0,722	0,457	0,973	0,992	0,995	0,826	1,000	0,988	0,989
21	0,716	0,730	0,467	0,975	0,994	0,996	0,832	1,000	0,988	0,989
22	0,724	0,740	0,482	0,978	0,996	0,996	0,839	1,000	0,988	0,991
23	0,732	0,747	0,497	0,979	0,996	0,996	0,845	1,000	0,988	0,992
24	0,741	0,753	0,497	0,981	0,997	0,996	0,850	1,000	0,988	0,992
25	0,747	0,760	0,498	0,982	0,998	0,996	0,855	1,000	0,988	0,993
26	0,753	0,766	0,500							
27	0,761	0,772	0,500							
28	0,767	0,777	0,499							
29	0,772	0,782	0,506							
30	0,777	0,788	0,509							
31	0,783	0,792	0,513							
32	0,788	0,797	0,513							
33	0,793	0,801	0,520							
34	0,797	0,805	0,523							
35	0,802	0,809	0,525							
36	0,806	0,812								
37	0,810	0,816								
38	0,814									
39	0,818									
40	0,822									
41	0,825									
42	0,829									
43	0,832									
44	0,835									
45	0,839									
46	0,842									
47	0,845									
48	0,847									
49	0,850									

Tabela 41 – Medida independente de escala para *Classical Multidimensional Scaling*

40	1,000									
41	1,000									
42	1,000									
43	1,000									
44	1,000									
45	1,000									
46	1,000									
47	1,000									
48	1,000									
49	1,000									

Tabela 43 – Continuidade para *Principal Components Analysis*

.9.3 Erro de generalização

Na tabela .9.3 é exibido os resultados da métrica erro de generalização para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
	6.78%	14.36%	12.08%	4.70%	19.68%	53.60%	3.27%	79.68%	42.63%	9.23%
2	63,30%	75,29%	21,42%	59,73%	56,78%	59,38%	52,34%	81,96%	49,63%	10,69%
3	58,06%	65,29%	19,64%	39,83%	46,88%	50,12%	44,57%	81,27%	47,04%	11,08%
4	42,91%	50,96%	19,37%	26,40%	39,83%	45,30%	29,42%	79,42%	42,02%	10,12%
5	32,78%	40,33%	19,59%	17,50%	36,58%	45,12%	21,75%	80,32%	39,59%	9,53%
6	23,48%	34,07%	18,99%	15,28%	30,20%	47,90%	15,68%	80,32%	38,40%	9,61%
7	19,30%	30,46%	19,23%	13,08%	26,83%	47,15%	12,07%	80,37%	37,78%	9,88%
8	15,34%	27,64%	18,29%	10,63%	25,13%	48,13%	9,43%	79,52%	37,57%	9,29%
9	14,25%	27,72%	17,41%	9,80%	23,70%	48,02%	8,42%	81,16%	38,11%	9,03%
10	13,04%	26,89%	17,06%	9,35%	21,63%	48,81%	7,71%	80,11%	37,49%	9,57%
11	12,21%	24,89%	16,64%	9,50%	21,23%	49,92%	6,35%	80,63%	36,71%	9,19%
12	10,72%	24,29%	15,56%	8,17%	22,13%	50,19%	5,89%	80,63%	37,90%	8,82%
13	9,20%	23,20%	15,05%	7,83%	21,20%	51,57%	5,34%	80,05%	37,94%	9,31%
14	9,01%	22,99%	15,48%	7,50%	19,90%	50,65%	5,20%	79,58%	38,89%	8,88%
15	9,33%	21,15%	14,28%	6,23%	20,30%	52,93%	4,84%	80,74%	35,88%	9,11%
16	7,77%	21,39%	13,99%	6,83%	20,58%	53,56%	4,79%	80,74%	37,20%	9,64%
17	8,01%	19,74%	14,12%	6,57%	19,70%	52,39%	4,56%	79,42%	38,85%	9,15%
18	7,14%	19,82%	14,33%	5,87%	20,93%	53,34%	4,38%	81,27%	37,12%	9,70%
19	7,16%	18,62%	14,39%	6,18%	19,78%	53,08%	4,49%	80,90%	38,15%	9,47%
20	7,43%	17,77%	13,81%	5,55%	19,85%	53,32%	3,85%	79,84%	38,48%	9,19%
21	6,78%	17,62%	14,67%	5,65%	20,18%	53,19%	4,08%	81,48%	37,37%	8,88%
22	6,59%	17,15%	14,10%	5,55%	19,65%	53,84%	3,99%	78,41%	38,27%	9,72%
23	6,63%	17,45%	13,02%	4,95%	19,73%	52,76%	3,79%	80,11%	39,26%	8,74%
24	6,37%	17,37%	12,84%	5,38%	18,58%	53,46%	3,70%	79,26%	37,45%	9,03%
25	6,17%	17,21%	12,91%	5,15%	19,75%	52,97%	3,73%	81,16%	39,88%	9,51%
26	5,98%	15,80%	13,53%							
27	5,94%	15,46%	13,19%							
28	6,10%	15,94%	13,08%							
29	6,14%	15,54%	12,83%							
30	5,78%	15,19%	13,26%							
31	6,10%	15,47%	12,99%							
32	6,02%	15,15%	13,97%							
33	5,92%	15,30%	12,61%							
34	5,42%	15,20%	12,84%							
35	5,99%	14,96%	12,90%							

36	5,60%	15,28%								
37	5,12%	14,96%								
38	5,18%									
39	5,64%									
40	5,79%									
41	5,79%									
42	5,46%									
43	5,94%									
44	5,55%									
45	5,89%									
46	6,17%									
47	5,97%									
48	5,82%									
49	5,90%									

Tabela 44 – Erro de generalização para *Principal Components Analysis*

.9.4 Medida independente de escala

Na tabela .9.4 é exibido os resultados da métrica independente de escala para os conjuntos de dados naturais.

D	MNIST	ISOLET	NOMAO	FACTORS	ZERNIKE	GAS	USPS	EATING	HILL	OZONE
2	0,117	0,136	0,260	0,302	0,269	0,300	0,170	0,935	0,977	0,471
3	0,197	0,242	0,287	0,471	0,420	0,396	0,291	1,000	0,980	0,663
4	0,278	0,345	0,306	0,633	0,524	0,450	0,395	1,000	0,981	0,766
5	0,346	0,412	0,316	0,717	0,593	0,513	0,478	1,000	0,983	0,878
6	0,423	0,455	0,324	0,780	0,675	0,580	0,532	1,000	0,984	0,909
7	0,462	0,485	0,336	0,814	0,736	0,666	0,581	1,000	0,984	0,924
8	0,500	0,522	0,346	0,847	0,784	0,696	0,629	1,000	0,985	0,942
9	0,531	0,547	0,355	0,867	0,828	0,733	0,657	1,000	0,985	0,953
10	0,557	0,569	0,366	0,892	0,860	0,789	0,697	1,000	0,986	0,960
11	0,573	0,594	0,374	0,910	0,888	0,831	0,728	1,000	0,986	0,965
12	0,597	0,611	0,381	0,922	0,905	0,871	0,751	1,000	0,986	0,969
13	0,619	0,632	0,392	0,930	0,922	0,906	0,774	1,000	0,986	0,973
14	0,638	0,649	0,398	0,940	0,940	0,932	0,790	1,000	0,987	0,976
15	0,648	0,663	0,407	0,947	0,951	0,965	0,806	1,000	0,987	0,980
16	0,664	0,675	0,415	0,952	0,969	0,994	0,822	1,000	0,987	0,983
17	0,673	0,690	0,419	0,958	0,978	0,994	0,835	1,000	0,987	0,985
18	0,685	0,704	0,449	0,966	0,984	0,995	0,844	1,000	0,987	0,986
19	0,697	0,712	0,455	0,970	0,989	0,995	0,853	1,000	0,987	0,987
20	0,706	0,722	0,458	0,973	0,992	0,995	0,866	1,000	0,987	0,989
21	0,716	0,730	0,466	0,975	0,994	0,996	0,874	1,000	0,987	0,989
22	0,724	0,740	0,482	0,978	0,995	0,996	0,883	1,000	0,988	0,991
23	0,732	0,747	0,497	0,979	0,996	0,996	0,892	1,000	0,988	0,992
24	0,741	0,753	0,497	0,981	0,997	0,996	0,898	1,000	0,988	0,992
25	0,747	0,760	0,498	0,982	0,998	0,996	0,904	1,000	0,988	0,993
26	0,753	0,766	0,500							
27	0,761	0,772	0,500							
28	0,767	0,777	0,499							
29	0,772	0,782	0,506							
30	0,777	0,788	0,509							
31	0,783	0,792	0,512							
32	0,788	0,797	0,512							

33	0,793	0,801	0,520							
34	0,797	0,805	0,522							
35	0,802	0,809	0,525							
36	0,806	0,812								
37	0,810	0,816								
38	0,814									
39	0,817									
40	0,822									
41	0,825									
42	0,829									
43	0,832									
44	0,835									
45	0,839									
46	0,842									
47	0,845									
48	0,847									
49	0,850									

Tabela 45 – Medida independente de escala para *Principal Components Analysis*

Referências

- ALCALÁ-FDEZ, J. et al. . keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic and Soft Computing*, v. 17, p. 255–287, 2011. Citado na página 92.
- ALCALÁ-FDEZ, J. et al. Keel: A software tool to assess evolutionary algorithms to data mining problems. *Soft Computing*, v. 13, p. 307–318, 2009. Citado na página 92.
- ARÉVALO, F. L. G. *Redução de Dimensionalidade usando Isomap Aplicada ao Áudio Espacial*. 2014. Citado na página 50.
- BELKIN, M. *Problems of Learning on Manifolds*. Tese (Doutorado) — The Faculty of the Division of the Physical Sciences, Chicago, Illinois, 2003. Citado na página 58.
- BELKIN, M.; NIYOGI, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, v. 15, p. 1373–1396, 2003. Citado 2 vezes nas páginas 23 e 57.
- BELKIN, M.; NIYOGI, P. Towards a theoretical foundation for laplacian-based manifold methods. *Journal of Computer and System Sciences*, v. 74, p. 1289–1308, 2008. Citado na página 58.
- BENGIO, Y. et al. Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering. In: *Proceedings of the 16th International Conference on Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2003. (NIPS'03), p. 177–184. Disponível em: <<http://dl.acm.org/citation.cfm?id=2981345.2981368>>. Citado na página 97.
- C., G. J. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, v. 3 and 4, n. 53, p. 325–338, 1966. Citado na página 44.
- CAMASTRA, F. Data dimensionality estimation methods: A survey. *Pattern Recognition*, v. 36, p. 2945–2954, 2003. Citado na página 19.
- CHEN, L.; BUJA, A. *Local Multidimensional Scaling for Nonlinear Dimension Reduction, Graph Layout and Proximity Analysis*. Tese (Doutorado) — University of Pennsylvania, 2006. Citado na página 83.
- CHOI, H.; CHOI, S. Robust kernel isomap. *The journal of the pattern recognition society*, v. 40, p. 853–862, 2007. Citado na página 95.
- COIFMAN, R. R.; LAFON, S. Diffusion maps. *Applied and Computational Harmonic Analysis*, v. 21, p. 5–30, 2006. Citado 2 vezes nas páginas 23 e 73.
- COIFMAN, R. R. et al. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *In PNAS*, v. 102, p. 7426–7431, 2005. Citado na página 23.
- DIJKSTRA, E. W. A note on two problems in connexion with graphs. *Numerische Mathematik*, v. 1, n. 1, p. 269–271, 1959. Citado 2 vezes nas páginas 22 e 50.

- DONOHO, D.; GRIMES, C. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. In: *Proceedings of the National Academy of Sciences*. [S.l.: s.n.], 2003. Citado 2 vezes nas páginas 23 e 66.
- FUKUNAGA, K. Intrinsic dimensionality extraction. In: _____. *Handbook of Statistics*. North Holland: [s.n.], 1982. p. 347–362. Citado na página 19.
- HINO, H. ider: Intrinsic dimension estimation with r. *The R Journal*, v. 9, p. 329–341, 2017. Citado na página 86.
- HOTELLING, H. *Analysis of a Complex of Statistical Variables into Principal Components*. [S.l.]: Warwick York Inc. - Baltimore, 1933. Citado na página 20.
- KRUSKAL, J. B. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, v. 29, n. 2, p. 115–129, 1964. Citado na página 21.
- KRUSKAL, J. B.; HILL, N. J. M. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, v. 29, n. 1, p. 1–27, 1964. Citado na página 21.
- LAFON, S.; LEE, A. B. Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 28, n. 9, p. 1393–1403, 2006. Citado na página 23.
- LAFON, S. S. *Diffusion Maps and Geometric Harmonics*. Tese (Doutorado) — Faculty of the Graduate School of Yale University, USA, Yale, 2004. Citado na página 23.
- LEE, J. A.; VERLEYSSEN, M. Nonlinear dimensionality reduction of data manifolds with essential loops. *Neurocomputing*, v. 67, p. 29–53, 2005. Citado na página 90.
- LEE, J. A.; VERLEYSSEN, M. Quality assessment of dimensionality reduction: Rank-based criteria. *Neurocomputing*, v. 72, p. 1431–1443, 2009. Citado 3 vezes nas páginas 81, 82 e 83.
- LEE, J. A.; VERLEYSSEN, M. Scale-independent quality criteria for dimensionality reduction. *Pattern Recognition Letters*, v. 31, p. 2248–2257, 2010. Citado 2 vezes nas páginas 82 e 83.
- MAATEN, L. van der. Matlab toolbox for dimensionality reduction. *Proceedings of the Belgian-Dutch Artificial Intelligence Conference*, 2007. Citado na página 86.
- PEARSON, K. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, v. 2, n. 11, p. 559–572, 1901. Citado 2 vezes nas páginas 20 e 37.
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011. Citado na página 85.
- PLESS, R. Image spaces and video trajectories: Using isomap to explore video sequence. p. 1–8, 2003. Citado na página 50.
- ROWEIS, S. T.; SAUL, L. K. Nonlinear dimensionality reduction by locally linear embedding. *Science*, v. 290, n. 5500, p. 2323–2325, 2000. Citado 2 vezes nas páginas 22 e 54.
- SAXENA, A.; GUPTA, A.; MUKERJEE, A. Non-linear dimensionality reduction by locally linear isomaps. *Lecture Notes in Computer Science*, v. 3316, p. 1038–1043, 2004. Citado na página 95.

- SHEPARD, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function i. *Psychometrika*, v. 27, n. 2, p. 125–140, 1962. Citado na página 21.
- SHEPARD, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function ii. *Psychometrika*, v. 27, n. 3, p. 219–246, 1962. Citado na página 21.
- T., K.; H., T. International comparisons of prices and quantities consumed. *Econometrica*, n. 33, p. 535–556, 1965. Citado na página 44.
- TENENBAUM, J. B.; SILVA, V. de; LANGFORD, J. C. A global geometric framework for nonlinear dimensionality reduction. *Science*, v. 290, n. 2319, p. 2319–2323, 2000. Citado 2 vezes nas páginas 22 e 50.
- TORGERSON, W. S. Multidimensional scaling: I. theory and method. *Psychometrika*, v. 17, n. 4, p. 401–419, 1952. Citado na página 20.
- TORGERSON, W. S. *Theory and Methods of Scaling*. [S.l.]: Wiley - New York, 1958. Citado na página 44.
- VANSCHOREN, J. et al. Openml: networked science in machine learning. *CoRR*, abs/1407.7722, 2014. Disponível em: <<http://arxiv.org/abs/1407.7722>>. Citado na página 85.
- WANG, C. et al. Wireless sensor networks localization with isomap. p. 1–5, 2009. Citado na página 50.
- WANG, J. *Analysis of a Complex of Statistical Variables into Principal Components*. [S.l.]: Higher Education Press and Springer Heidelberg Dordrecht London - Beijing and New York, 2012. Citado 2 vezes nas páginas 24 e 40.
- WEINBERGER, K. Q.; PACKER, B. D.; SAUL, L. K. Nonlinear dimensionality reduction by semidefinite programming and kernel matrix factorization. In: *Proceedings of the 10th International Workshop on AI and Statistics*. [S.l.: s.n.], 2005. Citado 2 vezes nas páginas 22 e 77.
- WEINBERGER, K. Q.; SAUL, L. K. Unsupervised learning of image manifolds by semidefinite programming. In: *IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2004. Citado 2 vezes nas páginas 22 e 77.
- YANG, L. Building connected neighborhood graphs for isometric data embedding. In *KDD*, p. 722–728, 2005. Citado na página 48.
- YOUNG, G.; HOUSEHOLDER, A. S. Discussion of a set of points in terms of their mutual distances. *Psychometrika*, v. 3, n. 1, p. 1–3, 1938. Citado na página 20.
- ZHANG, Z.; ZHA, H. Principal manifolds and nonlinear dimension reduction via local tangent space alignment. *SIAM Journal on Scientific Computation*, v. 26, p. 313–338, 2004. Citado 2 vezes nas páginas 22 e 60.