

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

Análise comparativa de métodos de redução de dimensionalidade de dados multivariados

Alexandre Cançado Cardoso

JUIZ DE FORA
FEVEREIRO, 2014

Análise comparativa de métodos de redução de dimensionalidade de dados multivariados

ALEXANDRE CANÇADO CARDOSO

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Bacharelado em Ciência da Computação

Orientador: Custódio Gouvêa Lopes da Motta

JUIZ DE FORA
FEVEREIRO, 2014

ANÁLISE COMPARATIVA DE MÉTODOS DE REDUÇÃO DE DIMENSIONALIDADE DE DADOS MULTIVARIADOS

Alexandre Cançado Cardoso

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTEGRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

Custódio Gouvêa Lopes da Motta
D. Eng.

Raul Fonseca Neto
D. Eng.

Carlos Cristiano Hasenclever Borges
D. Eng.

JUIZ DE FORA
07 DE FEVEREIRO, 2014

Para Neusa Cançado.

Resumo

Atualmente o volume de dados e o número de variáveis avaliadas em problemas das mais diversas áreas crescem rapidamente. Assim, aumenta cada vez mais a dificuldade para extrair informações e conhecimentos destes dados. A utilização de métodos computacionais para a manipulação de dados torna-se cada vez mais necessária e frequente. É neste contexto que a redução de dimensionalidade se insere como uma ferramenta poderosa para tornar os dados multivariados manipuláveis, interpretáveis e compreensíveis tanto por humanos, quanto por máquinas, de forma que facilite a obtenção de conhecimentos a partir dos dados, e não gere (ou minimize) a perda de informações relevantes. Este trabalho apresenta diversos métodos de redução de dimensionalidade existentes dentro do contexto de descoberta de conhecimento em bancos de dados e faz uma análise comparativa através de avaliações das características, vantagens, desvantagens e aplicações de cada um. Finalmente, são definidas metodologias e formas de uso, cenários e aplicações mais adequadas a cada um desses métodos, visando à melhoria do processo de descoberta de conhecimento tanto para identificação de padrões pela mineração dos dados, quanto para a visualização destes dados no pré e no pós-processamento.

Palavras-chave: descoberta de conhecimento em bancos de dados, mineração de dados, redução de dimensionalidade.

Abstract

Currently the volume of data and number of variables evaluated in problems from various areas grow rapidly. And more and more, increases the difficulty of extracting information and knowledge from them. Thus, the use of computational methods for the manipulation of data is ever more necessary and frequent. It is in this context that the dimensionality reduction fits as a powerful tool to make multivariate data manageable, interpretable and understandable either by humans as well as by machines. It is in this context that the dimensionality reduction fits as a tool to make the data manageable, interpretable and understandable either by humans as well as by machines, in order to facilitate the obtaining of knowledge from data, and does not generate (or minimize) the loss of relevant information. This paper will present several existing methods of dimensionality reduction within the context of knowledge discovery in databases and analyze them comparatively through the assessments of its features, advantage, disadvantages and application. Defining, then, methodologies and forms of use, settings and applications better suited to each method, aiming to improve a process of knowledge discovery both for pattern discovery through data mining and for viewing this as the pre-and post-processing.

Keywords: dimensionality reduction, knowledge discovery in databases, data mining.

Agradecimentos

Primeiramente ao Senhor por ter me guiado através das dificuldades mesmo quando eu teimava em não segui-lo. Aos meus pais e avós pelo exemplo e suporte para a realização do curso. A minha namorada pela dedicação, apoio e motivação para a conclusão. Aos familiares pela torcida e aos colegas pelo companheirismo nos estudos e trabalhos.

Aos professores pelos valiosos ensinamentos, os quais não se restringem ao âmbito acadêmico e profissional. Em especial ao Professor Custódio Motta, pela orientação nesta monografia. E, em conjunto aos demais colaboradores da Universidade Federal de Juiz de Fora, por me terem proporcionado este ensino de excelência.

*"Ainda que eu ande pelo vale da sombra
da morte, não temerei mal nenhum, por-
que Tu estás comigo; o teu bordão e o teu
cajado me consolam." Sl 23.4*

Sumário

Lista de Figuras	8
Lista de Tabelas	12
Lista de Abreviações	13
1 Introdução	14
1.1 CONTEXUALIZAÇÃO	14
1.1.1 Variáveis Aleatórias, Vetores Aleatórios e Matriz de Dados	14
1.1.2 Dimensão do Espaço de Características	15
1.1.3 Redução de Dimensionalidade	15
1.2 JUSTIFICATIVA	16
1.3 OBJETIVOS	18
1.3.1 Objetivos Gerais	18
1.3.2 Objetivos Específicos	18
1.4 METODOLOGIA	19
1.5 ESTRUTURA DO TRABALHO	20
2 Métodos	21
2.1 ANÁLISE DE COMPONENTES PRINCIPAIS	21
2.1.1 Definição Matemática	25
2.1.2 Obtenção Das Componentes Principais	28
2.1.3 Obtenção Das Componentes Principais Para Dados Padronizados .	33
2.1.4 Obtenção Das Componentes Principais Através Da Decomposição Em Valor Singular	35
2.1.5 Redução De Dimensionalidade A Partir Das Componentes Principais	38
2.2 DECOMPOSIÇÃO EM VALOR SINGULAR	40
2.3 ESCALONAMENTO MULTIDIMENSIONAL	43
2.3.1 Escalonamento Multidimensional Clássico	45
2.3.2 Escalonamento Multidimensional Não Clássico	49
2.4 FASTMAP	52
3 Experimentos	56
3.1 BASE DE DADOS	57
3.2 MÉTODOS	57
3.3 MÉTRICAS	60
3.4 RESULTADOS	64
4 Análise Comparativa	77
4.1 COMPLEXIDADE E TEMPO DE EXECUÇÃO	77
4.2 PROPORÇÃO DA VARIÂNCIA ACUMULADA	82
4.3 STRAIN	85
4.4 STRESS-1	87
4.4.1 Métrica Euclidiana Para Matriz De Distância Dos Scores	87

4.4.2	Mesma Métrica da Matriz de Distância De Entrada Para Matriz De Distância Dos Scores	88
4.5	SSTRESS	89
4.5.1	Métrica Euclidiana Para Matriz De Distância Dos Scores	89
4.5.2	Mesma Métrica da Matriz de Distância De Entrada Para Matriz De Distância Dos Scores	90
4.6	ACURÁCIA DO CLASSIFICADOR	90
4.7	VISUALIZAÇÃO DOS DADOS	95
5	Considerações Finais	100
A	Resultados	102
B	Gráficos dos resultados	151
C	Visualização dos Dados	195

Lista de Figuras

2.1 <i>Scree plot</i> do PCA para base de dados Iris não padronizada.	39
B.1 Legendas utilizada nos gráficos de tempo de execução e das métricas de qualidade das representações obtidas	152
B.2 Tempo de execução para base Triagem de Crédito (Legenda Fig. B.1) . . .	152
B.3 Tempo de execução para base Imagens (Legenda Fig. B.1)	153
B.4 Tempo de execução para base Iris (Legenda Fig. B.1)	154
B.5 Tempo de execução para base Sonar (Legenda Fig. B.1)	155
B.6 Tempo de execução para base Espiral (Legenda Fig. B.1)	156
B.7 Tempo de execução para base Vogal (Legenda Fig. B.1)	157
B.8 Tempo de execução para base Triagem de Crédito - Excluindo métodos baseados em otimização	158
B.9 Tempo de execução para base Imagens - Excluindo métodos baseados em otimização	159
B.10 Tempo de execução para base Iris - Excluindo métodos baseados em otimização	160
B.11 Tempo de execução para base Sonar - Excluindo métodos baseados em otimização	161
B.12 Tempo de execução para base Espiral - Excluindo métodos baseados em otimização	162
B.13 Tempo de execução para base Vogal - Excluindo métodos baseados em otimização	163
B.14 Proporção da variância acumulada para base Triagem de Crédito (Legenda Fig. B.1)	164
B.15 Proporção da variância acumulada para base Imagens (Legenda Fig. B.1) .	165
B.16 Proporção da variância acumulada para base Iris (Legenda Fig. B.1) . .	166
B.17 Proporção da variância acumulada para base Sonar (Legenda Fig. B.1) .	167
B.18 Proporção da variância acumulada para base Espiral (Legenda Fig. B.1) .	168
B.19 Proporção da variância acumulada para base Vogal (Legenda Fig. B.1) .	169
B.20 <i>Strain</i> para base Triagem de Crédito (Legenda Fig. B.1)	170
B.21 <i>Strain</i> para base Imagens (Legenda Fig. B.1)	171
B.22 <i>Strain</i> para base Iris (Legenda Fig. B.1)	172
B.23 <i>Strain</i> para base Sonar (Legenda Fig. B.1)	173
B.24 <i>Strain</i> para base Espiral (Legenda Fig. B.1)	174
B.25 <i>Strain</i> para base Vogal (Legenda Fig. B.1)	175
B.26 <i>Stress</i> – 1 (formulação tradicional) para base Triagem de Crédito (Legenda Fig. B.1)	176
B.27 <i>Stress</i> – 1 (formulação tradicional) para base Imagens (Legenda Fig. B.1)	177
B.28 <i>Stress</i> – 1 (formulação tradicional) para base Iris (Legenda Fig. B.1) . .	178
B.29 <i>Stress</i> – 1 (formulação tradicional) para base Sonar (Legenda Fig. B.1) .	179
B.30 <i>Stress</i> – 1 (formulação tradicional) para base Espiral (Legenda Fig. B.1) .	180
B.31 <i>Stress</i> – 1 (formulação tradicional) para base Vogal (Legenda Fig. B.1) .	181
B.32 <i>Stress</i> – 1 (formulação alternativa) para base Triagem de Crédito (Legenda Fig. B.1)	181

B.33 <i>Stress</i> – 1 (formulação alternativa) para base Imagens (Legenda Fig. B.1)	182
B.34 <i>Stress</i> – 1 (formulação alternativa) para base Iris (Legenda Fig. B.1)	182
B.35 <i>Stress</i> – 1 (formulação alternativa) para base Sonar (Legenda Fig. B.1)	182
B.36 <i>Stress</i> – 1 (formulação alternativa) para base Espiral (Legenda Fig. B.1)	183
B.37 <i>Stress</i> – 1 (formulação alternativa) para base Vogal (Legenda Fig. B.1)	184
B.38 <i>Sstress</i> (formulação tradicional) para base Triagem de Crédito (Legenda Fig. B.1)	184
B.39 <i>Sstress</i> (formulação tradicional) para base Imagens (Legenda Fig. B.1)	185
B.40 <i>Sstress</i> (formulação tradicional) para base Iris (Legenda Fig. B.1)	186
B.41 <i>Sstress</i> (formulação tradicional) para base Sonar (Legenda Fig. B.1)	187
B.42 <i>Sstress</i> (formulação tradicional) para base Espiral (Legenda Fig. B.1)	188
B.43 <i>Sstress</i> (formulação tradicional) para base Vogal (Legenda Fig. B.1)	189
B.44 <i>Sstress</i> (formulação alternativa) para base Triagem de Crédito (Legenda Fig. B.1)	189
B.45 <i>Sstress</i> (formulação alternativa) para base Imagens (Legenda Fig. B.1)	190
B.46 <i>Sstress</i> (formulação alternativa) para base Iris (Legenda Fig. B.1)	190
B.47 <i>Sstress</i> (formulação alternativa) para base Sonar (Legenda Fig. B.1)	190
B.48 <i>Sstress</i> (formulação alternativa) para base Espiral (Legenda Fig. B.1)	191
B.49 <i>Sstress</i> (formulação alternativa) para base Vogal (Legenda Fig. B.1)	191
B.50 Legendas utilizada nos gráficos de acurácia do classificador nos cenários X e D_{Euclidiana} e X e D_{Manhattan}	191
B.51 Legendas utilizada nos gráficos de acurácia do classificador nos cenários Z e D_{Euclidiana} e Z e D_{Manhattan}	192
B.52 Acurácia do classificador para base Triagem de Crédito	192
B.53 Acurácia do classificador para base Imagens	192
B.54 Acurácia do classificador para base Iris	193
B.55 Acurácia do classificador para base Sonar	193
B.56 Acurácia do classificador para base Espiral	194
B.57 Acurácia do classificador para base Vogal	194
C.1 <i>Matrix plot</i> para base Iris	196
C.2 <i>Matrix plot</i> para base Espiral	197
C.3 <i>Matrix plot</i> para base Vogal	198
C.4 Gráfico de dispersão para as representações unidimensionais a partir da base Triagem de Crédito não padronizada	199
C.5 Gráfico de dispersão para as representações unidimensionais a partir da base Imagens não padronizada	200
C.6 Gráfico de dispersão para as representações unidimensionais a partir da base Iris não padronizada	201
C.7 Gráfico de dispersão para as representações unidimensionais a partir da base Sonar não padronizada	202
C.8 Gráfico de dispersão para as representações unidimensionais a partir da base Espiral não padronizada	203
C.9 Gráfico de dispersão para as representações unidimensionais a partir da base Vogal não padronizada	204
C.10 Gráfico de dispersão para as representações unidimensionais a partir da base Triagem de Crédito padronizada	205
C.11 Gráfico de dispersão para as representações unidimensionais a partir da base Imagens padronizada	206

C.12 Gráfico de dispersão para as representações unidimensionais a partir da base Iris padronizada	207
C.13 Gráfico de dispersão para as representações unidimensionais a partir da base Sonar padronizada	208
C.14 Gráfico de dispersão para as representações unidimensionais a partir da base Espiral padronizada	209
C.15 Gráfico de dispersão para as representações unidimensionais a partir da base Vogal padronizada	210
C.16 Gráfico de dispersão para as representações bidimensionais a partir da base Triagem de Crédito não padronizada	211
C.17 Gráfico de dispersão para as representações bidimensionais a partir da base Imagens não padronizada	213
C.18 Gráfico de dispersão para as representações bidimensionais a partir da base Iris não padronizada	215
C.19 Gráfico de dispersão para as representações bidimensionais a partir da base Sonar não padronizada	217
C.20 Gráfico de dispersão para as representações bidimensionais a partir da base Espiral não padronizada	219
C.21 Gráfico de dispersão para as representações bidimensionais a partir da base Vogal não padronizada	221
C.22 Gráfico de dispersão para as representações bidimensionais a partir da base Triagem de Crédito padronizada	223
C.23 Gráfico de dispersão para as representações bidimensionais a partir da base Imagens padronizada	225
C.24 Gráfico de dispersão para as representações bidimensionais a partir da base Iris padronizada	227
C.25 Gráfico de dispersão para as representações bidimensionais a partir da base Sonar padronizada	229
C.26 Gráfico de dispersão para as representações bidimensionais a partir da base Espiral padronizada	231
C.27 Gráfico de dispersão para as representações bidimensionais a partir da base Vogal padronizada	233
C.28 Gráfico de dispersão para as representações tridimensionais a partir da base Triagem de Crédito não padronizada	235
C.29 Gráfico de dispersão para as representações tridimensionais a partir da base Imagens não padronizada	237
C.30 Gráfico de dispersão para as representações tridimensionais a partir da base Iris não padronizada	239
C.31 Gráfico de dispersão para as representações tridimensionais a partir da base Sonar não padronizada	241
C.32 Gráfico de dispersão para as representações tridimensionais a partir da base Vogal não padronizada	243
C.33 Gráfico de dispersão para as representações tridimensionais a partir da base Triagem de Crédito padronizada	245
C.34 Gráfico de dispersão para as representações tridimensionais a partir da base Imagens padronizada	247
C.35 Gráfico de dispersão para as representações tridimensionais a partir da base Iris padronizada	249
C.36 Gráfico de dispersão para as representações tridimensionais a partir da base Sonar padronizada	251

C.37 Gráfico de dispersão para as representações tridimensionais a partir da base Vogal padronizada	253
--	-----

Lista de Tabelas

3.1	Bases de dados	57
3.2	Execução dos Métodos por Cenário de Experimentos	59
3.3	Acurácia da classificação da matriz de dados original e padronizada por base de dados	65
3.4	Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de X da base Triagem de Crédito	65
3.5	Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de X da base Imagens	66
3.6	Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de X da base Iris	67
3.7	Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de X da base Sonar	67
3.8	Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de X da base Espiral	69
3.9	Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de X da base Vogal	70
3.10	Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de Z da base Triagem de Crédito	71
3.11	Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de Z da base Imagens	72
3.12	Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de Z da base Iris	73
3.13	Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de Z da base Sonar	73
3.14	Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de Z da base Espiral	75
3.15	Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de Z da base Vogal	76
4.1	Complexidade dos métodos	77
A.1	Métricas de qualidade da representação e tempo de execução	102

Lista de Abreviações

Ac	Acumulada
Aprox	Aproximação
CMDS	<i>Classical Multidimensional Scaling</i>
Dev. P.	Desvio Padrão
DR	<i>Dimension/Dimensionality Reduction</i>
E	Distância Euclidiana
Eq.	Equação
Euc	Distância Euclidiana
Fig.	Figura
KDD	<i>Knowledge Discovery in/from Databases</i>
M	Distância de Manhattan
MDS	<i>Multidimensional Scaling</i>
Man	Distância de Manhattan
PC	<i>Principal Component(s)</i>
PCA	<i>Principal Component Analysis</i>
Prop	Proporção
SVD	<i>Singular Value Decomposition</i>
Tab.	Tabela
UFJF	Universidade Federal de Juiz de Fora
V.A.	Variável(eis) Aleatória(s)
Var	Variância

1 Introdução

Na área de descoberta de conhecimento em bancos/base de dados (*Knowledge Discovery in/from* – KDD), um importante tema de estudo são os métodos de redução de dimensionalidade (*Dimension/Dimensionality Reduction* – DR) de dados multivariados. Os quais são analisados comparativamente no decorrer deste trabalho.

1.1 CONTEXTUALIZAÇÃO

1.1.1 Variáveis Aleatórias, Vetores Aleatórios e Matriz de Dados

Um conjunto de dados (*Dataset*) multivariados, \mathbf{X} , é definido por diversas observações/itens/medidas de um vetor aleatório, X . Este, por sua vez, é um vetor de variáveis/atributos/características aleatórias/randômicas (V.A.).

Seja o vetor aleatório X (Eq. 1.1) abaixo:

$$X^T = X_1, X_2, \dots, X_k, \dots, X_p \quad (1.1)$$

Onde $X_1, X_2, \dots, X_k, \dots, X_p$ são variáveis aleatórias.

Os dados são ditos multivariados, quando o vetor aleatório é composto por múltiplas variáveis aleatórias ($p > 1$). Normalmente quando X é composto por uma única V.A., é chamado de monovariado e quando composto por duas V.A. é bivariado.

Dados bivariados são multivariados, pois p maior que um. Os métodos estatísticos (e computacionais) de análise e manipulação de dados multivariados são igualmente válidos para bivariados. Estes também podem ser estudados por uma abordagem específica e simplificada, não sendo esta utilizada neste trabalho, visto que o mesmo visa um estudo de redução de dimensionalidade de dados multivariados para qualquer valor de p .

Fazendo n observações das p variáveis aleatórias, ou seja, n medidas do vetor aleatório X , obtemos a matriz de dados \mathbf{X} (Eq. 1.2) abaixo:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{j1} & x_{j2} & \dots & x_{jp} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad (1.2)$$

1.1.2 Dimensão do Espaço de Características

A dimensão do espaço de características dos dados é o número de variáveis aleatórias observadas, ou seja, a dimensão do vetor aleatório. Portanto, para a base de dados \mathbf{X} tem-se que a dimensão do espaço de característica de seus dados (ou a dimensão da base de dados) é p .

1.1.3 Redução de Dimensionalidade

“Redução de dimensionalidade de dados refere-se ao processo de derivação de um conjunto de graus de liberdade que pode ser ajustado para reproduzir grande parte da variabilidade observada” (GERING, 2002). Ou seja, segundo ROWEIS (2000), redução de dimensionalidade é o processo de reduzir o número de variáveis aleatórias consideradas.

A redução de dimensionalidade pode ser realizada por duas abordagens: seleção de características e extração de características.

A primeira consiste em reduzir as V.A. para um subconjunto das V.A. originais. Ou seja, o novo espaço será um subespaço das características original retirando algumas dimensões e perdendo por completo as informações expressas por elas.

Normalmente, por esta abordagem, o problema de reduzir a dimensionalidade é tratado como um problema de otimização combinatória, que visa obter a melhor combinação das dimensões a serem mantidas, orientado por uma estratégia específica. Essas estratégias podem ser: *filter* (orientado pelo ganho de informação – *information gain*), ou *wrapper* (orientado pela precisão – *accuracy*).

A outra abordagem consiste em transformar o espaço de alta dimensão dos dados em um espaço com menos dimensões. Esta transformação pode ser linear ou não linear.

Um exemplo de transformação linear é o PCA (*Principal Component Analysis – Análise de Componentes Principais*) (JOHNSON, 2007), que é o método de redução de dimensionalidade mais tradicional, conhecido e difundido.

ZHANG (2004) apresenta um método de redução de dimensionalidade não linear através de alinhamento espaço tangente. E o método de redução de dimensionalidade não linear por incorporação localmente linear (ROWEIS, 2000).

Diferentemente da maioria dos métodos de redução de dimensionalidade, existem métodos que combinam redução de dimensionalidade e aprendizado não supervisionado (agrupamento – *clustering*), a fim de aprimorar a redução do espaço de características dos dados adaptativamente. Estes métodos são chamados de redução de dimensionalidade adaptativa (*Adaptive Dimension/Dimensionality Reduction - ADR*) e um exemplo é o exposto em (DING, 2002).

1.2 JUSTIFICATIVA

Atualmente o volume de dados e o número de variáveis avaliadas em problemas das mais diversas áreas crescem rapidamente. E cada vez mais aumenta a dificuldade para extrair informação e conhecimento destes dados. Sendo assim, a utilização de métodos computacionais para a manipulação de dados é cada vez mais necessária e frequente.

Consequentemente, diversos novos desafios são postos à ciência da computação. Conforme exposto no documento da Sociedade Brasileira de Computação sobre os desafios da computação até 2016 (SBC, 2006), dois dos principais desafios definidos neste relatório são: “Gestão da Informação em grandes volumes de dados multimídia distribuídos” e “Modelagem computacional de sistemas complexos artificiais, naturais e sócio-culturais e da interação homem-natureza”.

Portanto, “as pesquisas relativas a este processo [KDD] ganharam rápido crescimento a partir da última década” (FAYYAD, 1996b) *apud* (MOTTA, 2010).

Em relação ao primeiro desafio, SBC (2006) define que o “objetivo deste desafio é (...) desenvolver soluções para o tratamento, a recuperação e a disseminação de informação relevante (...) a partir de volumes exponencialmente crescentes de dados (...). E ressalta os “problemas técnicos e científicos a abordar para fazer frente a estes desafios”,

dos quais a redução de dimensionalidade pode auxiliar a solucionar três: a) “Redução (abstração e sumarização) das massas de dados (...); b) “Estudos em modelos e mecanismos de conciliação e integração de dados altamente heterogêneos”; e c) “Estudo de formas alternativas de disponibilização da informação (...).”(SBC, 2006)

Por sua vez, o segundo desafio, a Ciência Computacional (*Computational Science*) é definido como:

“(...) modelos, algoritmos e ferramentas computacionais para solução de sistemas complexos de diferentes naturezas. Isto (...) possibilita que cientistas em outros domínios do conhecimento possam investigar problemas que até recentemente não poderiam ser tratados – pelo excesso de dados manipulados, (...). Exemplos são estudos de genômica de processos bioquímicos, da física de partículas, [de astronomia, das engenharias, da estatística,] de mudanças climáticas de fenômenos epidemiológicos ou de interações sociais com milhões de participantes (...). O objetivo deste desafio é criar, avaliar, modificar, compor, gerenciar e explorar modelos computacionais para todos esses domínios e aplicações. (...) A complexidade deste tipo de pesquisa aumenta à medida que crescem o volume de dados e/ou os parâmetros a serem considerados.” (SBC, 2006)

O documento ainda afirma que “a modelagem computacional envolve (...) técnicas de simulação, manipulação de dados, mineração de dados (...)" e que:

Os mesmos fatores que “(...) impulsionam a pesquisa em modelagem computacional de sistemas complexos, também trazem problemas para o seu desenvolvimento, pelo volume de dados gerados e pela complexidade crescente do processamento demandado. Por exemplo, o conjunto diário de dados recebidos de satélites e diferentes tipos de sensores já atinge petabytes (1015), exigindo pesquisa em técnicas para redução de dados.” (SBC, 2006)

O relatório afirma também que um dos principais desafios específicos relacionado com a modelagem computacional é o tratamento “dos problemas de aumento na dimensionalidade (número crescente de variáveis em um problema) e escalabilidade, que contribuem para tempos de processamento exponencial” (SBC, 2006).

Neste contexto percebe-se a importância direta do estudo de métodos para a realização de redução de dimensionalidades que minimizem a perda de informação (princi-

palmente de informação importante, podendo não se saber a priori quais são). Buscando facilitar a manipulação de grandes volumes de dados complexos por outros processos (tanto reduzindo o volume, quanto a complexidade), evidenciar as informações relevantes a métodos e técnicas de mineração de dados, aprendizado de máquina, e descoberta de conhecimento em geral (tanto computacional, quanto por humanos, ou ainda assistida por computador) e, também, possibilitar a visualização dos dados por humanos, e consequentemente sua análise.

Ainda permite que, com a redução de certa quantidade de características irrelevantes, as características relevantes dos dados sejam evidenciadas, facilitando a identificação das informações importantes (seja por humanos – visualização de dados –, ou algoritmos – de mineração de dados, por exemplo) colaborando com a geração de melhores modelos. E ao diminuir a dimensão do espaço de variáveis dos dados, também, diminui o custo computacional de manipulá-los.

Muitos dos métodos de redução de dimensionalidade, como o PCA (JOHNSON, 2007), para definir quais características dos dados são mais relevantes, calculam a interação/relação entre estas. Colaborando, desta forma, com a interpretação da base de dados, ao expressarem as estruturas de relacionamento de suas características.

1.3 OBJETIVOS

1.3.1 Objetivos Gerais

Realizar análise comparativa de diferentes métodos de redução de dimensionalidade de dados multivariados, tendo em vista a aplicação destes na etapa de pré-processamento de um processo de descoberta de conhecimento em banco de dados (KDD). Buscando uma melhoria na mineração de dados e/ou para possibilitar a visualização da distribuição dos dados por humanos.

1.3.2 Objetivos Específicos

- Levantamento de diferentes métodos de redução de dimensionalidade existentes;

- Seleção dos métodos levantados a serem estudados no trabalho;
- Descrever, exemplificar e definir a aplicação e necessidade da redução de dimensionalidade para um processo de KDD e para a visualização da distribuição dos dados por humanos;
- Estudar os métodos de redução de dimensionalidade quanto à minimização da perda de informação;
- Estudar os métodos de redução de dimensionalidade quanto à contribuição para uma melhor mineração de dados, em um processo de KDD;
- Estudar os métodos de redução de dimensionalidade quanto à contribuição para uma melhor análise exploratória dos dados;
- Descrever as características, vantagens e desvantagens de cada método estudado e definir os cenários mais adequados a cada um;
- Propor maneiras adequadas de utilização da redução de dimensionalidade em relação aos cenários;

1.4 METODOLOGIA

Com o direcionamento e auxílio dos professores Custódio Motta (orientador) e Lupércio Bessagato (Departamento de Estatística – UFJF), foram inicialmente estudados os fundamentos teóricos, as técnicas e os métodos estatísticos de análise multivariada de dados. Tais assuntos formam a base da área de redução de dimensionalidade.

Concomitantemente, foram levantados da literatura diversos métodos de redução de dimensionalidade, dentre os quais, alguns foram selecionados para serem mais profundamente estudados neste trabalho.

Para esses métodos selecionados foi dado uma maior atenção em seus funcionamentos, bem como em suas bases teóricas estatísticas e computacionais, ressaltando os seguintes quesitos: capacidade de minimizar a perda de informação, capacidade de auxiliar na interpretação dos dados, computacionalmente ou por humanos, contribuição para

uma melhor mineração de dados, contribuição para uma melhor análise exploratória dos dados (tanto na entrada do processo de KDD, quanto na saída).

Com este estudo em mãos, definiu-se as vantagens e desvantagens de cada um dos métodos selecionados. Por fim, realizou-se a análise comparativa destes métodos de redução de dimensionalidade e definiu-se os cenário de uso e as maneiras de uso mais adequados a cada um.

1.5 ESTRUTURA DO TRABALHO

Neste primeiro capítulo, foram introduzidas as bases estatísticas da área e definidas as necessidades da redução de dimensionalidade para um processo de KDD e para a visualização da distribuição dos dados por humanos.

O segundo capítulo apresenta a fundamentação teórica do trabalho, complementando as bases estatísticas e explicando o processo de KDD e a participação dos métodos de redução de dimensionalidade neste processo.

No terceiro capítulo os diversos métodos levantados são apresentados e em seguida os estudos dos métodos selecionados são expostos, ressaltando as respectivas vantagens e desvantagens.

O quarto capítulo expõe a análise comparativa dos métodos e a definição dos cenários e maneiras de utilização mais adequadas para cada um. Concluindo, o quinto capítulo apresenta as considerações finais e possíveis trabalhos futuros.

2 Métodos

Nesta secção abordaremos os principais métodos de redução de dimensionalidade utilizados na etapa de pré-processamento ou para visualização dos dados em uma descoberta de conhecimento em banco de dados.

2.1 ANÁLISE DE COMPONENTES PRINCIPAIS

"A técnica dominante de transformação de características que transforma o dado para um espaço reduzido, o qual captura a maior parte da variância destes, é a Análise de Componentes Principais"(CUNNINGHAM, 2007), do inglês, *Principal Component Analysis* (PCA), "originalmente conhecida como Transformada de Karhunen-Loève"(MAATEN, 2009)e, também, chamada em outras áreas por Transformada de Hotelling e método da função ortogonal empírica (*Empirical Orthogonal Function*), segundo FODOR (2002).

Ainda, de acordo com SHLENS (2003), por ser um método simples e não paramétrico de extração de informação relevante de conjuntos confusos de dados, com mínimo pós-processamento pode ser utilizado para redução da dimensionalidade e ser aplicado em diversas áreas, da neurociência a computação gráfica, o PCA é o mais valioso resultado da álgebra linear aplicada. Portanto, será o primeiro método que abordaremos.

Uma análise de componentes principais é um método de estatística multivariada com o objetivo de "explicar a estrutura de variância-covariância de um conjunto de variáveis aleatórias através de poucas combinações lineares destas."(JOHNSON, 2007)

Do ponto de vista da álgebra linear, o PCA encontra uma nova base para o espaço de características. A qual é composta pelos vetores de coeficientes das combinações lineares das variáveis originais, sendo estes vetores ortogonais entre si. Estas combinações lineares são chamadas de componentes principais (*Principal Components* - PC). E são definidas de forma que a primeira componente principal (PC_1) maximize a variância (o seu vetor de coeficientes - a_1 - será o primeiro vetor da base), a segunda componente (PC_2) maximize a variabilidade restante (o seu vetor de coeficientes - a_2 - será o segundo

vetor da base), e assim por diante.

"Geometricamente, estas combinações lineares $[(PC_1, PC_2, \dots, PC_p)]$ representam a seleção de um novo sistema de coordenadas obtido por rotação do sistema com [as V.A.] X_1, X_2, \dots, X_p como eixos coordenados [(original)]"(JOHNSON, 2007). Sendo os vetores de coeficientes de cada componente (a_i) o vetor direcional do respectivo eixo coordenado.

Por exemplo, seja o conjunto de dados, \mathbf{X} , com dez amostras e duas variáveis aleatórias ($X^T = [X_1 X_2]$) definido abaixo:

$$\mathbf{X} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

Sabemos, da álgebra linear, que a base canônica para o espaço cartesiano de duas dimensões, e logo para o espaço das características em questão ($X^T = [X_1 X_2]$), é $B = (0; 1), (1; 0) = \hat{i}, \hat{j}$. Aplicando o PCA uma nova base para o mesmo espaço é obtida: $B_{PCA} = a_1, a_2 = (-0,70711; 0,70711), (-0,70711; -0,70711)$.¹ Onde os vetores são ortogonais entre si ($a_1^T \cdot a_2 = (-0,70711; 0,70711)_T \cdot (-0,70711; -0,70711) = 0$).

O PCA, também, obtém a quantidade de variabilidade representada por cada componente. Que para o exemplo é 100% para a primeira e 0% para a segunda. Nota-se que de fato a primeira componente maximizou a representação da variância (e que a segunda maximiza a variância restante, no caso 0%).

¹ Dependendo da forma como os cálculos do PCA foram realizados pode-se obter uma base equivalente, onde o vetor de coeficiente seja o oposto do obtido, no exemplo. Mas, da geometria, sabemos que a direção do eixo coordenado definida tanto pelo a_1 quanto $-a_1$ são as mesmas.

Entretanto, apesar de obtermos um novo espaço de características $X_{PCA}^T = [PC_1 \ PC_2]$, que é muito útil para interpretação da estrutura de variança-covariança e análise dos dados, a dimensão continua a mesma ($\text{Dim}(X) = 2 = \text{Dim}(X_{PCA})$). Observando a variabilidade explicada por cada componente é perceptível que não há necessidade de representar os dados de \mathbf{X} por duas V.A. (características). E que poderíamos reduzir a dimensão de X com uma perda mínima de informação representando os em um novo espaço formado apenas pela(s) componente(s) que captura(m) a maior parte da variância.

Para tanto, definimos um novo espaço de característica (X_{red}) de menor dimensão (no caso, $\text{Dim}(X_{red}) = 1 = k < p = 2 = \text{Dim}(X)$), sendo as k primeiras componentes principais os eixos deste novo espaço. Desta forma, para o exemplo, a base do novo espaço será: $B_{red} = v_1 = a_1 = (-0,70711; 0,70711)$. Ou seja, o novo espaço (vetor aleatório, vetor de características) é composto por apenas a primeira componente: $X_{red}^T = [PC_1]$.

Note que como utilizamos a componente que captura a maior parte da variabilidade dos dados, esta continuará a ser representada no novo espaço. No caso, a variância representada por PC_1 é de 100%, logo o novo espaço continuará a representar toda a variabilidade dos dados (ou seja, não perde, praticamente, qualquer informação).

Com o novo espaço determinado é hora de obter a matriz de Escore ($Score$) dos dados, isto é, transformar a matriz de dados original (\mathbf{X}) para a matriz dos dados no novo espaço ($\mathbf{X}_{red} = Score$).² Segue a matriz obtida para o exemplo:

²O mesmo procedimento pode ser utilizado para gerar os escores, a matriz de dados, para o espaço definido utilizando todas as componentes principais (X_{PCA}), antes da redução. Isto é muito utilizado na estatística multivariada para interpretação destes.

$$\mathbf{X}_{\text{red}} = \text{Score}_{X_{\text{red}}}(\mathbf{X}) = \begin{bmatrix} 0.70711 \\ 0.70711 \\ -0.70711 \\ -0.70711 \\ -0.70711 \\ 0.70711 \\ -0.70711 \\ -0.70711 \\ 0.70711 \\ 0.70711 \end{bmatrix}$$

Perceba que ficou mais simples identificar os dois agrupamentos das amostras, o dos positivos (onde nos dados originais, primeira coluna era 0 e segunda 1) e o dos negativos (onde nos dados originais, primeira coluna era 1 e segunda 0). Isto ilustra o uso da redução de dimensionalidade como ferramente de pré-processamento de dados em um processo de KDD, pois, da mesma forma que facilitou a tarefa de agrupar os dados manualmente, provavelmente, também, tornaria mais simples uma mineração destes dados.

Além de a representação na dimensão permitir uma melhora na acurácia e na identificação de padrões/informações mais interessantes, ao ressaltar as relações mais importantes entre os dados. Ela ainda pode melhorar o desempenho computacional da técnica de mineração por diminuir a quantidade de dados a serem manipulados e, também, na economia de espaço de armazenamento.

Este foi um exemplo didático com uma relação muito simples entre os dados. Mas ele serve para mostrar os objetivos, o funcionamento e os resultados obtidos com o uso do método PCA e consequentemente o motivo de seu uso (tal qual de outros métodos de redução de dimensionalidade). Pois em bases de dados enormes, até mesmo esta simples relação de negação entre as V.A. pode ser difícil, nem que em custo de tempo de processamento, de ser identificada.

Imagine fazer este agrupamento manualmente de uma matriz de dados com mil

amostras e cem atributos com a mesma relação de negação entre si. Com certeza este padrão não nos apareceria simplesmente e demandaria muito mais tempo de análise dos dados para extraímos alguma informação relevante.

Diferentemente do exemplo, onde conseguimos reduzir a dimensão com perda de 0% da representação da variabilidade dos dados, em casos reais, normalmente, isto não é possível. Sendo necessário utilizar todas as componentes para se reproduzir a variabilidade do sistema por completo.

Entretanto, "frequentemente a maior parte da variabilidade pode ser explicada [(representada, capturada)] por um pequeno número k de componentes principais" (JOHNSON, 2007). Logo, frequentemente é possível reduzir a dimensão do espaço de características de p para k com perdas mínimas de informação.

2.1.1 Definição Matemática

Tradicionalmente na estatística a análise de componentes principais é definida populacionalmente para posteriormente se derivar sua definição amostral. Entretanto, abordaremos este e os demais métodos diretamente do ponto de vista amostral. Pois, este trabalho trata da aplicação destes em um processo de descoberta de conhecimento em base de dados, onde se pretende extraír (minerar, aprender) padrões relevantes (informações interessantes) a partir da análise dos dados amostrais da entrada. Para as definições segundo a abordagem populacional e a derivação da amostral sugere-se consultar JOHNSON (2007).

Dada uma matriz de dados \mathbf{X} com n amostras, observações $(x_i; i = 1, 2, n)$, com dimensão p , ou seja, $x_i^T = [x_{i1} x_{i2} \cdots x_{ip}]; i = 1, 2, n$. É X , seu respectivo espaço de características, vetor de variáveis aleatórias (ou ainda, apenas, vetor aleatório) com p variáveis aleatórias, V.A., X_1, X_2, \dots, X_p (Eq. 2.1),

$$X^T = \begin{bmatrix} X_1 & X_2 & \dots & X_p \end{bmatrix} \quad (2.1)$$

E S a sua matriz de variância-covariância amostral não viciada³.

³Escolhemos a matriz de variância-covariância amostral não viciada ($S = (n - 1)^{-1} \mathbf{X}_{\text{centralizado}}^T \mathbf{X}_{\text{centralizado}}$) ao invés da viciada ($S_n = n^{-1} \mathbf{X}_{\text{centralizado}}^T \mathbf{X}_{\text{centralizado}}$) conforme sugerido em JOHNSON (2007) e utilizado na bibliografia de área de redução de dimensionalidade,

Consideremos as combinações lineares em 2.2.

$$\begin{aligned}
 Y_1 &= a_1^T X = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\
 Y_2 &= a_2^T X = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \\
 &\vdots \\
 Y_p &= a_p^T X = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p
 \end{aligned} \tag{2.2}$$

Onde, $a_i^T = [a_{i1} \ a_{i2} \ \cdots \ a_{ip}]$ é o vetor de coeficientes da combinação linear Y_i .

Segundo (JOHNSON, 2007), obtemos a variância (Eq. 2.3) e a covariância (Eq. 2.4) das combinações lineares de (Eq. 2.2).

$$Var(Y_i) = a_i^T S a_i \quad i = 1, 2, \dots, p \tag{2.3}$$

$$Cov(Y_i, Y_j) = a_i^T S a_j \quad i, j = 1, 2, \dots, p \tag{2.4}$$

As componentes principais são, então, definidas como as combinações lineares ortogonais, tal que a variância (Eq. 2.3) seja tão grande possível. Onde a primeira componente principal será a com maior variância.

Entretanto, perceba, que $Var(Y_i) = a_i^T S a_i$ pode ser aumentado indefinidamente multiplicando-se o vetor de coeficientes a_i por uma constante. A fim de eliminar esta indeterminação, adicionamos a restrição de que o vetor de coeficientes deve ser unitário.

Portanto, definindo o método PCA com uma otimização, temos Eq. 2.5 (JOHNSON, 2007):

mineração de dados e KDD (HAN, 2006) (TAN, 2005)e implementações dos métodos (MATHWORKS, 2012)(SHLENS, 2003)(NG, 2013)

$$\begin{aligned}
\text{Primeira componente} & \quad \text{combinação linear } Y_1 = a_1^T X \\
\text{principal } (PC_1) & = \text{que maximiza } Var(Y_1) = Var(a_1^T X) \\
& \text{s.a. } a_1^T a_1 = 1 \\
\text{Segunda componente} & \quad \text{combinação linear } Y_2 = a_2^T X \\
\text{principal } (PC_2) & = \text{que maximiza } Var(Y_2) = Var(a_2^T X) \\
& \text{s.a. } a_2^T a_1 = 1 \\
& \quad Cov(Y_1, Y_2) = Cov(a_1^T X, a_2^T X) = 0 \\
& \quad \vdots \\
\text{i-ésima componente} & \quad \text{combinação linear } Y_i = a_i^T X \\
\text{principal } (PC_i) & = \text{que maximiza } Var(Y_i) = Var(a_i^T X) \\
& \text{s.a. } a_i^T a_i = 1 \\
& \quad Cov(Y_j, Y_i) = Cov(a_j^T X, a_i^T X) = 0 ; \\
& \quad \forall j < i \\
& \quad \vdots \\
\text{Última (p-ésima) componente} & \quad \text{combinação linear } Y_p = a_p^T X \\
\text{principal } (PC_p) & = \text{que maximiza } Var(Y_p) = Var(a_p^T X) \\
& \text{s.a. } a_p^T a_p = 1 \\
& \quad Cov(Y_j, Y_p) = Cov(a_j^T X, a_p^T X) = 0 ; \\
& \quad \forall j < i
\end{aligned} \tag{2.5}$$

E consequentemente as notações PC_i e Y_i são equivalentes para denotar a combinação linear que é a i-ésima componente principal.

Nota-se que o PCA pode ser realizado por diferentes técnicas que encontrem a matriz de vetores de coeficientes ($A = [a_1 a_2 \cdots a_p]$) resolvam o problema de otimização em Eq. 2.5 dêis de as mais simples até outras mais complexas (mas menos utilizadas, e portanto, além do escopo deste trabalho), como uso de redes neurais artificiais. Também existe uma formulação não linear para o PCA, sendo o Kernel PCA o seu método de solução mais conhecido (MAATEN (2009)apresenta mais informações sobre está técnica).

2.1.2 Obtenção Das Componentes Principais

O problema da obtenção das componentes principais é tradicionalmente solucionado através de ferramentas da álgebra linear e propriedades da estatística multivariada. A pesar de ser definido como uma maximização podemos utilizar de certos resultados analíticos a partir das áreas citadas sua resolução.

A seguir derivaremos esta forma de obter as componentes (conforme JOHNSON (2007)).

Sabe-se que "as matrizes de variâncias-covariâncias (ou de correlações) são sempre matrizes semi-definidas positivas, ou seja, matrizes simétricas ($M_{p \times p}$) cujas formas quadráticas ($x^T M x$) são sempre não-negativas."(CADIMA, [20__]). E que maximização em um espaço unitário da forma quadrática de uma matriz M semi-definida qualquer, isto é, com autovalores não negativos ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$) é igual ao seu maior autovalor (λ_1):

$$\max_{x \neq 0} \frac{x^T M x}{x^T x} = \lambda_1$$

Sendo que o valor de x que soluciona esta maximização é o autovetor unitário associado a este maior autovalor, ou seja, e_1 . Como os autovetores são perpendiculares (ortogonais) entre si, podemos definir genericamente que

$$\max_{x \neq 0, x \perp e_1, \dots, e_i} \frac{x^T M x}{x^T x} = \lambda_{i+1} ; i = 0, 1, 2, \dots, p-1 \quad (2.6)$$

É atingido quando $x = e_{i+1}; i = 1, 2, \dots, p-1$ respectivamente, onde p é a dimensão da matriz M .⁴. Ou seja, $x = e_1$ soluciona:

$$\max_{x \neq 0} \frac{x^T M x}{x^T x} = \lambda_1$$

Em seguida, a solução é $x = e_2$ (e o valor máximo λ_2):

$$\max_{x \neq 0, x \perp e_1} \frac{x^T M x}{x^T x} = \lambda_2$$

⁴A prova pode ser obtida em JOHNSON (2007)

E assim por diante até:

$$\max_{a \neq 0, a \perp e_1, \dots, e_{p-1}} \frac{x^T M x}{x^T x} = \lambda_p$$

Onde $x = e_p$ e o valor máximo é λ_p , o menor autovalor associado a M .

Portanto, seja S a matriz de variância-covariância associada com o vetor aleatório X . E $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$ os pares de autovalores-autovetores de S . Onde e_i são autovetores normalizados e $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Então, da Eq.2.6, com $M = S$ e $x = a$, temos:

$$\max_{a \neq 0, a \perp e_1, \dots, e_i} \frac{a^T S a}{a^T a} = \lambda_{i+1}; i = 0, 1, 2, \dots, p-1 \quad (2.7)$$

Onde, o a que soluciona a maximização é $a = e_{i+1}$. Ou seja:

$$\max_{a \neq 0, a \perp e_1, \dots, e_i} \frac{a^T S a}{a^T a} = \lambda_{i+1} = \frac{e_{i+1}^T S e_{i+1}}{e_{i+1}^T e_{i+1}}; i = 0, 1, 2, \dots, p-1$$

Entretanto, os autovetores são vetores unitários, logo, $e_j^T e_j = 1; j = 1, 2, \dots, p$.

Consequentemente:

$$\max_{a \neq 0, a \perp e_1, \dots, e_i} \frac{a^T S a}{a^T a} = \lambda_{i+1} = \frac{e_{i+1}^T S e_{i+1}}{e_{i+1}^T e_{i+1}} = e_{i+1}^T S e_{i+1}; i = 0, 1, 2, \dots, p-1$$

Da Eq. 2.3 temos que a variância de uma componente principal (lembrando $Y_j = PC_j$) é exatamente $Var(PC_j) = Var(Y_j) = e_j^T S e_j$. Portanto:

$$\begin{aligned} \max_{a \neq 0, a \perp e_1, \dots, e_i} \frac{a^T S a}{a^T a} &= \lambda_{i+1} = \frac{e_{i+1}^T S e_{i+1}}{e_{i+1}^T e_{i+1}} = e_{i+1}^T S e_{i+1} = Var(Y_{i+1}) = \\ &= Var(PC_{i+1}); i = 0, 1, 2, \dots, p-1 \end{aligned}$$

Ou seja,

$$Var(PC_j) = Var(Y_j) = e_j^T S e_j = \frac{e_j^T S e_j}{e_j^T e_j} = \lambda_j; j = 1, 2, \dots, p \quad (2.8)$$

Ainda, como os autovetores são perpendiculares entre si ($e_k^T \perp e_j = e_k^T e_j = 0; k \neq j$) e $Se_j = \lambda e_j$, temos:

$$Var(PC_k, PC_j) = Var(Y_k, Y_j) = e_k^T Se_j = e_k^T \lambda_j e_j = \lambda_j e_k^T e_j = 0 ; j = 1, 2, \dots, p \quad (2.9)$$

E como o a_i que soluciona a i -ésima maximização (Eq.2.7) é igual ao respectivo autovetor ($a_i = e_i$), temos Eq.2.10:

$$\begin{aligned} PC_1 &= Y_1 = a_1^T X = e_1^T X = e_{11}X_1 + \dots + e_{1p}X_p \\ PC_2 &= Y_2 = a_2^T X = e_2^T X = e_{21}X_1 + \dots + e_{2p}X_p \\ &\vdots \\ PC_p &= Y_p = a_p^T X = e_p^T X = e_{p1}X_1 + \dots + e_{pp}X_p \end{aligned} \quad (2.10)$$

Onde o novo espaço de características $PC_T = [PC_1 \ PC_2 \ \dots \ PC_p]$ tem a mesma variância total que o espaço original. Conforme demonstrado a seguir:

Cada elemento da matriz de variância-covariância associado a um vetor aleatório $M_T = [M_1 M_2 \ \dots \ M_p]$ qualquer é definido por $Cov(M_i, M_j)$, logo os elementos da sua diagonal são: $Cov(M_i, M_i) = Var(M_i)$. Portanto temos que a variância total de M , ou seja, o somatório das variâncias de suas p V.A., é o traço da sua respectiva matriz de variância-covariância.

Para o caso do espaço de características original (X com matriz de variância-covariância S) temos:

$$Var_{tot}(X) = \sum_{i=1}^p Var(X_i) = tr(S)$$

Como vimos, realizando a decomposição espectral de S obtivemos os autovalores $\lambda_1, \lambda_2, \dots, \lambda_p$ e os autovetores $e_1 e_2 \ \dots \ e_p$. Esta decomposição também pode ser descrita da forma matricial: $S = P \Lambda P_T$, onde Λ é a matriz diagonal dos autovetores e $P = [e_1 e_2 \ \dots \ e_p]$ (observe que como os autovetores são os coeficientes das componentes lineares $A = [a_1 a_2 \ \dots \ a_p] = [e_1 e_2 \ \dots \ e_p] = P$). Desta forma, $PP_T = P_T P = I$, sendo I a matriz identidade. Portanto:

$$\text{tr}(S) = \text{tr}(P\Lambda P_T) = \text{tr}(\Lambda P_T P) = \text{tr}(\Lambda I) = \text{tr}(\Lambda) = \lambda_1 + \lambda_2 + \cdots + \lambda_p$$

Da Eq. 2.8 temos que $\text{Var}_i = \lambda_i$, logo

$$\begin{aligned} \sum_{i=1}^p \text{Var}(X_i) &= \text{tr}(S) = \text{tr}(\Lambda) = \lambda_1 + \lambda_2 + \cdots + \lambda_p = \\ &= \text{Var}(Y_1) + \text{Var}(Y_2) + \cdots + \text{Var}(Y_p) = \sum_{i=1}^p \text{Var}(Y_i) \end{aligned}$$

Ou seja,

$$\sum_{i=1}^p \text{Var}(X_i) = \sum_{i=1}^p \text{Var}(Y_i) = \sum_{i=1}^p \text{Var}(PC_i) \quad (2.11)$$

A partir desta demonstração podemos derivar uma fórmula para o cálculo da proporção da variabilidade (variância total) explicada por cada componente principal (Eq.2.12), informação essencial para a redução da dimensionalidade.

$$\begin{pmatrix} \text{Proporção da} \\ \text{variância total} \\ \text{segundo } PC_i \end{pmatrix} = \frac{\text{Var}(PC_i)}{\text{Var}_{tot}(PC)} = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \quad (2.12)$$

Consequentemente a proporção acumulada da variabilidade representada pelas k primeiras componentes (tal que $k \leq p$) é dado pela Eq.2.13 abaixo:

$$\begin{pmatrix} \text{Proporção acumulada da} \\ \text{variância total segundo} \\ \text{as } k \text{ primeiras PC} \end{pmatrix} = \sum_{i=1}^k \frac{\text{Var}(PC_i)}{\text{Var}_{tot}(PC)} = \sum_{i=1}^k \frac{\lambda_i}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} ; k \leq p \quad (2.13)$$

Desta forma, temos um algoritmo (Alg. 1) para obtenção das componentes principais associadas a um espaço de características X a partir de uma matriz de dados \mathbf{X} com n amostras neste espaço.

Onde a matriz de variância-covariância amostral não viciada, S , pode ser obtida

Algoritmo 1: Obtenção das componentes principais pela decomposição espectral da matriz de variância-covariância.

Entrada: matriz de dados (\mathbf{X}).

Saída: matriz de coeficientes das componentes principais (A); vetor das variâncias explicadas por cada componente principal (λ).

início

Obtenha a matriz de variância-covariância S associada a X ; Realize a decomposição espectral da matriz S ; Ordene os pares autovalor-autovetor (λ_i, e_i) , tal que, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$; $A \leftarrow [e_1 \ e_2 \ \dots \ e_p]$; $\lambda \leftarrow [\lambda_1 \ \lambda_2 \ \dots \ \lambda_p]^T$;

fim

segundo o algoritmo Alg. 2 (ou pelo uso de bibliotecas de estatística⁵).

Algoritmo 2: Calcula a matriz de variância-covariância amostral não viciada.

Entrada: matriz de dados (\mathbf{X}).

Saída: matriz de variância-covariância amostral não viciada (S).

início

Obtenha a matriz de dados centralizada \mathbf{C} ; $S \leftarrow \frac{1}{n-1} \mathbf{C}^T \mathbf{C}$;

fim

Por sua vez a matriz de dados centralizada ($\mathbf{C} = [c_1 c_2 \dots c_n]^T$) é obtida subtraindo de cada amostra (x_i) o vetor de médias amostrais ($\bar{x} = [\bar{x}_1 \bar{x}_2 \dots \bar{x}_p]^T$; \bar{x}_i é a média amostral da i-ésima V.A., X_i). Conforme ilustrada no Alg. 3.

Algoritmo 3: Centraliza a matriz de dados.

Entrada: matriz de dados (\mathbf{X}).

Saída: matriz de dados centralizada (\mathbf{C}).

início

Calcule o vetor de médias amostrais $\bar{x} = [\bar{x}_1 \ \bar{x}_2 \ \dots \ \bar{x}_p]^T$; $\mathbf{C} \leftarrow [x_1 - \bar{x} \ x_2 - \bar{x} \ \dots \ x_n - \bar{x}]^T$;
--

fim

⁵Como a função $cov(X)$ da biblioteca *statistics* da linguagem de programação MatLab (MATHWORKS, 2012)

2.1.3 Obtenção Das Componentes Principais Para Dados Padronizados

Entretanto a solução apresentada está sujeita a influência da diferenças das unidades e escalas das características. Podendo ocorrer a dominância de algumas sobre as outras, não por suas relevância para o sistema, mas pela discrepância entre as respectivas escalas. Sendo este um cenário comum e problemático em um processo de KDD.

A fim de solucionar esta dependência as ações mais tradicionais utilizados no processo são realizar a normalização ou a padronização dos dados, conforme HAN (2006) e TAN (2005). Especificamente a padronização tem uma relação com os cálculos do PCA que é bastante útil.

Inicialmente aplicaremos o método tradicional (utilizando a decomposição espectral da matriz de variância-covariância) ao espaço padronizado e posteriormente utilizaremos de suas propriedades para obter uma outra forma de solucionar o PCA para este cenário.

Seja $Z_T = [Z_1 Z_2 \cdots Z_p]$ o vetor característico obtido da padronização do espaço original ($X_T = [X_1 X_2 \cdots X_p]$) e $\mathbf{Z} = [z_1 z_2 \cdots z_n]^T$ a matriz de dados padronizada, tal que:

$$z_{ij} = \frac{(x_{ij} - \bar{x}_j)}{\sqrt{s_{jj}}}; i = 1, 2, \dots, n; j = 1, 2, \dots, p$$

Onde, z_{ij} é a coordenada na dimensão da j-ésima V.A. do espaço padronizado (Z_j) para a i-ésima amostra da matriz de dados padronizada (z_i). Da mesma forma, x_{ij} é a coordenada na dimensão da j-ésima V.A. do espaço original (X_j) para a i-ésima amostra da matriz de dados original (x_i). E \bar{x}_j é a média amostral da j-ésima variável aleatória do espaço original (X_j), a qual, também, tem variância amostral $s_{jj} = Cov(X_j, X_j) = Var(X_j) = s_j^2$.

O Alg. 4 realiza a padronização dos dados de uma matriz de dados de entrada:

Sendo que a centralização dos dados pode ser feita utilizando-se do Alg. 3.

Após a padronização, podemos, então, utilizar a técnica desenvolvida na Secção 2.1.2 para se obter as componentes principais em relação ao espaço padronizado. Para tanto, basta informar a matriz de dados padronizada, \mathbf{Z} ao invés da matriz de dados

Algoritmo 4: Padroniza os dados

Entrada: matriz de dados (\mathbf{X}).
Saída: matriz de dados padronizada (\mathbf{Z}).
início

fim	Obtenha a matriz de dados centralizada \mathbf{C} ; Calcule o vetor de desvios padrões amostrais $s^{1/2} = [s_1^{1/2} \ s_2^{1/2} \ \dots \ s_p^{1/2}]^T$; $\mathbf{Z} \leftarrow [c_1 - s^{1/2} \ c_2 - s^{1/2} \ \dots \ c_n - s^{1/2}]^T$;
------------	--

original, X , como entrada para o Alg. 2.

Entretanto, este processo pode ser realizado de uma outra forma, aproveitando-se da uma relação entre a matriz de variância-covariância do espaço padronizado e da matriz de correlação dos dados originais.

JOHNSON (2007) demonstra que a matriz de variância-covariância amostral do espaço padronizado Z é igual a matriz de correlação amostral (R) do espaço original X , isto é, $S_Z = R_X$. Portanto, a partir deste resultado, podemos definir uma nova forma de solucionar o método do PCA.

Ao invés de se padronizar o espaço para então calcularmos sua matriz de variância-covariância e então realizar a decomposição espectral desta para definição das respectivas componentes principais. Podemos obter as componentes principais associadas ao espaço padronizado diretamente da matriz de dados original, conforme descrito no Alg. 5.

Algoritmo 5: Obtenção das componentes principais associadas ao espaço padronizado pela decomposição espectral da matriz de correlação dos dados originais

Entrada: matriz de dados original (\mathbf{X}).
Saída: matriz de coeficientes das componentes principais associadas ao espaço padronizado (A); vetor das variâncias explicadas por cada componente principal associada ao espaço padronizado (λ).
início

fim	Obtenha a matriz de correlação R associada a X ; Realize a decomposição espectral da matriz R ; Ordene os pares autovalor-autovetor ((λ_i, e_i)), tal que, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$; $A \leftarrow [e_1 \ e_2 \ \dots \ e_p]$; $\lambda \leftarrow [\lambda_1 \ \lambda_2 \ \dots \ \lambda_p]^T$;
------------	---

Onde, a matriz de correlação é definida matricialmente por $R = (D^{1/2})^{-1}S(D^{1/2})^{-1} =$

$D^{-1/2}SD^{-1/2}$, tal que, $D^{1/2}$ é a matriz de desvios padrões amostrais (uma matriz diagonal, composta pelos elementos do vetor de desvios padrões amostrais, $s^{1/2}$).

Da mesma forma que demonstramos que a variância explicada pela i -ésima componente principal é o autovalor associado ao autovetor que por sua vez é o vetor de coeficientes da componente, ou seja, $Var(PC_i) = \lambda_1$, pode-se derivar é obtidas pela decomposição espectral de S_X ou R_X ⁶

2.1.4 Obtenção Das Componentes Principais Através Da Decomposição Em Valor Singular

Entretendo, a multiplicação $\mathbf{X}^T\mathbf{X}$, utilizada no cálculo da matriz de variância-covariância, esta sujeita a um erro numérico quando alguns elementos da matiz de dados são muito pequenos.⁷ Então, para evitar-lo, pode-se solucionar o PCA utilizando-se da decomposição em valor singular, do inglês, *singular value decomposition* (SVD), conforme verificaremos nesta secção.

A decomposição em valor singular de uma matriz $M_{n \times p}$ qualquer, segundo a álgebra linear é dada pela Eq. 2.14.

$$M = U\Gamma V^T \quad (2.14)$$

Onde, Γ ⁸ é a matriz diagonal retangular ($n \times p$), e os elementos de sua diagonal (s_{ii}) são não negativos; U é a matriz ($n \times n$) dos vetores singulares pela esquerda ($U = [u_1 u_2 \cdots u_n]$); e V é a matriz ($p \times p$) dos vetores singulares pela direita ($V = [v_1 v_2 \cdots v_p]$). Tanto as colunas de U (os vetores singulares pela esquerda), quanto as de V (vetores singulares pela direita), são ortonormais entre si, portanto, $U^T U = I$ e $L^T L = I$.

Multiplicando 2.14 por M^T a esquerda obtemos a decomposição em valor singular de $M^T M$ (Eq.2.15):

⁶JOHNSON (2007)apresenta a demonstração detalhada.

⁷Um exemplo de matriz que ocasiona este erro é a matriz de Läuchli (LÄUCHLI, 1961).

⁸No contexto de SVD, normalmente se utiliza S para se representar a matriz Γ , entretanto nesta secção utilizaremos esta notação para não ocasionar confusão com a matriz de variância-covariância amostral, S .

$$M^T M = (U\Gamma V^T)^T (U\Gamma V^T) = (V\Gamma^T U^T)(U\Gamma V^T) = V\Gamma^T U^T U S V^T = V\Gamma^T I \Gamma V^T = V(\Gamma^T \Gamma)V^T$$
(2.15)

Note que a decomposição em vetor singular de $M^T M = V W V^T; W = \Gamma^T \Gamma$, também é sua decomposição espectral. Onde, V é a matriz de autovetores associada a $M^T M$ e $\Gamma^T \Gamma$ é a matriz diagonal dos autovalores associados.

Isto é, $V = [v_1 v_2 \cdots v_p] = [e_1 e_2 \cdots e_p] = P$ e $W = \Gamma^T \Gamma = \Lambda$, ou seja, o i-ésimo elemento da diagonal de Γ é igual a raiz quadrada do i-ésimo elemento da diagonal de λ ($diagonal(\Gamma^T \Gamma) = [\gamma_1 \gamma_1 \gamma_2 \gamma_2 \cdots \gamma_p \gamma_p] = [\gamma_1^2 \gamma_2^2 \cdots \gamma_p^2] = [\lambda_1 \lambda_2 \cdots \lambda_p] = \Lambda$; $digonal(\Gamma) = [\gamma_1 \gamma_2 \cdots \gamma_p]$).

Lembrando que por convenção os autovalores são ordenados da forma $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ (e P é ordenada para que e_i seja o autovetor associado a λ_i). Da mesma forma, utilizamos esta convenção para os valores singulares: $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_p$ (e, consequentemente, V é ordenada para que v_i seja o autovetor associado a γ_i).

Retomando a matriz de dados \mathbf{X} com espaço característico $X^T = [X_1 X_2 \cdots X_p]$ temos que a sua matriz de variância-covariância amostral não viciada (S) é:

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix}; s_{ij} = s_{ji}; i, j = 1, 2, \dots, p$$
(2.16)

Onde, a s_{ij} é a covariância não viciada entre X_i e X_j ($Cov(X_i, X_j)$), sendo $s_{ii} = Cov(X_i, X_i) = Var(X_i) = s_i^2$. O qual é determinado conforme segue:

$$s_{ij} = \frac{1}{n-1} \sum_{l=0}^n [(x_{li} \bar{x}_i)(x_{lj} \bar{x}_j)]$$

Sendo, \bar{x}_i a média amostral da i-ésima V.A. (X_i). Então podemos reescrever a Eq.2.16 na forma matricial (Eq. 2.17).

$$\begin{aligned}
S &= \frac{1}{n-1} \begin{bmatrix} \sum_{l=0}^n (x_{l1}\bar{x}_1)^2 & \sum_{l=0}^n [(x_{l1}\bar{x}_1)(x_{l2}\bar{x}_2)] & \cdots & \sum_{l=0}^n [(x_{l1}\bar{x}_1)(x_{lp}\bar{x}_p)] \\ \sum_{l=0}^n [(x_{l2}\bar{x}_2)(x_{l1}\bar{x}_1)] & \sum_{l=0}^n (x_{l2}\bar{x}_2)^2 & \cdots & \sum_{l=0}^n [(x_{l2}\bar{x}_2)(x_{lp}\bar{x}_p)] \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{l=0}^n [(x_{lp}\bar{x}_p)(x_{l1}\bar{x}_1)] & \sum_{l=0}^n [(x_{lp}\bar{x}_p)(x_{l2}\bar{x}_2)] & \cdots & \sum_{l=0}^n (x_{lp}\bar{x}_p)^2 \end{bmatrix} = \\
&= \frac{1}{n-1} \mathbf{C}^T \mathbf{C}
\end{aligned} \tag{2.17}$$

Tal que, \mathbf{C} é a matriz de dados centralizada (ou seja, $c_{ij} = x_{ij} - \bar{x}_j$; $i, j = 1, 2, \dots, p$).

Vimos na Secção 2.1.2 que as componentes principais são obtidas realizando a decomposição espectral da matriz de variância-covariância do espaço. Isto é o mesmo que a decomposição espectral de $Cov(\mathbf{X}) = S = \frac{1}{n-1} \mathbf{C}^T \mathbf{C}$. Ou seja, $P \Lambda P = S = \frac{1}{n-1} \mathbf{C}^T \mathbf{C} = \Gamma^T \Gamma = V(\Gamma^T \Gamma) V^T$.

Anteriormente, derivamos a decomposição em valor singular de uma matriz M qualquer (Eq.2.14) de forma a obter a decomposição associada a $M^T M$ (Eq.2.15). Para obtermos uma outra técnica de solução, que evite o problema numérico da multiplicação $\mathbf{X}^T \mathbf{X}$ ou $\mathbf{C}^T \mathbf{C}$ ou $\Gamma^T \Gamma$ intrínseca ao cálculo da matriz de variância-covariância, precisamos definir uma matriz M , tal que $M^T M = S = \frac{1}{n-1} \mathbf{C}^T \mathbf{C}$.

Então, seja, $M = \sqrt{n-1}^{-1} \mathbf{C}$ (onde n é o número de amostras de \mathbf{C}), onde \mathbf{C} é a matriz de dados \mathbf{X} centralizada. Temos, então:

$$M^T M = \left(\frac{1}{\sqrt{n-1}} C^T \right) \left(\frac{1}{\sqrt{n-1}} C \right) = \frac{1}{n-1} C^T C = P \Lambda P = V(\Gamma^T \Gamma) V^T$$

Lembramos que a decomposição em vetor singular de M é $M = U \Gamma V^T$. Portanto, ao invés de se calcular a decomposição espectral da matriz de variância-covariância de \mathbf{X} ($Cov(\mathbf{X}) = S = \frac{1}{n-1} \mathbf{C}^T \mathbf{C}$), podemos obter a matriz de componentes principais (PC) e matriz diagonal com suas respectivas variâncias (Λ), apenas realizando a decomposição em valor singular de $\sqrt{n-1}^{-1} \mathbf{C}$. Onde a matriz de vetores singulares pela esquerda (V) serão as componentes principais ($e_i = v_i$), ou seja, $PC_i = V_i$) e os valores singulares (γ_i) ao

quadrado são as variâncias explicadas por cada componente ($\gamma_i^2 = \lambda_i$, ou seja, $\Gamma^T \Gamma = \Lambda$).

Desta forma podemos obter as componentes principais para os dados não padronizados evitando os problemas numéricos do cálculo da matriz de variância-covariância.

Da mesma forma esta solução utilizando o SVD pode ser aplicada a dados padronizado substituído a matriz de dados \mathbf{X} pela matriz de dados padronizados \mathbf{Z} . Neste caso não a a necessidade de centralizar Z , pois isto já é feito durante o processo de padronização (mas a execução desta centralização não geraria qualquer erro).

O Alg. 6 soluciona o PCA, evitando erros numéricos, através do uso do SVD (independente dos dados terem sido ou não previamente centralizados).

Algoritmo 6: Obtém as componentes principais utilizando pela decomposição em valor singular

Entrada: matriz de dados (\mathbf{X}).

Saída: matriz de coeficientes das componentes principais (A); vetor das variâncias explicadas por cada componente principal (λ).

início

$\mathbf{C} \leftarrow \mathbf{X}$ /* Se X centralizada, esta linha é opcional. */ ; $\mathbf{M} \leftarrow \sqrt{n - 1}^{-1} \mathbf{C}$ Realize a decomposição em valor singular da matriz \mathbf{M} ; Ordene os pares valor singular-vetor singular a esquerda (γ_i, v_i) , tal que, $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_p$; $A \leftarrow [v_1 \ v_2 \ \dots \ v_p]$; $\lambda \leftarrow [\gamma_1^2 \ \gamma_2^2 \ \dots \ \gamma_p^2]^T$;	
--	--

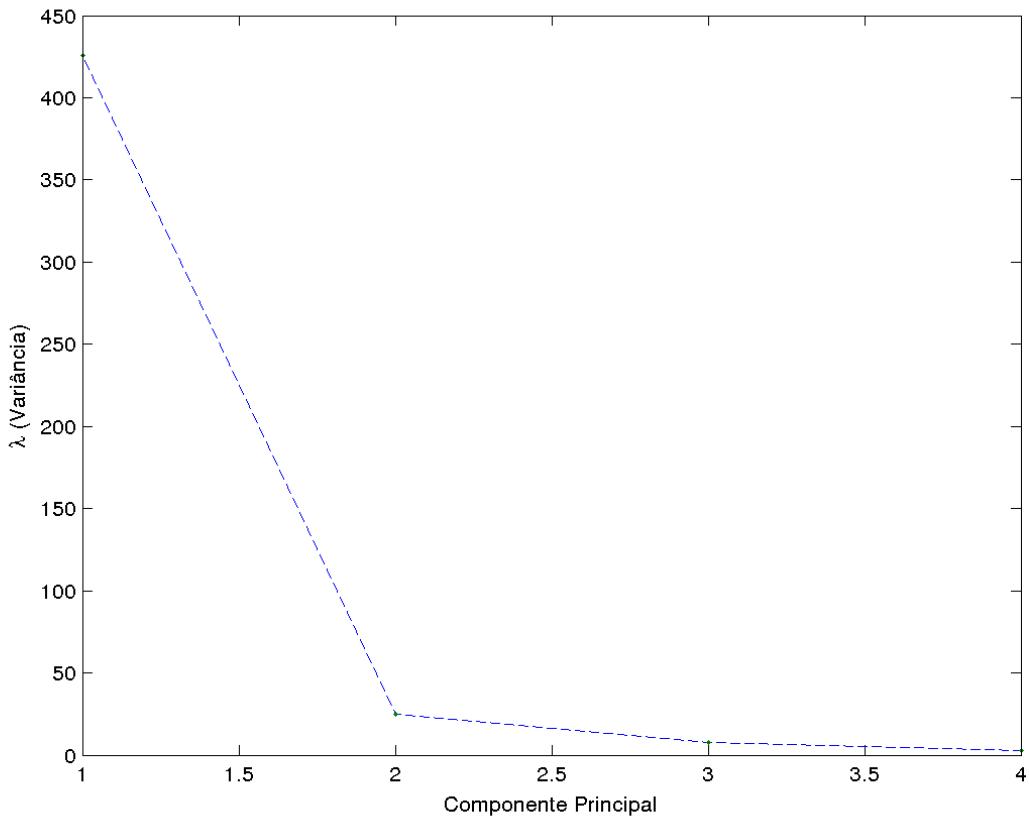
fim

2.1.5 Redução De Dimensionalidade A Partir Das Componentes Principais

Após obtermos as p componentes principais, independentemente da técnica utilizada, podemos realizar a redução da dimensionalidade. Nesta secção abordaremos os critério de decisão da dimensão k mais adequada para o novo espaço.

Um importante critério é a proporção da variabilidade explicada por uma componente (Eq. 2.12) e a proporção acumulada da variabilidade explicada pelas k primeiras componentes (Eq. 2.13). Pois, desta forma podemos definir qual é o menor k , tal qual, a variabilidade capturada por esta componentes seja satisfatória, ou seja, que a perda de

Figura 2.1: *Scree plot* do PCA para base de dados Iris não padronizada.



informação (variabilidade explicada pelas $p - k$ últimas componentes) seja suficientemente pequena.

Esta proporção satisfatória varia de aplicação para aplicação. Normalmente, representações que expliquem 85% da variabilidade dos dados são consideradas boas, em alguns casos este limite é aumentado para 95% (JOHNSON, 2007). Outra técnica é a análise do *scree plot* (gráfico de encosta⁹) do índice da componente pela sua respectiva variância. Onde, procura-se, principalmente, encontrar o *cotovelo* (*elbow*) da curva, o local onde uma alteração brusca na curvatura da curva.

A Fig. 2.1.5 mostra o *scree plot* de análise de componentes principais para a base de dados Iris (NEWMAN, 1998) *apud* (MOTTA, 2010) a partir dos dados não padronizados. Pela análise deste gráfico temos que uma boa dimensão para o espaço reduzido seria $k = 2$, pois é neste ponto que temos o *cotovelo* da curva.

Entretanto nem sempre o gráfico de declive obtido poderá indicar a dimensão

⁹Em referência ao cascalhos de rocha ao pé de penhascos

mais adequada, seja pela curva não apresentar um alteração brusca na sua curvatura, ou por apresentar várias alterações. Nesta última situação, os diversos pontos onde há esta variação brusca são bons candidatos a serem a dimensão mais adequada.

Em casos onde a redução seja aplicada para obtenção de uma representação onde a menor quantidade de dados é necessária para se economizar memória ou custo de manipulação destes, pode-se definir um limite ao tamanho dos dados e se utilizar a dimensão k , tal que, forneça a maior variabilidade acumulada dentro do limite de espaço imposto.

Outra estratégia de decisão para escolha do k que, também, não foca na quantidade de variabilidade explicada mas nos limites para k é a visualização de dados. Onde, devido a natureza da aplicação, o número de dimensões está sujeito a uma, duas, ou três, $k = 1, 2, 3$.

Definido a quantidade de dimensões desejadas ($k \leq p$), temos então que o novo espaço é definido pelo vetor característico $X_{reduzido} = [PC_1 PC_2 \cdots PC_k]^T$, onde as k primeiras componentes principais são as variáveis aleatórias. E os respectivos coeficientes, $a_i, i = 1, 2, \dots, k$ (ou seja, os autovetores, $e_i, i = 1, 2, \dots, k$), são os vetores unitários que dão a direção (e convencionalmente, também, o sentido) dos eixos deste espaço.

Por fim, com o espaço de características definido, os dados \mathbf{X} pode ser representados no novo espaço. Tradicionalmente esta representação dos dados é chamada de escore (*score*) dos dados, e portanto a nova matriz de dados é a matriz de Escores, **Score**. Mas outros nomes como matriz reduzida, matriz de dados reduzida, também são utilizados (**Score** $\equiv \mathbf{X}_{reduzido} \equiv \mathbf{X}_{red}$, ou ainda, indicando a dimensão do espaço, \mathbf{X}_{kD}).

Independentemente do nome utilizado, a transformação dos dados do espaço original para o reduzido é feito de acordo com a Eq. 2.18.

$$Score = X A_{reduzido} = X[a_1 a_2 \cdots a_k] = X[e_1 e_2 \cdots e_k]; k \leq p \quad (2.18)$$

2.2 DECOMPOSIÇÃO EM VALOR SINGULAR

Como vimos na Secção 2.1.4 a decomposição em valor singular, do inglês, *singular value decomposition* (SVD), também conhecida como teorema de Eckart-Young (*Eckart-Young theorem*), é uma das técnicas mais comuns de resolução do PCA. Entretanto, o

PCA via SVD ou pelas demais técnicas exige que os dados sejam centralizados, seja explicitamente (caso do SVD) ou implicitamente no cálculo da matriz de variância-covariância ou dos dados padronizados (Z). E, nem sempre, é desejável realizar este processo de obtenção da média das V.A. e posterior subtração dela de todos os dados, pois dependendo do número de amostras (n) isto pode ser um processo custoso, ou indesejado se os dados forem muito dispersos.

Nestes casos, pode-se aproveitar de algumas propriedades da decomposição em valor singular para se realizar a redução de dimensionalidade. Como vimos, o SVD de uma matriz qualquer é dada pela Eq. 2.14. Portanto, seja esta a matriz de dados ($M = \mathbf{X}$), temos:

$$\mathbf{X} = U S V = \sum_{i=1}^{\text{posto}(\mathbf{X})} s_i u_i v_i^T$$

Onde, $U = [u_1 u_2 \cdots u_n]$ é a matriz de vetores singulares pela direita, S a matriz retangular diagonal dos valores singulares, e $V = [v_1 v_2 \cdots v_p]$ a matriz de vetores singulares pela esquerda.

TAN (2005) ressalta algumas propriedades desta decomposição:

- Os vetores singulares pela direita (u_i) capturam padrões entre as amostras (objetos);
- Os vetores singulares pela esquerda (v_i) capturam padrões entre os atributos (V.A.);
- A matriz decomposta (no caso, \mathbf{X}) pode ser aproximada de forma ótima pegando-se em ordem os termos da decomposição. Isto é, os autovalores são naturalmente ordenados ($s_1 \geq s_2 \geq \cdots \geq s_p$), e a matriz original é representada da melhor forma aproximada utilizando-se os k primeiros valores e vetores singulares (evidentemente, que quanto mais valores e vetores forem utilizados melhor será a precisão da aproximação).
- Pode-se obter uma nova matriz de dados da mesma forma que obtivemos a matriz *Score* no PCA (Eq. 2.18), substituindo a matriz dos k primeiros coeficientes ($[a_1 a_2 \cdots a_k]$) pela matriz dos k primeiros vetores singulares pela esquerda, $[v_1 v_2 \cdots v_k]$.

Destas afirmações, percebe-se que é possível se obter um espaço de menor dimensão através da aplicação direta do SVD. Onde, de forma análoga ao PCA, o novo espaço é definido pelo vetor de características $X_{reduzido} = [v_1 X_1 v_2 X_2 \cdots v_k X_k]; k \leq p$. E o eixo de cada dimensão é definido pela direção (e convencionalmente, o sentido) do respectivo vetor singular a esquerda.

Por sua vez, a transformação da matriz de dados original (\mathbf{X}) para a matriz de dados no novo espaço (ou matriz de escores, $\mathbf{X}_{reduzida} = \mathbf{Score}$) é obtida pela Eq. 2.19.

$$Score = X V_{reduzido}^= X [v_1 v_2 \cdots v_k] ; k \leq p \quad (2.19)$$

Enquanto os autovalores representam a variabilidade explicada pela componente principal a qual estão associados, os valores singulares representam o quanto a decomposição nesta dimensão se aproxima da matriz original. Portanto, $s_i > s_j$ significa que $s_i u_i v_i^T = Fator_i$ aproxima-se mais de \mathbf{X} do que $s_j u_j v_j^T = Fator_j$.

Então, analogamente, a proporção de variabilidade explicada por uma componente (Eq. 2.12) e a proporção acumulada das k primeira componentes (Eq. 2.13), temos, a proporção da aproximação da matriz original (Eq. 2.20).

$$\begin{pmatrix} \text{Proporção da aproximação} \\ \text{segundo } Fator_i = s_i u_i v_i^T \end{pmatrix} = \frac{s_i}{s_1 + s_2 + \cdots + s_p} \quad (2.20)$$

Consequentemente a proporção acumulada da aproximação representada pelos k primeiros fatores (tal que $k \leq p$) é dado pela Eq. 2.21 abaixo:

$$\begin{pmatrix} \text{Proporção acumulada da aproximação} \\ \text{segundo os } k \text{ primeiros fatores} \end{pmatrix} = \sum_{i=1}^k \frac{s_i}{s_1 + s_2 + \cdots + s_p} ; k \leq p \quad (2.21)$$

Para se definir o número de dimensões utiliza-se as mesmas técnicas desenvolvidas para o PCA (Secção 2.1.5), adaptadas para o SVD:

- Menor k , tal que, a proporção de aproximação acumulada seja satisfatório (normalmente, maior que 85% ou 95%);

- Análise da *scree plot* dos valores singulares pelo índice do respectivo fator, escolhendo k de acordo com o *cotovelo* da curva;
- Maior k , tal que, o volume de dados não exceda um limite pré-estabelecido;
- $k = 1, 2, 3$ para aplicações de visualização dos dados.

2.3 ESCALONAMENTO MULTIDIMENSIONAL

O escalonamento multidimensional, do inglês, *multidimensional scaling* (MDS), é uma família de métodos de projeção de dados em um espaço com p dimensões em um espaço k -dimensional, tal que $k \leq p$. Sendo o processo de projeção é chamado de escalonamento (*scaling*) ou mapeamento (*mapping*).

Estes métodos diferem dos discutidos anteriormente pelo seu objetivo primário que é ajustar os dados originais em um sistemas de coordenadas com menor dimensão, tal que, qualquer distorção ocasionada pela redução de dimensionalidade seja minimizada. Onde, distorção geralmente se refere a distorção da similaridade ou dissimilaridades entre as amostras. Ou seja, um processo de escalonamento multidimensional ideal é aquele que a projeção obtida mantenha no espaço reduzido exatamente as mesmas similaridade ou dissimilaridade que os objetos tinham entre si no espaço original.

Portanto, o problema do MDS pode ser definido como: "Para um conjunto de similaridade (ou dissimilaridades) observadas entre pares de n itens [(amostras, objetos, pontos)], encontre uma representação destes itens, tal que, a proximidade inter-item 'corresponda aproximadamente' a similaridade (dissimilaridade) original." (JOHNSON, 2007)

Observa-se que se trata de um problema de otimização, onde a similaridade deve ser maximizada ou dissimilaridade minimizada. Sabendo ainda que a estas medidas são opostas, mas proporcionais entre si, temos que ambos os problemas de otimização são equivalentes. Tradicionalmente as técnicas de otimização são solucionadas como uma minimização (convertendo previamente maximizações em minimizações), então, utilizaremos no decorre do trabalho a definição do MDS pela minimização da dissimilaridade. Entretanto, todos os resultados a serem obtidos para tal são, também, aplicáveis a maximização da similaridade.

Evidentemente, normalmente não será possível entrar uma representação que corresponda exatamente a dissimilaridade original. Portanto, a otimização é feita em relação a uma função objetivo, ou de perda (*loss function*) que mensure o quanto uma projeção corresponde, se aproxima, possível aos dados originais.

Diversas funções objetivos, ou seja, métricas de correspondência, proximidade (*closeness*) podem ser utilizadas. É de acordo com que os métodos de escalonamento multidimensional são diferenciados. O tipo do método MDS é definido de acordo a dois critérios, valor da dissimilaridade e ajuste da dissimilaridade. Quanto ao valor da dissimilaridade as possibilidades são:

- Escalonamento métrico (*metric scaling*) - quando se utiliza os valores de dissimilaridade. Por exemplo, a distância em quilômetros de uma cidade a outra no mapa.
- Escalonamento não métrico (*nonmetric scaling*) - quando se utiliza apenas a ordem das dissimilaridades. Por exemplo, o quanto diferentes os consumidores avaliam, com notas de 1 (muito parecidos) à 5 (muito diferentes), ser pares de produtos. Esta técnica é largamente utilizada em casos onde se deseja analisar a relação subjetiva da dissimilaridade entre objetos, com inúmeras aplicações nas ciências sociais (BUJA, 2008)(COX, 2001). É executada por uma estimando uma transformação monotônica ótima da dissimilaridade, $f(\delta_{ij})$, simultaneamente ao .

Por sua vez, quanto ao ajuste da dissimilaridade:

- Escalonamento clássico (*classical Torgerson-Gower inner-product scaling*) - o ajuste da dissimilaridade é realizado em função do produto interno $b_{ij} = \langle x_i, x_j \rangle$, tal que, $\delta_{ij}^2 = b_{ii} - 2b_{ij} + b_{jj}$.
- Escalonamento não clássico (*Kruskal-Shepard distance scaling*) - o ajuste da dissimilaridade é realizado pelas distâncias d_{ij} . Indiferentemente da métrica de distância utilizada, comumente se utiliza as derivadas da distância Minkowski, principalmente a Euclidiana, seguida da Manhattan (ou *city block*).

Os métodos de MDS são definidos de acordo com a combinação do tipo de valor e ajuste de dissimilaridade da sua função objetivo. Podendo ser um escalonamento métrico

e clássico, ou métrico e não clássico, ou não métrico e não clássico, ou, por fim, não métrico e clássico. Sendo o último muito raramente utilizado, independentemente da área. Tanto que, normalmente é desconsiderado e comumente o escalonamento métrico e clássico é chamado apenas por escalonamento clássico, e será desta forma que o denominaremos no decorrer deste trabalho.

A partir deste ponto só consideraremos os MDS métricos por serem os normalmente utilizados em aplicações dentro de um processo de KDD (TAN, 2005). Para maiores informações sobre o escalonamento não métrico e não clássico consulte (KRUSKAL, 1964b)(trabalho no qual o escalonamento não métrico foi inicialmente proposto), e para o não métrico e clássico sugere-se BUJA (2008).

2.3.1 Escalonamento Multidimensional Clássico

Chamado de escalonamento clássico (*classical scaling* - CMDS) por ter sido o primeiro método prático desenvolvido para o escalonamento multidimensional por TORGERSON (1952), TORGERSON (1958)e GOWER (1966), também é conhecido pelos nomes de escalonamento de Torgerson (*Torgerson scaling*) e escalonamento de Torgerson-Gower (*Torgerson-Gower scaling*). Seus fundamentos surgiram com os trabalhos de ECKART (1936)e (YOUNG, 1938)sobre a reconstrução, aproximada, dos dados originais, a partir de sua matriz de similaridade (ou dissimilaridade).¹⁰

A ideia básica do MDS clássico é assumir que as dissimilaridades são distâncias, e então, encontrar as coordenadas que expliquem estas distâncias, obtendo, assim, as coordenadas do novo espaço.

Seja Δ_Y a dissimilaridade explicada pelo espaço característico $Y^T = [Y_1 Y_2 \cdots Y_p]$, \mathbf{Y} a matriz de dados neste espaço, e D_Y a distância entre as amostras de \mathbf{Y} . Assumimos que $\Delta_Y = D_Y$. E como as distâncias não são afetadas por translações dos eixos coordenados consideraremos que nosso espaço está centralizado (todas as V.A. tem média zero). Segundo BORG (2005)a distância pode ser computada pela Eq. 2.22.

$$D_Y = c\mathbf{1}^T + \mathbf{1}\mathbf{c}^T - 2\mathbf{Y}\mathbf{Y}^T = c\mathbf{1}^T + \mathbf{1}\mathbf{c}^T - 2\mathbf{B} \quad (2.22)$$

¹⁰COX (2001)apresenta detalhadamente como esta recuperação dos dados originais é realizada e sua base teórica.

Onde, B é a matriz de produtos internos (ou matriz de produtos escalares), logo, $B = \mathbf{Y}\mathbf{Y}^T$. E, c é o vetor com os elementos da diagonal de B ($c = \text{diag}(B)$). Isto é o mesmo que:

$$d_{ij}^2 = b_{ii} - 2b_{ij} + b_{jj}; b_{lk} = \langle y_l, y_k \rangle; \forall l, k = 1, 2, \dots, n$$

Multiplicando a Eq. 2.22 pelo escalar $-1/2$ e em ambos os lados pela matriz de centralização, $J = I - \mathbf{1}n^{-1}\mathbf{1}^T$, onde, n é o número de amostras de \mathbf{Y} , que é o mesmo que o número de colunas e de linhas de D_Y , temos:

$$\begin{aligned} \frac{-1}{2}JD_Y &= \frac{-1}{2}J(c\mathbf{1}^T + \mathbf{1}c^T - 2\mathbf{B})J = \frac{-1}{2}Jc\mathbf{1}^TJ + \frac{-1}{2}J\mathbf{1}c^TJ - J\mathbf{B}J = \\ &= \frac{-1}{2}Jc\mathbf{0}^T + \frac{-1}{2}\mathbf{0}c^TJ - J\mathbf{B}J = JBJ = B \end{aligned}$$

É evidente, que ao se centralizar um vetor de uns, $\mathbf{1}$ ou $\mathbf{1}^T$ ($J\mathbf{1}$ ou $\mathbf{1}^TJ$, respectivamente), obtemos um vetor de zeros, $\mathbf{0}$ ou $\mathbf{0}^T$. E, como \mathbf{Y} é centralizada, consequentemente B , também o é, logo a dupla centralização de B é ela própria ($JBJ = B$).

Finalmente, as coordenadas do espaço Y são dadas pela direção (e convencionalmente o sentido) dos autovalores da matriz de produto interno. Então, seja a decomposição espectral $B = \mathbf{Y}\mathbf{Y}^T = Q\Lambda Q^T$, onde $Q = [e_1 e_2 \dots e_p]$, tal que, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. Os eixos coordenados são e_1, e_2, \dots, e_p .

Aplicando o resultado obtido a definição geral do MDS, temos que, dada uma matriz de dissimilaridade Δ_X entre as amostras da matriz de dados \mathbf{X} no espaço característico X , o escalonamento objetiva encontrar uma projeção desses dados (\mathbf{Y}) em um espaço característico Y , tal que, a dissimilaridade entre os itens do novo espaço (Δ_Y) seja o mais próximo da original (Δ_X). Isto é, deseja-se minimizar a discrepância, perda na representação, entre as dissimilaridades entre os amostras no espaço X e no espaço Y . Uma métrica comum para mensurar a discrepancia entre duas matrizes é, a soma residual de quadrados:

$$L = \|\Delta_Y^2 - \Delta_X^2\|^2$$

Como, no escalonamento multidimensional clássico assumimos que a dissimilaridade são distâncias, e as distâncias entre as observações não são afetadas pela translação dos eixos, logo:

$$L = \|D_{\mathbf{Y}}^2 - D_{\mathbf{X}}^2\|^2 = \left\| \frac{-1}{2} J(D_{\mathbf{Y}}^2 - D_{\mathbf{X}}^2)J \right\|^2 = \left\| \left(\frac{-1}{2} JD_{\mathbf{Y}}^2 J \right) + \left(\frac{1}{2} JD_{\mathbf{X}}^2 J \right) \right\|^2$$

Da Eq. 2.22 (o mesmo pode ser obtido analogamente para os espaço original X), temos:

$$L = \|B_Y - B_X\| = \|\mathbf{Y}\mathbf{Y}^T - \mathbf{X}\mathbf{X}^T\|$$

Onde L é uma função de perda que mensura a falta de ajuste entre os produtos internos B_Y e B_X .

Toda função de perda que mensure a falta da ajuste entre produtos internos é chamada de *Strain*, portanto, a soma residual de quadrados, função L , é uma função *Strain* e pode ser definida por Eq. 2.23.

$$\begin{aligned} Strain &= \|B_Y - B_X\| = \left(\sum_{i=1}^n \sum_{j=1}^n \|\langle y_i, y_j \rangle - \langle x_i, x_j \rangle\|^2 \right)^{1/2} = \\ &= \left[\sum_{i=1}^n \sum_{j=1}^n (\langle y_i, y_j \rangle - \langle x_i, x_j \rangle)^2 \right]^{1/2} \end{aligned} \quad (2.23)$$

Além da Eq. 2.23, outras funções *Strain* também podem ser derivadas como funções objetivos para o escalonamento clássico. A mais comum entre elas é a soma residual de quadrados padronizada, definida na Eq. 2.24 (BUJA, 2008). A qual conta com a vantagem de ser padronizada, e por isto, livre de influências de escala e unidade.

$$Strain = \left[\frac{\sum_{i=1}^n \sum_{j=1}^n (\langle y_i, y_j \rangle - \langle x_i, x_j \rangle)^2}{\sum_{i=1}^n \sum_{j=1}^n \langle x_i, x_j \rangle^2} \right]^{1/2} \quad (2.24)$$

Existem diversas implementações iterativas que visão minimizar o *Strain*. Entretanto, segundo BORG (2005), o escalonamento multidimensional clássico é popular pois possui uma solução analítica, não necessitando de iterações. Ele, também, deriva esta so-

lução a partir dos teoremas citados acima e consequentemente, obtém um algoritmo para solução do MDS clássico (Alg. 7), o qual não necessita de aplicar técnicas de otimização.

Algoritmo 7: Executa o escalonamento multidimensional clássico

Entrada: matriz de dissimilaridades dos dados originais (Δ)

Saída: matriz de coordenada escalada (\mathbf{Y})

início

```

    Calcula a matriz de dissimilaridades quadradas,  $\Delta^2$ ;
    /* Aplica a centralização dupla à matriz  $\Delta^2$  */ 
     $B_\Delta \leftarrow \frac{-1}{2} J \Delta^2 J; J = I - \mathbf{1} n^{-1} \mathbf{1}^\mathbf{T}; n$  é o número de linhas de  $\Delta^2$ ;
    Realiza a decomposição espectral de  $B_\Delta$ ;
    Ordene os pares autovalor-autovetor  $(\lambda_i, e_i)$ , tal que,  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ ;
     $\lambda_+ \leftarrow [\lambda_1 \dots \lambda_m]; \lambda_i \geq 0; i = 1, \dots, m \leq p$ 
    Defina  $\Lambda_+$  como a matriz diagonal com os elementos de  $\lambda_+$ 
     $P_+ \leftarrow [e_1 \dots e_m] \mathbf{Y} \leftarrow P_+ Q_+^{1/2}$ 

```

fim

Onde, a notação de Δ_X foi simplificada para apenas Δ . E, $J = I - \mathbf{1} n^{-1} \mathbf{1}^\mathbf{T}$ é a matriz de centralização, tal que, n seja o número de linhas (e colunas) de Δ^2 (que é o mesmo de Δ), ou seja, o número de amostras, objetos.

Diferentemente dos outros métodos de escalonamento multidimensional que veremos o MDS minimizando o *Strain* (ou seja, o escalonamento clássico) gera soluções aninhadas. Ou seja, o espaço com $k + 1$ dimensões composto exatamente pelas dimensões do espaço com k dimensões mais a $(k + 1)$ -ésima dimensão, e assim consecutivamente.

Outra, propriedade particular do escalonamento clássico é que analogamente ao PCA é possível se obter a variabilidade explicada por cada uma de suas coordenadas Y_i (COX (2001) deriva formalmente esta propriedade). Portanto, podemos adicionar uma saída ao Alg. 7, que seria a matriz Λ_+ . E a variabilidade explicada por uma coordena ou a variabilidade explicada acumulada são definidas pelas Eq. 2.25 e Eq. 2.26, respectivamente.

$$\begin{pmatrix} \text{Proporção da} \\ \text{variância total} \\ \text{segundo } Y_i \end{pmatrix} = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_m} \quad (2.25)$$

Onde, $m \leq p$ é a quantidade de autovalores não negativos obtidos na decomposição espectral de B_Δ .

$$\left(\begin{array}{l} \text{Proporção acumulada da} \\ \text{variância total segundo} \\ \text{as } k \text{ primeiras } Y \end{array} \right) = \sum_{i=1}^k \frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_m}; k \leq m \quad (2.26)$$

Portanto, as estratégias de decisão do número de dimensões a se preservar em uma redução são as mesmas abordadas na Secção 2.1.5 para o PCA (não havendo a necessidade de discorrermos sobre elas novamente). Note, ainda que a matriz de coordenadas escaladas, \mathbf{Y} , é análoga a matriz de *Score* do PCA, e, por isto, também é chamada por este nome.

No caso das dissimilaridades entre as observações originais serem dadas pela distância euclidiana todos os autovalores serão não negativos e portanto $m = p$. Entretanto, ao se utilizar outras distâncias, ou outras métricas de dissimilaridade, pode-se ocorrer autovalores negativos, os quais são considerados como erros pelo MDS clássico e devem ser descartados.

Ainda, quando dissimilaridade de entrada é a distância euclidiana temos outra consequência importante. Conforme demonstrado por COX (2001)e BORG (2005)o escalonamento multidimensional clássico, neste caso, é equivalente a análise de coordenadas principais, do inglês, *principal coordinates analysis* (PCO), e ao análise de componente principais (PCA). Ou seja, o MDS clássico quando sua entrada for uma matriz de distâncias euclidianas obtém o mesmo resultado que o primeiro método que vimos neste trabalho (Secção 2.1) TAN (2005).

2.3.2 Escalonamento Multidimensional Não Clássico

A forma mais utilizada de escalonamento multidimensional é a não clássica e métrica (*nonclassical metric multidimensional scaling*), por isto, que comumente é chamado apenas por MDS não clássico (nome que utilizaremos para identificá-lo a partir deste ponto), ou ainda, apenas por MDS. Esta formulação foi proposta por SHEPARD (1966)e KRUSKAL (1964a). Por este motivo, também é chamado de escalonamento de Kruskal-Shepard (*Kruskal-Shepard scaling*).

Neste método de escalonamento as dissimilaridades δ_i são ajustadas pelas distâ-

cias d_i . Sendo assim, a definição do MDS pode ser enunciada enunciada como: dada uma matriz de distância $D_{\mathbf{X}}$ que representa dissimilaridade $\Delta_{\mathbf{X}}$ entre as amostras da matriz de dados \mathbf{X} no espaço de características X , o escalonamento tem por objetivo encontrar uma projeção destes dados (\mathbf{Y}) em um espaço Y , tal que, a distância entre os itens no novo espaço ($D_{\mathbf{Y}}$) que representa a dissimilaridade $\Delta_{\mathbf{Y}}$ seja o mais próximo da original.

Isto é, deseja-se minimizar a discrepância, perda na representação, entre as distâncias das observações no espaço X e no espaço Y . Qualquer métrica utilizada para mensurar a falta de ajuste (discrepância) entre distâncias é denominada *Stress*. Uma fórmula comum para se calcular o *Stress* entre duas matrizes de distâncias é, a soma residual de quadrados (Eq. 2.27).

$$\text{Stress} = [\|D_{\mathbf{Y}} - D_{\mathbf{X}}\|^2]^{1/2} \quad (2.27)$$

Entretanto a Eq. 2.27 sofre influência da unidade e da escala dos dados. Desta forma, não significa que um valor de *Stress* alto indique que a representação é ruim. Por exemplo, sejam os itens cidades, e as distâncias entre elas dada em metros, o *Stress* obtido seria mil vezes maior do que se estes dados fossem mensurados em quilômetros. Portanto, para solucionar isto a função de *Stress* mais utilizada é a soma residual de quadrados padronizada, a qual é descrita na equação Eq. 2.28, esta função de *Stress* é comumente conhecido pelos nomes de *Kruskal – Stress*, *Kruskal – Stress – 1* e, o q utilizaremos, *Stress – 1*.

$$\text{Stress} - 1 = \left[\frac{\sum_{i=0}^n \sum_{j=0}^n (\hat{d}_{ij} - d_{ij})^2}{\sum_{i=0}^n \sum_{j=0}^n d_{ij}^2} \right]^{1/2} \quad (2.28)$$

Onde, \hat{d}_{ij} é a distância entre os itens y_i e y_j , ou seja, $\hat{D} = D_{\mathbf{Y}}$.

KRUSKAL (1964a)sugere que a qualidade do ajuste seja interpretado a partir do *Stress* (Eq. 2.28) seguindo a Tab. 2.3.2.

Stress Qualidade do ajuste

20% Ruim

10% Razoável

5% Bom

2, 5% Excelente

0% Perfeito

Outra função de perda, também, utilizada para o mds não clássico é o *SSTress* introduzido por TAKANE (1977) *apud* (JOHNSON, 2007), definida na Eq. 2.29.

$$SSTress = \left[\frac{\sum_{i=0}^n \sum_{j=0}^n (\hat{d}_{ij}^2 - d_{ij}^2)^2}{\sum_{i=0}^n \sum_{j=0}^n d_{ij}^4} \right]^{1/2} \quad (2.29)$$

O *SSTress* é sempre um valor entre zero e um ($0 \leq SSTress \leq 1$). E, tipicamente, são consideradas boas representações aquelas que tenham $SSTress < 0,1$.

Diferentemente do MDS clássico, o não clássico não tem uma solução analítica e portanto para sucioná-lo devemos resolver a otimização *minStress* (seja o *Stress* definido conforme a (Eq. 2.27, Eq. 2.28, Eq. 2.29, ou outras existentes) utilizando alguma técnica de otimização não linear. Portanto, obtemos um algoritmo para o MDS não clássico (Alg. 8).

Algoritmo 8: Soluciona o MDS não clássico

Entrada: matriz de dissimilaridades dos dados originais (D); número de dimensões desejadas para o espaço objetivo (k).

Saída: matriz de coordenadas escaladas (\mathbf{Y}).

início

Inicializa a matriz de coordenadas escaladas \mathbf{Y} com n amostras de k dimensões;

$minStress$;

Retorna a matriz de coordenadas escaladas \mathbf{Y} final obtida durante a otimização;

fim

Onde, normalmente a minimização é feita utilizando-se do método do gradiente (do inglês, *gradient descent* ou *steepest descent*). E a inicialização pode ser feita de diversas formas, sendo as mais comuns:

- Gerar uma matriz $n \times k$ aleatoriamente;

- Executar o escalonamento clássico e utilizar a matriz de dados escalados $n \times k$ obtida por este.

Observe, que o Alg. 8 exige ser informado previamente do número de dimensões desejado para o espaço objetivo. Isto se deve ao fato de que os espaços obtidos para o MDS não clássico não são aninhados, diferentemente do MDS clássico. Ou seja, o espaço $k+1$ não é necessariamente composto pelas dimensões do espaço k -dimensional mais uma dimensão.

Então pare se decidir sobre quanta dimensões utilizar para o espaço reduzido uma técnica é executar o Alg. 8 para diferentes k e escolher o que obtiver o menor valor de Stress. Para isto, adicionamos o *Stress* final como uma saída ao Alg. 8.

Vale ainda ressaltar que a matriz de coordenadas escaladas, \mathbf{Y} , é a matriz de dados no espaço reduzido, e portanto também é chamada de matriz de escores ($\mathbf{Y} = Score$), da mesma forma que utilizamos no PCA e no CMDS.

2.4 FASTMAP

O FastMap é um dos métodos mais recentes da área de redução de dimensionalidade e foi proposto por FALOUTSOS (1995) como uma alternativa ao MDS, com as vantagens de ser mais rápido (complexidade linear) e poder operar de forma incremental (TAN, 2005).

Diferentemente de todos os métodos que foram apresentados (PCA, SVD, CMDS e MDS não clássico), quando uma amostra é incluída aos dados não se é necessário executar novamente todo o processo de obtenção do espaço reduzido e a posterior transformação da nova amostra para o espaço reduzido, ou ainda apenas transformar a nova amostra para o espaço reduzido obtido anteriormente, o que pode implicar em inadequações da representação. O FastMap consegue gerar um novo espaço reduzido com apenas uma iteração, contanto que tenha como entrada as distâncias do novo ponto (amostra, item) para todos os demais da base, por isto dizemos que ele tem complexidade linear na inserção.

A ideia básica deste método dada a matriz de distância de um conjunto de dados e a dimensão do espaço objetivo é:

1. Determine os dois itens (amostras, observações, pontos) de mais distantes, os quais são chamados de pivôs;
2. O eixo principal será o segmento que liga estes dois vetores;
3. Projete todos os pontos sobre este eixo (isto é feito utilizando a lei dos cossenos);
4. Se a matriz de dados obtida pela projeção sobre o eixo dos pivôs é da dimensão desejada, ela será a matriz \mathbf{Y} . Caso contrario continue;
5. Defina um hiper-plano H ortogonal (perpendicular) ao eixo dos pivôs;
6. Projete todos os pontos sobre o plano H ;
7. Defina a matriz de distância \hat{D} para as projeções sobre o plano ortogonal;
8. Repita os passos anteriores utilizando a matriz de distância obtida no item (6).

A projeção dos dados sobre o eixo dos pivôs pode ser calculada sem a necessidade de se conhecer a matriz de dados, apenas com a matriz de distâncias. Conforme a Eq. 2.30, a qual foi derivada da lei dos cossenos (FALOUTSOS, 1995).

$$y_i = \frac{d_{ai}^2 + d_{ab}^2 + d_{bi}^2}{2d_{ab}} \quad (2.30)$$

Onde, y_i é o i-ésimo item da matriz de dados reduzida (\mathbf{Y}), a e b são o índice dos itens escolhidos como pivôs, e d_{jk} a distância entre os itens j e k no espaço original.

Para se obter a distancia da projeções dos itens no plano perpendicular ao eixo dos pivôs não é necessário que definamos o plano ortogonal H , nem as projeções neste plano. Podemos obtê-las diretamente da matriz de distância original e das projeções no eixo dos pivôs utilizando o teorema de Pitágoras, conforme definido na Eq. 2.31 (FALOUTSOS, 1995).

$$\hat{d}_{ij} = \sqrt{d_{ij}^2 - (y_i - y_j)^2} \quad (2.31)$$

Sendo, \hat{d}_{ij} a distância entre as projeções de x_i e x_j no plano H .

Alg. 9 apresenta o algoritmo iterativo do FastMap (a versão recursiva pode ser encontrada em (FALOUTSOS, 1995)).

Algoritmo 9: Algoritmo FastMap

Entrada: matriz de distância original (\mathbf{D}); dimensão desejada para o espaço reduzido (k).
Saída: matriz de dados projetados no espaço k -dimensional (\mathbf{Y}).
início

$D \leftarrow$ matriz de distâncias entre os itens de \mathbf{X} ;

repita

Determine os pivôs p_1, p_2 ;

$\hat{\mathbf{Y}} \leftarrow$ projeções dos itens sobre o segmento que liga os pivôs;

$\mathbf{D} \leftarrow$
distâncias da projeções dos itens no plano perpendicular ao eixo dos pivôs;

até dimensão de $\hat{\mathbf{Y}} > k$;

$\mathbf{Y} \leftarrow \hat{\mathbf{Y}}$;

fim

Onde, $k < p$ e a determinação dos pivôs é realizada segundo Alg. 10.

Algoritmo 10: Determina os pontos mais distantes (pivôs)

Entrada: matriz de distância (\mathbf{D}).
Saída: índice do primeiro pivô (a); índice do segundo pivô (b).
início

$a \leftarrow 1$;

$b \leftarrow$
o índice da coluna com a maior distância dentre as da (a)-ésima linha de \mathbf{D} ;

repita

$c \leftarrow$
o índice da coluna com a maior distância dentre as da (b)-ésima linha de \mathbf{D} ;

se a e c são iguais **então**

return ;

fim se

$a \leftarrow b$;

$b \leftarrow c$;

até Infinitamente;

fim

Nota-se que o FastMap é baseado na escolha dos pontos mais distantes, uma heurística que visa tentar obter a direção de maior variabilidade. Como matriz de distâncias da próxima iteração é gerada sobre o plano ortogonal ao eixo dos pivôs, logo, a direção que se espera ser a direção de maior variabilidade, todos os eixos obtidos são perpendiculares entre si. Por isto que este método é apelidado de PCA métrico, pois utilizando-se da ideia das projeções do MDS, ele gera um espaço característico onde todos os eixos são

ortogonais entre si e estão nas direções de maior variabilidade, segundo a heurística.

Por se tratar de um método heurístico ele não garante que esta escolha dos pivôs realmente dê a direção de maior variância dos dados (diferentemente do PCA). Entretanto, em casos reais, apesar de normalmente não ser obtida a solução ótima, a projeção final gerada pelo FastMap costuma ser boa. Sendo preferível para muitas aplicações, como as sobre base de dados multimídia (CANO, 2002), se obter esta representação satisfatória, do que pagar o custo de se executar o PCA, SVD, ou MDS (os quais são $O(n^2)$) para obtenção da representação ótima.

Entretanto, o FastMap tem uma limitação além da intrínseca aos métodos heurísticos. O cálculo da distância projetada no plano ortogonal pelo teorema de Pitágoras (Eq. 2.31 só é válida quando a matriz de distâncias originais (D) passada como argumento ao método for uma matriz de distâncias euclidianas. Isto se deve, ao fato de que estas projeções de distâncias não euclidianas podem ser um número complexo. Atualmente, existe extensões do FastMap que modificam o cálculo da distâncias projetadas quando D não é euclidianas. Mas estes métodos ainda são pouco difundidos.

3 Experimentos

Neste capítulo apresentaremos seis bases de dados de problemas de classificação comumente utilizados para comparação de métodos e algoritmos de KDD e aprendizado de máquina. Definiremos as configurações dos métodos utilizados e métricas para a avaliação destes. Posteriormente, aplicaremos os métodos abordados (Cap. 2) para realizar a transformação do espaço de características a cada uma das bases, realizaremos a redução de dimensionalidade para todas as dimensões k possíveis ($1 \leq k \leq p$) e finalmente apresentaremos os valores da métricas para cada tripla base de dado, algoritmo e dimensão.

Neste trabalho, utilizamos tanto a matriz de dados originais (\mathbf{X}) quanto a padronizada (\mathbf{Z}) e duas das métricas de distâncias mais comuns na área de KDD (HAN, 2006), a Euclidiana e a Manhattan. E, está além do escopo deste trabalho definir qual é a forma mais adequada para matriz de dados¹¹ ou a métrica de distância mais adequada as bases. Portanto, os métodos CMDS, MDS Stress-1 e MDS Sstress serão executados para as duas distâncias. Mas os métodos serão comparados entre as métricas de avaliação calculadas para cada cenário. Desta forma, temos quatro cenários:

- Cenário \mathbf{X} e $\mathbf{D}_{\text{Euclidiana}}$: Matriz de dados original e matriz de distâncias dos dados originais segundo métrica de distância Euclidiana;
- Cenário \mathbf{X} e $\mathbf{D}_{\text{Manhattan}}$: Matriz de dados original e matriz de distâncias dos dados originais segundo métrica de distância Manhattan;
- Cenário \mathbf{Z} e $\mathbf{D}_{\text{Euclidiana}}$: Matriz de dados padronizada e matriz de distâncias dos dados padronizados segundo métrica de distância Euclidiana;
- Cenário \mathbf{Z} e $\mathbf{D}_{\text{Manhattan}}$: Matriz de dados padronizada e matriz de distâncias dos dados padronizados segundo métrica de distância Manhattan.

¹¹Os motivos que definem isto foram relatados no Cap. 2

3.1 BASE DE DADOS

As bases utilizadas foram as dos problemas de classificação: Triagem de crédito (*Credit Screening*), Imagens (*Images*), Iris (*Iris*, Sonar (*Sonar*), Espiral (*Spiral*) e Vogal (*Vowel*) (NEWMAN, 1998) *apud* (MOTTA, 2010). A tabela 3.1 apresenta algumas características destas bases.

Tabela 3.1: Bases de dados

Nome	Número de Amostras (n)	Número de Dimensões (p)	Número de Classes (m)
Triagem de Crédito	653	15	2
Imagens	210	18	7
Iris	150	4	3
Sonar	208	60	2
Espiral	92	2	2
Vogal	990	10	11

Além da matriz de dados original, \mathbf{X} , obteve-se a matriz de dados padronizada, \mathbf{Z} , (utilizando escore-Z, *Z-score*) e as matrizes de distâncias Euclidianas, $\mathbf{D}_{\text{Eucladiana}}$, e de Manhattan, $\mathbf{D}_{\text{Manhattan}}$ (*City Block*, $\mathbf{Deuc} = \mathbf{Dcb}$) referentes tanto aos dados originais quanto ao padronizados (ou seja, quatro matrizes de distâncias distintas: $\mathbf{D}_{\text{Eucladiana}}(\mathbf{X})$, $\mathbf{D}_{\text{Eucladiana}}(\mathbf{Z})$, $\mathbf{D}_{\text{Manhattan}}(\mathbf{X})$ e $\mathbf{D}_{\text{Manhattan}}(\mathbf{Z})$).¹²

3.2 MÉTODOS

Como vimos os métodos análise de componentes principais e escalonamento multidimensional podem ser executados de diferentes formas. Nestes experimentos, a obtenção das componentes principais do PCA foi feita através da decomposição em valor singular conforme descrito na Secção 2.1.4.¹³ Pois esta forma de resolução evita o erro numérico da obtenção das componentes pela matriz de variância-covariância ou pela matriz de correlação.

Por sua vez, se utilizou três técnicas de escalonamento multidimensional, todos na formulação métrica (matriz de dissimilaridade de entrada é uma matriz de distâncias).

¹²A padronização e cálculo das distâncias foram feitas utilizando, respectivamente, as funções *zscore* e *pdist*, do ambiente/linguagem *MatLab 2012a* (MATHWORKS, 2012).

¹³O PCA foi executado utilizando-se da função *princomp* do *MatLab 2012a* (MATHWORKS, 2012).

O primeiro é o escalonamento clássico (CMDS) analítico¹⁴, pois esta maneira de resolução garante se encontrar as coordenadas que minimizam o *Strain* (ótimo global), é livre da pré-determinação do número de dimensões desejado (k), e não está sujeita aos problemas da solução de uma otimização, por exemplo a definição de parâmetros, divergência e custo computacional (tanto em desempenho quanto memória) da otimização.

Entretanto, nesta configuração o método não calcula no seu decorrer a valor da função de perda. Uma vez que temos a matriz de dados que gerou a matriz de dissimilaridade (distâncias) de entrada e o CMDS retorna os dados transformados (**Score** ou **Y**), podemos obter o valor desta função externamente a execução através da Eq. 2.24.

Vale ressaltar que a função *Strain* tem um valor para cada dimensão $k \leq p$. Portanto, definindo um k , a matriz de **Y** utilizada para o cálculo do Strain_k deve ser a matriz de dados transformada nesta dimensão. Como as dimensões do CMDS são aninhadas e a resolução analítica obtém todas elas, isto significa que para o Strain_k deve-se utilizar as k primeiras colunas de **Y**. Lembrando que a matriz de dados utilizada deve ser sempre a matriz com todas as p dimensões (a mesma que foi utilizada para se definir a matriz de distâncias recebida pelo execução).

Os outros dois escalonamentos são não clássico, onde a função de perda (*Stress*) utilizada por um foi o $\text{Stress} - 1$ (Eq. 2.28) e pelo outro foi o *Sstress* (Eq. 2.29). Ambos não tem solução analítica, e portanto obtiveram a solução através da minimização das respectivas funções de perda. Esta otimização foi realizada utilizando o método do gradiente com solução inicial (matriz **Y** inicial) randômica, número máximo de iterações igual a 200 e tolerância de 1^{-4} para função objetivo, gradiente e passo mínimo.

Como as dimensões do MDS não clássico não são aninhadas e cada execução do escalonamento multidimensional por otimização obtém apenas um espaço de coordenadas, o referente ao k pré-definido. Para cada base de dados o MDS com $\text{Stress} - 1$ e com *Sstress* foram executados p vezes, uma para cada $k \leq p$.¹⁵ Ressalta-se, ainda, que a fim de se evitar mínimos locais e devido a inicialização randômica, o que referimos aqui por

¹⁴O CMDS analítico foi executado utilizando-se da função *cmdscale* do *MatLab 2012a* (MATHWORKS, 2012).

¹⁵O MDS não clássico com $\text{Stress} - 1$ e com *Sstress* foram executados utilizando-se da função *mdscale* do *MatLab 2012a* (MATHWORKS, 2012). E a função de perda foi escolhida definindo o parâmetro referente como '*metricstress*' e '*metricsstress*', respectivamente.

uma execução, no caso dos escalonamentos multidimensionais não clássicos foram três execuções mantendo os resultados apenas da que obteve o melhor (menor) valor para função de perda e descartando as outras duas.

Além do PCA, CMDS, MDS Stress-1 e MDS Sstress, também foram utilizados os métodos decomposição em valor singular (SVD)¹⁶ e FastMap¹⁷. Onde o primeiro obtém em apenas uma execução todas as dimensões pois elas são aninhadas. E o segundo, necessita de ser realizado para cada dimensão $k \leq p$, entretanto vale lembrar que o FasMap não é realizado por uma otimização, mas sim por uma heurística iterativa, obtendo sempre o mesmo mapeamento para uma dimensão, diferentemente do MDS não clássico, ele é determinístico.

A tabela 3.2 relaciona os métodos utilizados em relação aos quatro cenário descritos anteriormente e a quantidade de execuções necessárias, para se obter uma representação para cada dimensão (por exemplo, quando as dimensões são aninhadas, em uma única execução se obtém os dados em todas as dimensões). As células mescladas indicam que os resultados obtidos para ambos os cenários são os mesmos, e consequentemente, não é necessário repetir os experimentos.

Tabela 3.2: Execução dos Métodos por Cenário de Experimentos

	Cenários de Experimentos			
	X		Z	
	Euclidiana	Manhattan	Euclidiana	Manhattan
PCA		1		1
SVD		1		1
CMDS	1	1	1	1
MDS Stress-1	p	p	p	p
MDS Sstress	p	p	p	p
FastMap	p		p	

Importante dizer que todos os experimentos foram realizados em um computador com sistema operacional *Linux Ubuntu 12.04 LTS*, 2GB de memória e uma única cpu *Intel Core 2 Duo* de 2.0GHz. Utilizando-se do ambiente, linguagem de programação e pacotes nativos do *MatLab 2012a* (MATHWORKS, 2012), sendo que o algorítimo do FastMap, escrito em C e disponibilizado pelo inventor deste método (FALOUTSOS, 1995),

¹⁶O SVD foi executado utilizando-se da função *svd* do *MatLab 2012a* (MATHWORKS, 2012).

¹⁷O FastMap foi executado pelo programa disponibilizado pelo inventor do método (FALOUTSOS, 1995), o qual foi escrito em C.

foi executado no terminal pelo comando de execução de *shell Unix* do *MatLab*.

3.3 MÉTRICAS

Este trabalho trata de uma análise comparativa dentre cinco dos principais métodos de redução de dimensionalidade (e configurações destes), entretanto, cada um utiliza uma métrica distinta. O PCA otimiza a proporção de variância acumulada, enquanto o SVD a proporção de aproximação da matriz dados de entrada. Por sua vez, o CMDS e o MDS otimizam suas respectivas funções de perda: *Strain*, *Stress* – 1 e *Sstress*. E por último, temos o FastMap, que apesar de não ter nenhuma métrica de qualidade da transformação das características do espaço e consequentemente da redução da dimensionalidade, pretende fazer isto rapidamente (como seu próprio nome indica), ou seja, podemos dizer que ele visa otimizar o tempo de execução.

É evidente, que o tempo de execução pode ser facilmente obtido (em código) independente do método. Mas observando a definição das demais métricas (Cap. 2) percebe-se que, exceto pela proporção de aproximação utilizada pelo SVD, é possível se calcular manualmente a proporção de variância acumulada, o *Strain* e ambos os *Stress* para todos. Desta forma, podemos comparar os métodos segundo as mesmas métricas além da visualização dos dados redimensionados nos espaços característicos produzidos visíveis ($k < 3$).

A seguir definiremos estas métricas de forma que possam ser calculadas para todos os métodos a serem analisados. As definições serão feitas em relação a matriz de dados originais, \mathbf{X} , e suas respectivas matrizes de distâncias ($D_{Euclidiana}(X)$ e $D_{Manhattan}(X)$). Entretanto, podem ser analogamente definidas para a matriz de dados padronizada, \mathbf{Z} , apenas substituindo uma pela outra. E utilizaremos \mathbf{Y} para representar a matriz dados transformados (e se, $k < p$, a matriz de dados na dimensão reduzida), ou seja, a matriz de escores (**Score** = \mathbf{Y}).

A proporção da variância acumulada é definida como soma das variâncias explicadas por cada característica do espaço transformado (ou reduzido) pela variância total do dados antes da transformação (JOHNSON, 2007). Conforme a Eq. 3.1.

$$\text{Proporção Variância Acumulada} = \frac{\text{Variância}(\mathbf{Y}_{n \times k})}{\text{Variância}(X)} \quad (3.1)$$

No caso do PCA obtemos o valor da proporção facilmente através dos autovalores associados as k primeira componentes principais. Entretanto, para os outros métodos a variância de cada V.A. tem que ser calcula a partir das matrizes de dados.¹⁸

Nomeia-se de *Strain* qualquer função de perda que relate o produto interno. Neste trabalho utilizaremos a função definida pela Eq. 2.24 (também chamada de *Strain* – 1 por sua similaridade com o *Stress* – 1). Como temos a matriz de originais, podemos calcular seu produto interno, $B_{\mathbf{X}} = \mathbf{XX}^T$, da mesma forma para cada base-método-dimensão podemos obter o respectivo produto interno dos dados, $B_{\mathbf{Y}} = \mathbf{YY}^T$. Vale ressaltar que o CMDS não exige que tenhamos a matriz de dados originais, e sim, apenas a matriz de dissimilaridades (no caso do métrico, a matriz de distâncias) destes.

Por fim, por termos as matriz \mathbf{Y} para todas as execuções, é possível definir a matriz de distâncias euclidianas para cada base-método-dimensão, e como, também, se tem a matriz de distância dos dados originais, podemos aplicar a Eq. 2.28 para obtenção do *Stress* – 1 e a Eq. 2.29 para o cálculo do *Sstress*. Observe que a pesar da métrica de distância da matriz de distâncias dos dados transformados (ou reduzidos) ser fixada como Euclidiana, ao seguir as definições de *Stress* utilizada pelos MDS não métricos, a métrica da distância para os dados originais é livre. Quando o *Stress* é calculado dentro do método MDS isto é definido pela matriz de distâncias informada como entrada (e o método nem saberá qual é a distância utilizada para a construção desta).

Portanto, ao utilizarmos o *Stress* para os demais métodos temos: que o *Stress* do CMDS será análogo ao do MDS, a métrica da distância da matriz de distância referente aos dados originais será a mesma informada como entrada para o execução do escalonamento clássico. Da mesma forma, no caso do MDS Stress-1 e do MDS Sstress basta calcular o *Sstress* ou o *Stress* – 1, respectivamente, utilizando a mesma matriz de distância que foi a entrada de sua execução.

Por sua vez o FastMap aceita apenas a entrada de matrizes de distâncias Eucli-

¹⁸O cálculo da variâncias das matrizes de dados foi feito pela função *var* do *MatLab 2012a* (MATHWORKS, 2012).

anas, mesmo que se saiba que as distâncias entre os dados são melhor representadas por outra métrica, por exemplo, Manhattan. Neste cenário, consideraremos que o problema trata de reduzir a dimensão de dados em um espaço mais adequado a métrica de distância utilizada pelo MDS e CMDS. E, portanto, o método é incapaz de lidar com isto e utiliza na sua execução distâncias Euclidianas, entretanto para a avaliação da qualidade das transformações será utilizada a métrica que melhor se adequa aos dados.

Analogamente, usaremos desta abordagem para o PCA e o SVD, os quais não fazem uso de matrizes de distâncias em sua solução. Consideramos que a métrica mais adequada aos dados é a utilizada pelo MDS e CMDS e que os métodos do PCA e SVD, a pesar de ter este conhecimento, desconsideraram isto e realizaram a transformação (e consequentemente a redução de dimensionalidade) se o uso das distâncias. Mas para a avaliação da qualidade dos métodos será utilizada a matriz de distâncias dos dados originais construída pela métrica mais adequada aos dados.

Ainda pode-se obter outras métricas derivadas das funções de perda do MDS não clássico métrico, *Stress*. Como descrito na Secção 2.3.2 este escalonamento tem como entrada uma matriz de distâncias e tenta obter uma matriz de dados \mathbf{Y} tal que a distância euclidiana destes seja o mais próximo possível das originais, de acordo com uma função de perda que relate as (algumas das funções de *Stress*). Nota-se que a matriz de entrada não é necessariamente uma matriz de distâncias euclidianas, podendo ser definida por qualquer outra métrica que seja uma métrica de distância (para uma definição formal de distância consulte JOHNSON (2007)).

Por mais que normalmente quando um processo de descoberta de conhecimento é realizado por um cientista da computação ou engenheiro a decisão entre estas duas seja feita na empiricamente. Para muitos casos esta decisão não é viável, seja pelo custo ou pela impossibilidade de se testar ambas. Sendo a escolha da métrica de distância uma importante questão quando analisada do ponto estatístico, que deve levar em conta diversas propriedades, estruturas e características dos dados e dos espaços.

Desta forma, pode ser interessante comparar os métodos de acordo com uma função de perda que relate os dados de entrada e os transformados de acordo com a mesma métrica de distância. Ou seja, se as distâncias de um conjunto de amostras

é definida por uma matriz de distâncias de Manhattan ter uma função de perda que relate esta a matriz de distâncias de Manhattan dos dados transformados, e não a matriz de distâncias Euclidianas dos dados transformados, conforme é feito pela função otimizada nos métodos MDS não clássicos e métricos.

Entretanto, tem-se que uma função de perda que relate distâncias é chamada de função de *Stress*, independentemente da métrica de distância. Consequentemente podemos utilizar a mesma equação do $Stress - 1$ (Eq. 2.28) e $Sstress$ (Eq. 2.29) definidas no capítulo anterior só que ao invés de utilizar para os dados transformados a matriz de distâncias Euclidianas, utilizaremos a matriz de distância segundo a mesma métricas da matriz de distância dos dados não transformados.

Para identificação destas utilizaremos as seguinte nomenclatura, seja \blacksquare a métrica de distância utilizada para os dados não transformados:

Simbolo	Significado
$Stress - 1_{\blacksquare \rightarrow Eucliana}$	Função $Stress - 1$ tradicional
$Sstress_{\blacksquare \rightarrow Eucliana}$	Função $Sstress$ tradicional
$Stress - 1_{\blacksquare \rightarrow \blacksquare}$	Função $Stress - 1$ formulação alternativa
$Sstress_{\blacksquare \rightarrow \blacksquare}$	Função $Stress - 1$ formulação alternativa

Como neste trabalho estamos utilizando apenas as distâncias Eucliana e de Manhattan, temos dois casos. Se métricas dos dados não transformados for a Eucliana ($\blacksquare = Eucliana$) as novas definições das funções de perda serão as mesmas das tradicionais. Mas se $\blacksquare = Manhattan$, teremos tanto as funções tradicionais, $Stress - 1_{Manhattan \rightarrow Eucliana}$ e $Sstress_{Manhattan \rightarrow Eucliana}$, quanto nas novas definições, $Stress - 1_{Manhattan \rightarrow Manhattan}$ e $Sstress_{Manhattan \rightarrow Manhattan}$.

Acima listamos as métricas relacionadas ao qualidade da representação dos dados no espaço reduzido (ou transformado, quando $k = p$), além de uma métrica de desempenho computacional, o tempo. Outra métrica de desempenho é a ordem de complexidade de cada algoritmo, as quais foram obtidas da literatura. Mas ainda, falta os dois principais critérios de avaliação. Qual método apresenta a melhor visualização dos dados e qual método mais auxilia, ou menos atrapalha, uma tarefa de mineração de dados ou aprendizado de máquina (por exemplo, classificação), em relação aos resultados desta quando realizada sobre os dados pré-processados por um dos métodos de redução de dimensionalidade.

Seja este pré-processamento opcional para tentar se realçar as informações importantes e melhorar o desempenho do algoritmo, ou, obrigatório devido ao volume de dados, caso em que se deseja diminuir o volume de forma que atrapalhe ao mínimo os resultados das etapas posteriores do KDD.

A comparação quanto a visualização será feita através dos gráficos dos dados reduzidos para as dimensões visíveis (1D, $k = 1$; 2D, $k = 2$; e 3D, $k = 3$), os quais serão apresentados no capítulo seguinte. A métrica utilizada para mensurar a ajuda ou o quanto atrapalhou foi a acurácia obtida na classificação dos dados reduzidos. O algoritmo de mineração utilizado foi a Árvore de Decisão sem qualquer estratégia de poda.¹⁹ Com validação cruzada 3 – *folds*, e acurácia final sendo a média aritmética de cada uma das execuções da validação cruzada.

3.4 RESULTADOS

As métricas de qualidade da representações obtidas e o tempo de execução para as seis bases de dados por cada método em todas as dimensão possíveis ($k = 1, \dots, p$) em cada cenário estão relatadas no Apêndice A, onde as linhas em vermelho indicam que o método (no caso dos de otimização, MDS Stress-1 e MDS Sstress) não convergiram para dada base na dimensão k especificada, entretanto, estes métodos mesmo não convergindo retornam a melhor configuração obtida (por mais que não seja um minimo).

Abaixo segue as tabelas (Tab. 3.4 à Tab. 3.4) com as acurárias da classificação por Árvore de Decisão utilizando validação cruzada 3 – *fold* a partir dos dados obtidos por cada um dos métodos para todas as dimensões possíveis.

¹⁹O treino e a predição da árvore de decisão foram executados através das funções *fit* e *predict* da classe *ClassificationTree* do *MatLab 2012a* (MATHWORKS, 2012).

Tabela 3.3: Acurácia da classificação da matriz de dados original e padronizada por base de dados

Base de dados	Matriz de dados não transformada	
	X	Z
Triagem de Crédito	0.84381	0.84381
Imagens	0.87143	0.86667
Iris	0.92667	0.92667
Sonar	0.92667	0.70642
Spiral	0.92667	0.52222
Vogal	0.92667	0.78182

Tabela 3.4: Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de **X** da base Triagem de Crédito

k	PCA	SVD	FastMap
15	0.80246	0.79484	0.7734
14	0.80246	0.78562	0.76727
13	0.78256	0.79021	0.77184
12	0.7704	0.78875	0.71671
11	0.73819	0.79177	0.72591
10	0.73353	0.68146	0.71824
9	0.7382	0.69983	0.6876
8	0.69377	0.68762	0.6723
7	0.69682	0.66925	0.65696
6	0.70599	0.62483	0.66614
5	0.64314	0.62633	0.64327
4	0.65239	0.65079	0.64474
3	0.66614	0.65844	0.63405
2	0.59119	0.63555	0.6524
1	0.6248	0.62176	0.62173

k	CMDS		MDS Stress-1		MDS Sstress	
	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$
15	0.80246	0.63553	0.62632	0.63093	0.5988	0.57573
14	0.80246	0.64164	0.66006	0.66615	0.58655	0.6187
13	0.78256	0.64622	0.65234	0.64014	0.49772	0.49772
12	0.7704	0.61867	0.66762	0.65537	0.60793	0.62329
11	0.73819	0.62022	0.65702	0.64166	0.61109	0.58958
10	0.7366	0.61712	0.65083	0.64624	0.61101	0.59114
9	0.7382	0.60183	0.67541	0.62786	0.62947	0.62174
8	0.69377	0.61564	0.68297	0.66918	0.637	0.605
7	0.69682	0.64928	0.65394	0.65693	0.53911	0.53911
6	0.70599	0.62939	0.66155	0.65078	0.63103	0.63708
5	0.64314	0.63092	0.66918	0.66148	0.63239	0.59875
4	0.65239	0.6447	0.65696	0.62322	0.64169	0.65097
3	0.66614	0.63401	0.62172	0.63093	0.49771	0.49771
2	0.59119	0.6279	0.60178	0.62786	0.50533	0.50533
1	0.6248	0.5973	0.55285	0.60178	0.62169	0.60644

Tabela 3.5: Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de \mathbf{X} da base Imagens

k	PCA	SVD	FastMap
18	0.80952	0.79048	0.7381
17	0.80952	0.79048	0.74762
16	0.80952	0.79048	0.74762
15	0.80952	0.78095	0.74762
14	0.82857	0.78095	0.74762
13	0.82857	0.77143	0.74286
12	0.82857	0.77143	0.74286
11	0.83333	0.74762	0.77143
10	0.83333	0.69524	0.75238
9	0.83333	0.72857	0.77143
8	0.82857	0.73333	0.77619
7	0.82381	0.74762	0.79048
6	0.77619	0.67619	0.72857
5	0.7619	0.69048	0.6619
4	0.74286	0.64762	0.68095
3	0.5	0.54286	0.50476
2	0.49048	0.50952	0.37143
1	0.3	0.23333	0.37143

k	CMDS		MDS Stress-1		MDS Sstress	
	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$
18	0.82857	0.74762	0.81905	0.77143	0.73333	0.72381
17	0.82857	0.75238	0.79048	0.75714	0.77619	0.72381
16	0.82857	0.75238	0.82857	0.77619	0.75238	0.70476
15	0.82857	0.75238	0.80952	0.79048	0.73333	0.72857
14	0.82857	0.75714	0.79524	0.76667	0.70952	0.65714
13	0.82857	0.8	0.78571	0.75714	0.7381	0.72381
12	0.82857	0.8	0.80952	0.76667	0.71905	0.74286
11	0.83333	0.73333	0.81905	0.78095	0.70476	0.73333
10	0.83333	0.71905	0.82857	0.75714	0.76667	0.74762
9	0.83333	0.72381	0.8381	0.7619	0.73333	0.76667
8	0.82857	0.72381	0.80952	0.75238	0.78095	0.7619
7	0.82381	0.72381	0.8381	0.76667	0.78571	0.74762
6	0.77619	0.73333	0.79048	0.72381	0.77619	0.77143
5	0.7619	0.73333	0.78571	0.72381	0.76667	0.74286
4	0.74286	0.73333	0.77143	0.7	0.75714	0.73333
3	0.5	0.58095	0.77619	0.67619	0.50952	0.58571
2	0.49048	0.51429	0.5381	0.55238	0.50952	0.50952
1	0.3	0.41905	0.51905	0.47619	0.4381	0.35714

Tabela 3.6: Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de \mathbf{X} da base Iris

k	PCA	SVD	FastMap
4	0.90667	0.93333	0.89333
3	0.91333	0.92667	0.90667
2	0.92667	0.94667	0.92
1	0.89333	0.79333	0.90667

k	CMDS		MDS Stress-1		MDS Sstress	
	$D_{Eucladiana}$	$D_{Manhattan}$	$D_{Eucladiana}$	$D_{Manhattan}$	$D_{Eucladiana}$	$D_{Manhattan}$
4	0.90667	0.93333	0.91333	0.94667	0.90667	0.93333
3	0.91333	0.91333	0.88667	0.94667	0.9	0.92
2	0.92667	0.91333	0.90667	0.88667	0.92667	0.89333
1	0.89333	0.87333	0.90667	0.88	0.9	0.88

Tabela 3.7: Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de \mathbf{X} da base Sonar

60	0.68737	0.66384	0.74017	30	0.67336	0.71587	0.77881
59	0.69689	0.66384	0.74017	29	0.67336	0.71587	0.78364
58	0.69689	0.66384	0.74017	28	0.68765	0.71587	0.78364
57	0.70656	0.69241	0.74017	27	0.69731	0.71594	0.77888
56	0.70656	0.69241	0.74017	26	0.69731	0.7256	0.77888
55	0.70656	0.69241	0.74017	25	0.69731	0.7256	0.77405
54	0.70656	0.69241	0.74017	24	0.69731	0.71132	0.76929
53	0.70656	0.68758	0.76432	23	0.69731	0.71132	0.76929
52	0.70656	0.68758	0.76432	22	0.72588	0.72581	0.76929
51	0.69703	0.68758	0.76432	21	0.72574	0.72098	0.76929
50	0.70186	0.72091	0.76432	20	0.72574	0.70193	0.76929
49	0.68758	0.71125	0.76432	19	0.72098	0.70193	0.76446
48	0.68758	0.71125	0.76432	18	0.72098	0.70193	0.76446
47	0.68758	0.71125	0.76432	17	0.72098	0.70193	0.75963
46	0.68758	0.71125	0.76432	16	0.72098	0.70193	0.75963
45	0.68758	0.71125	0.76432	15	0.72098	0.71159	0.77391
44	0.68758	0.71125	0.76432	14	0.72574	0.71146	0.77881
43	0.68758	0.69193	0.76432	13	0.71608	0.71146	0.77398
42	0.68758	0.69193	0.76915	12	0.67764	0.70193	0.76915
41	0.68758	0.69193	0.77398	11	0.67757	0.72098	0.76915
40	0.68758	0.69193	0.77398	10	0.69683	0.72098	0.77867
39	0.68758	0.69193	0.77398	9	0.69676	0.71132	0.7882
38	0.68758	0.69676	0.76915	8	0.70166	0.72574	0.77847
37	0.68758	0.69676	0.77398	7	0.70649	0.70642	0.78371
36	0.68758	0.69676	0.77881	6	0.70186	0.71608	0.77398
35	0.68758	0.70152	0.77881	5	0.71139	0.72084	0.75459
34	0.68758	0.70152	0.77881	4	0.69703	0.60566	0.6971
33	0.69234	0.71587	0.77881	3	0.63961	0.60083	0.62954
32	0.66853	0.71587	0.77881	2	0.66349	0.66342	0.59634
31	0.67336	0.71587	0.77881	1	0.48068	0.68758	0.58668

k	CMDs		MDS Stress-1		MDS Sstress	
	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$
60	0.68737	0.69717	0.68737	0.66805	0.69206	0.64403
59	0.69689	0.69717	0.61063	0.6824	0.68275	0.67764
58	0.69689	0.69717	0.69241	0.6873	0.67805	0.62988
57	0.70656	0.69717	0.71622	0.6156	0.67322	0.67315
56	0.70656	0.69717	0.64921	0.68247	0.68275	0.65397
55	0.70656	0.69717	0.67308	0.65852	0.70207	0.66349
54	0.70656	0.69717	0.69745	0.65411	0.64444	0.67785
53	0.70656	0.69717	0.67288	0.66798	0.65404	0.67778
52	0.70656	0.69717	0.70697	0.70179	0.67322	0.67757
51	0.69703	0.69717	0.72153	0.69206	0.69199	0.64872
50	0.69703	0.69717	0.67329	0.67778	0.68268	0.66329
49	0.68275	0.69717	0.71146	0.6441	0.67819	0.65383
48	0.68275	0.69241	0.67778	0.65832	0.67778	0.66322
47	0.68275	0.69241	0.69724	0.69689	0.68268	0.65369
46	0.68275	0.69241	0.69268	0.6873	0.67315	0.64403
45	0.68275	0.68765	0.69268	0.6873	0.70179	0.66791
44	0.68275	0.69248	0.71643	0.66322	0.68785	0.64389
43	0.68275	0.69248	0.69248	0.68247	0.70656	0.64879
42	0.68275	0.69248	0.71643	0.68254	0.71146	0.6824
41	0.68275	0.69248	0.68295	0.67757	0.66342	0.66308
40	0.68275	0.69248	0.70214	0.68247	0.69213	0.66812
39	0.68275	0.69248	0.69248	0.67274	0.67302	0.6873
38	0.68275	0.69248	0.65894	0.68723	0.71125	0.6873
37	0.68275	0.69248	0.70697	0.68723	0.69255	0.68247
36	0.68275	0.69248	0.69227	0.67302	0.72126	0.64389
35	0.68275	0.68765	0.67329	0.66329	0.67757	0.69206
34	0.68275	0.68765	0.68288	0.70145	0.73057	0.71615
33	0.68751	0.68765	0.70656	0.68709	0.7069	0.65859
32	0.6637	0.65908	0.66853	0.67764	0.7401	0.6824
31	0.66853	0.6637	0.68785	0.67771	0.73554	0.67736

k	CMDS		MDS Stress-1		MDS Sstress	
	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$
30	0.66853	0.67798	0.70676	0.66322	0.68282	0.68716
29	0.66853	0.67798	0.71153	0.70642	0.69683	0.67771
28	0.68282	0.67798	0.67792	0.69186	0.69724	0.70669
27	0.69248	0.69248	0.69234	0.65328	0.69696	0.68723
26	0.69248	0.69703	0.70669	0.66315	0.6971	0.69213
25	0.69248	0.69703	0.69724	0.69206	0.70186	0.70649
24	0.69248	0.69703	0.68282	0.70173	0.69234	0.67239
23	0.69248	0.69703	0.7069	0.70166	0.72616	0.69213
22	0.72105	0.69703	0.68268	0.69689	0.74983	0.71153
21	0.72091	0.69703	0.70179	0.69662	0.69234	0.70656
20	0.72091	0.68254	0.75003	0.70159	0.71622	0.71111
19	0.71615	0.68737	0.72574	0.70145	0.73554	0.67792
18	0.71615	0.6922	0.72105	0.65818	0.69683	0.75459
17	0.71615	0.6922	0.71159	0.68709	0.74037	0.69696
16	0.71615	0.68737	0.7499	0.71601	0.71649	0.70669
15	0.72098	0.67281	0.71139	0.70628	0.71139	0.72553
14	0.72574	0.67281	0.73099	0.70152	0.7401	0.67267
13	0.71608	0.67281	0.74983	0.6775	0.7548	0.68254
12	0.67764	0.71601	0.75963	0.75438	0.70656	0.72112
11	0.67757	0.7305	0.75024	0.72112	0.74023	0.73527
10	0.69683	0.71111	0.74044	0.68758	0.76411	0.73527
9	0.69676	0.69683	0.75473	0.75473	0.6971	0.73533
8	0.69213	0.69689	0.71594	0.73996	0.74507	0.74983
7	0.70166	0.71622	0.72091	0.73533	0.70663	0.79772
6	0.70186	0.72574	0.74976	0.75452	0.73071	0.73099
5	0.71139	0.71118	0.67302	0.76888	0.67274	0.76439
4	0.69703	0.6775	0.6637	0.66798	0.6441	0.67778
3	0.64444	0.69213	0.596	0.71118	0.59158	0.61056
2	0.64444	0.60097	0.65873	0.62043	0.56246	0.58206
1	0.48558	0.62519	0.58661	0.61546	0.58675	0.61532

Tabela 3.8: Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de \mathbf{X} da base Espiral

k	PCA	SVD	FastMap
2	0.51004	0.51004	0.54373
1	0.46667	0.46667	0.53262

k	CMDS		MDS Stress-1		MDS Sstress	
	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$
2	0.51004	0.54373	0.53226	0.55412	0.53226	0.59749
1	0.46667	0.54373	0.71792	0.72688	0.64122	0.38065

Tabela 3.9: Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de \mathbf{X} da base Vogal

k	PCA	SVD	FastMap
10	0.78081	0.78586	0.75253
9	0.77273	0.79293	0.75859
8	0.7798	0.79394	0.76364
7	0.78283	0.79192	0.75657
6	0.77071	0.77778	0.75152
5	0.78182	0.77778	0.74949
4	0.76566	0.77172	0.7404
3	0.69596	0.72525	0.67172
2	0.54949	0.59697	0.54444
1	0.33232	0.2404	0.27071

k	CMDS		MDS Stress-1		MDS Sstress	
	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$
10	0.78081	0.76162	0.78081	0.7798	0.78081	0.75051
9	0.77273	0.76263	0.78586	0.78586	0.78384	0.76162
8	0.7798	0.75253	0.78687	0.78283	0.77475	0.74242
7	0.78283	0.74646	0.80303	0.76869	0.79192	0.75556
6	0.77071	0.75152	0.77677	0.77172	0.78182	0.74444
5	0.78182	0.74141	0.78283	0.75051	0.77172	0.76465
4	0.76465	0.73838	0.76869	0.73333	0.77879	0.73838
3	0.7	0.66768	0.7303	0.6596	0.67374	0.67677
2	0.54646	0.55455	0.60303	0.59899	0.56061	0.55051
1	0.33232	0.31919	0.40808	0.40909	0.36061	0.36465

Tabela 3.10: Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de \mathbf{Z} da base Triagem de Crédito

k	PCA	SVD	FastMap
15	0.80859	0.80553	0.7657
14	0.80553	0.80553	0.7933
13	0.77488	0.77488	0.79174
12	0.79935	0.79935	0.76732
11	0.781	0.781	0.75501
10	0.7734	0.7734	0.73361
9	0.7688	0.7688	0.72284
8	0.77491	0.77491	0.72742
7	0.78564	0.78564	0.72741
6	0.79328	0.79328	0.72898
5	0.758	0.758	0.71371
4	0.77636	0.77636	0.72438
3	0.79177	0.79177	0.70599
2	0.76875	0.76875	0.7029
1	0.71357	0.71357	0.68306

k	CMDS		MDS Stress-1		MDS Sstress	
	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$
15	0.81164	0.80705	0.80706	0.79327	0.82083	0.81319
14	0.81164	0.7994	0.81164	0.79629	0.81317	0.80549
13	0.7856	0.80092	0.81167	0.81474	0.80554	0.80247
12	0.80704	0.7948	0.79938	0.81165	0.80093	0.79478
11	0.781	0.79788	0.79477	0.81777	0.79634	0.79325
10	0.7734	0.81009	0.79631	0.80394	0.76572	0.81162
9	0.7688	0.81013	0.77027	0.82235	0.78718	0.80855
8	0.77491	0.81165	0.80246	0.80701	0.78565	0.81166
7	0.78564	0.79944	0.77645	0.79777	0.77646	0.79789
6	0.79328	0.79485	0.77649	0.79174	0.76267	0.7994
5	0.758	0.76422	0.77345	0.77793	0.78713	0.77645
4	0.77636	0.76879	0.75494	0.77492	0.77638	0.76575
3	0.79177	0.75654	0.7627	0.77641	0.77953	0.74735
2	0.77028	0.77033	0.78405	0.78103	0.74585	0.77338
1	0.7105	0.7259	0.70752	0.70749	0.58659	0.67842

Tabela 3.11: Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de \mathbf{Z} da base Imagens

k	PCA	SVD	FastMap
18	0.82857	0.82857	0.76667
17	0.86667	0.86667	0.76667
16	0.86667	0.86667	0.76667
15	0.86667	0.86667	0.76667
14	0.86667	0.86667	0.75714
13	0.8619	0.8619	0.75714
12	0.8619	0.8619	0.72381
11	0.8	0.8	0.71905
10	0.80952	0.80952	0.69524
9	0.80952	0.80952	0.7
8	0.81429	0.81429	0.67619
7	0.73333	0.73333	0.68571
6	0.77143	0.77143	0.67143
5	0.74762	0.74762	0.68571
4	0.74762	0.74762	0.64286
3	0.76667	0.76667	0.61429
2	0.59524	0.59524	0.62857
1	0.53333	0.5381	0.40952

k	CMDS		MDS Stress-1		MDS Sstress	
	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$
18	0.86667	0.7619	0.8619	0.83333	0.84762	0.78095
17	0.86667	0.78095	0.85238	0.8381	0.82857	0.81905
16	0.86667	0.78095	0.8381	0.8381	0.81429	0.80952
15	0.86667	0.78571	0.8619	0.8381	0.83333	0.8
14	0.86667	0.78571	0.8619	0.8381	0.84286	0.74762
13	0.8619	0.77143	0.8619	0.8381	0.84762	0.78571
12	0.8619	0.8	0.84286	0.82857	0.82857	0.78571
11	0.8	0.79048	0.79524	0.84286	0.8	0.8
10	0.80952	0.79048	0.79048	0.83333	0.80952	0.77619
9	0.80952	0.80476	0.80952	0.8381	0.79048	0.75714
8	0.81429	0.79048	0.80476	0.85238	0.80476	0.8
7	0.73333	0.79048	0.7619	0.81905	0.78571	0.77619
6	0.77143	0.78095	0.82381	0.77619	0.7619	0.73333
5	0.74762	0.79048	0.75714	0.74286	0.70952	0.71905
4	0.74762	0.77619	0.8	0.77619	0.7619	0.75714
3	0.76667	0.75714	0.75714	0.82381	0.78095	0.71905
2	0.59524	0.60952	0.71429	0.73333	0.69524	0.65714
1	0.53333	0.53333	0.6381	0.66667	0.58571	0.51905

Tabela 3.12: Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de \mathbf{Z} da base Iris

k	PCA	SVD	FastMap
4	0.94	0.94	0.91333
3	0.92667	0.92667	0.91333
2	0.91333	0.91333	0.88
1	0.91333	0.91333	0.9

k	CMDS		MDS Stress-1		MDS Sstress	
	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$
4	0.94	0.93333	0.94	0.94	0.94	0.92
3	0.92667	0.89333	0.92	0.9	0.90667	0.92667
2	0.91333	0.88	0.88	0.88	0.9	0.87333
1	0.91333	0.89333	0.71333	0.75333	0.9	0.86667

Tabela 3.13: Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de \mathbf{Z} da base Sonar

k	PCA	SVD	FastMap
60	0.71159	0.72609	0.69676
59	0.71159	0.72609	0.69199
58	0.71159	0.72609	0.69199
57	0.6971	0.71159	0.69199
56	0.6971	0.71159	0.69199
55	0.6971	0.71159	0.69199
54	0.6971	0.71159	0.72091
53	0.6971	0.71159	0.72091
52	0.6971	0.71159	0.72091
51	0.6971	0.71159	0.72091
50	0.6971	0.71159	0.72091
49	0.72609	0.74058	0.72091
48	0.72609	0.74058	0.72091
47	0.72609	0.74058	0.72091
46	0.72609	0.74058	0.72091
45	0.72609	0.74058	0.72091
44	0.72609	0.74058	0.72091
43	0.72609	0.74058	0.72091
42	0.72609	0.74058	0.72091
41	0.72609	0.74058	0.72091
40	0.72609	0.74058	0.72574
39	0.72609	0.74058	0.72574
38	0.72609	0.74058	0.7354
37	0.72609	0.74058	0.73064
36	0.72609	0.74058	0.7354
35	0.72609	0.74058	0.7354
34	0.71643	0.73092	0.72574
33	0.71643	0.73092	0.72574
32	0.72126	0.73575	0.74003
31	0.72126	0.73575	0.74003

	CMDS		MDS Stress-1		MDS Sstress	
k	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$
60	0.71643	0.68709	0.70193	0.7352	0.71649	0.7352
59	0.71643	0.69193	0.71649	0.7207	0.70186	0.74003
58	0.71643	0.69193	0.69234	0.71104	0.72595	0.7207
57	0.70193	0.69193	0.71643	0.72553	0.71153	0.71118
56	0.70193	0.69193	0.69234	0.7207	0.72126	0.77378
55	0.70193	0.69193	0.71643	0.74486	0.73085	0.75438
54	0.70193	0.69193	0.73568	0.74003	0.72602	0.73527
53	0.70193	0.69193	0.68744	0.74486	0.69717	0.73989
52	0.70193	0.69193	0.72602	0.74486	0.73582	0.7109
51	0.70193	0.69193	0.70676	0.74486	0.74051	0.71594
50	0.70193	0.69193	0.71166	0.74486	0.72602	0.7158
49	0.73092	0.69193	0.71643	0.74486	0.70193	0.72077
48	0.73092	0.69193	0.73092	0.74486	0.73085	0.69689
47	0.73092	0.69193	0.69717	0.74479	0.69731	0.7158
46	0.73092	0.69193	0.73575	0.73996	0.71629	0.73023
45	0.73092	0.69193	0.72133	0.74003	0.72616	0.71118
44	0.73092	0.69193	0.71166	0.7352	0.71643	0.7256
43	0.73092	0.69193	0.72126	0.73506	0.74527	0.72077
42	0.73092	0.69193	0.71159	0.73513	0.72119	0.71587
41	0.73092	0.69193	0.72609	0.7207	0.74541	0.7207
40	0.73092	0.69193	0.74534	0.75445	0.72609	0.71594
39	0.73092	0.69193	0.72609	0.75921	0.72153	0.70614
38	0.73092	0.69193	0.71649	0.72077	0.73085	0.70138
37	0.73092	0.69193	0.70676	0.73989	0.72609	0.75921
36	0.73092	0.69193	0.702	0.74472	0.69227	0.7109
35	0.73092	0.69193	0.71649	0.7254	0.74044	0.7303
34	0.72126	0.70152	0.72126	0.73513	0.72139	0.70635
33	0.72126	0.70635	0.70214	0.70642	0.73575	0.70614
32	0.72609	0.70159	0.73085	0.71608	0.71636	0.71594
31	0.72609	0.70642	0.702	0.74479	0.71643	0.71111

k	CMDS		MDS Stress-1		MDS Sstress	
	$D_{Eucladiana}$	$D_{Manhattan}$	$D_{Eucladiana}$	$D_{Manhattan}$	$D_{Eucladiana}$	$D_{Manhattan}$
30	0.72609	0.70642	0.70642	0.71125	0.7403	0.7207
29	0.72609	0.70642	0.69696	0.73499	0.71643	0.73037
28	0.72609	0.72091	0.74023	0.74479	0.74541	0.72063
27	0.72609	0.72091	0.74997	0.74962	0.72139	0.72567
26	0.73092	0.72091	0.75466	0.77854	0.70676	0.73989
25	0.73092	0.72091	0.71173	0.75452	0.71643	0.73023
24	0.73092	0.73996	0.72602	0.77847	0.70193	0.7207
23	0.73568	0.73996	0.70207	0.77847	0.73078	0.73527
22	0.74044	0.73996	0.69717	0.75431	0.71643	0.74003
21	0.73561	0.73996	0.72581	0.75438	0.72126	0.74493
20	0.73561	0.73996	0.74037	0.73989	0.7452	0.77384
19	0.73561	0.73996	0.77391	0.74479	0.71622	0.71587
18	0.73078	0.72547	0.74023	0.77854	0.73071	0.71601
17	0.7403	0.7303	0.72091	0.77861	0.75942	0.76425
16	0.75459	0.7303	0.71629	0.73561	0.76894	0.75956
15	0.75459	0.7303	0.72581	0.77391	0.77378	0.75459
14	0.75459	0.7303	0.70635	0.75459	0.78344	0.74941
13	0.7499	0.73513	0.74023	0.71159	0.75452	0.75908
12	0.7499	0.73513	0.71649	0.75956	0.76418	0.74948
11	0.7499	0.73527	0.7499	0.77405	0.78351	0.75431
10	0.76908	0.75914	0.75466	0.73057	0.80759	0.81712
9	0.76901	0.77364	0.75942	0.75956	0.80283	0.74969
8	0.77391	0.76881	0.70663	0.69206	0.74513	0.7354
7	0.77391	0.78337	0.70656	0.68723	0.76922	0.7354
6	0.7401	0.77854	0.72595	0.72595	0.7256	0.72091
5	0.73057	0.75942	0.72098	0.745	0.69717	0.81222
4	0.74017	0.73989	0.70193	0.71146	0.76439	0.77398
3	0.71608	0.75914	0.69717	0.75452	0.69731	0.73547
2	0.62519	0.5285	0.66825	0.59627	0.61049	0.54327
1	0.52857	0.47122	0.56211	0.61028	0.58171	0.55721

Tabela 3.14: Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de \mathbf{Z} da base Espiral

k	PCA	SVD	FastMap
2	0.57563	0.59713	0.54373
1	0.54373	0.54373	0.53262

k	CMDS		MDS Stress-1		MDS Sstress	
	$D_{Eucladiana}$	$D_{Manhattan}$	$D_{Eucladiana}$	$D_{Manhattan}$	$D_{Eucladiana}$	$D_{Manhattan}$
2	0.57563	0.60932	0.619	0.57634	0.57563	0.68566
1	0.54373	0.60968	0.57634	0.73943	0.62975	0.36989

Tabela 3.15: Acurácia da classificação dos dados obtidos pelos métodos para cada dimensão a partir de \mathbf{Z} da base Vogal

k	PCA	SVD	FastMap
10	0.7697	0.7697	0.74949
9	0.76566	0.76465	0.76768
8	0.76263	0.76465	0.74949
7	0.76061	0.76162	0.75051
6	0.77071	0.77172	0.73838
5	0.76869	0.7697	0.67576
4	0.73232	0.73232	0.65455
3	0.60404	0.60404	0.56869
2	0.51414	0.51414	0.43636
1	0.17677	0.17677	0.22929

k	CMDS		MDS Stress-1		MDS Sstress	
	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$	$D_{Euclidian}$	$D_{Manhattan}$
10	0.76869	0.76061	0.76869	0.76465	0.7697	0.74343
9	0.76364	0.76667	0.76566	0.76667	0.77071	0.74949
8	0.76263	0.76869	0.78788	0.75657	0.77677	0.78384
7	0.76162	0.76667	0.77677	0.7798	0.79293	0.77576
6	0.77172	0.76566	0.76465	0.78485	0.77475	0.77172
5	0.76869	0.75657	0.74848	0.75455	0.76869	0.7697
4	0.73131	0.74141	0.75152	0.73838	0.71515	0.74141
3	0.60303	0.63333	0.64545	0.64646	0.59192	0.60101
2	0.51515	0.5	0.5596	0.55556	0.46162	0.49798
1	0.17677	0.21313	0.42222	0.37677	0.26667	0.24747

4 Análise Comparativa

Neste capítulo realizaremos uma análise comparativa dos métodos apresentados baseada nas métricas definidas na Secção 2 em relação aos resultados destes para cada uma das bases de dados utilizadas. Para melhor visualização dos dados da tabela de resultados (A) foram construídos gráficos, os quais estão dispostos no Apêndice B.

4.1 COMPLEXIDADE E TEMPO DE EXECUÇÃO

Iniciaremos comparando os métodos quanto ao custo computacional, de acordo com as duas métricas mais tradicionais para isto: a ordem de complexidade e o tempo de execução de cada um.

A Tab. 4.1 apresenta as complexidades para cada um dos métodos de redução de dimensionalidade abordados de acordo com a literatura. Onde, n e p são, respectivamente, o número de amostras e dimensões dos dados. Vale ressaltar que se o CMDS for executado como uma otimização (o que não é o caso deste trabalho) e não pela forma analítica, sua complexidade será a mesma dos outros métodos de escalonamento multidimensional, os quais, sendo resolvidos como uma minimização pelo método do gradiente tem a mesma complexidade independentemente da sua formulação da função de perda serem clássicos, não clássicos, métricos ou não métricos.

Tabela 4.1: Complexidade dos métodos

Método	Ordem de Complexidade
PCA	$O(p^3)$ (MAATEN, 2009)
SVD	$O(p^3)$ (NAVAB, 2014)
CMDS analítico	$O(n^3)$ (MAATEN, 2009)
MDS otimização	$O(n^3)$ (TZENG, 2008)
FastMap	$O(n)$ (FALOUTSOS, 1995)

Portanto temos que exceto o FastMap, o qual é linear, todos os demais tem complexidade cúbica. Nota-se que os métodos que recebem por entrada a matriz de distância, isto é, os de escalonamento (CMDS, MDS e FastMap), tem sua complexidade

em relação ao número de amostras, enquanto a do PCA e do SVD, os quais trabalham com a fatoração da matriz de variância-covariância ou de dados, é referente ao número de dimensões.

Tradicionalmente bases de dados apresentam quantidade de itens muito maior que o número de dimensões ($n >> p$), neste casos os métodos baseados em fatoração apresentam uma vantagem quanto a complexidade, podendo até mesmo se equiparar ao FastMap, que apesar de linear, sendo n muito grande e p suficientemente menor pode apresentar um desempenho pior que o PCA e o SVD. Entretanto, um uso comum da redução de dimensionalidade é exatamente quando esta relação de o número de amostras ser muito, ou apenas, maior que o de características. Caso no qual algoritmos de mineração de dados e aprendizado de máquina geralmente apresentam dificuldades em obter bons resultados (HAN, 2006). Sendo assim o cenário muda de figura quanto, e os métodos de projeções, mesmos os cúbicos, são assintoticamente melhores.

O Apêndice A apresenta os tempos de execução mensurados para cada método. Ressaltando que no caso do PCA, SVD e CMDS, por terem a propriedade de obter dimensões aninhadas, só é necessário executar uma única vez por base de dado, enquanto para o MDS e o FastMap deve-se executar para cada dimensão da base. Sendo que o MDS por ser uma otimização pode não convergir ou não encontrar um ótimo absoluto e pela matriz de soluções iniciais ser aleatória, pode a cada execução para uma mesma base e dimensão apresentar resultados distintos (o tratamento que demos para isto está descrito na Secção 3.2).

Para facilitar a comparação estes dados de todos os métodos foram plotados em um gráfico do tempo de execução pelo número de dimensões para cada par base de dado cenário (Fig. B.2 à Fig. B.7). A Fig. B.1 apresenta a legenda da representação de cada método.

Analisando os gráficos Fig. B.2 à Fig. B.7 nota-se que os algoritmos MDS Stress-1 e MDS Sstress apresentam tempos de execução extremamente piores que os do PCA, SVD, CMDS e FastMap, mesmo para bases com n pequeno e para as com p grandes, independentemente do valor de k . Mesmo para as menores bases, Iris e Espiral, onde conseguiram seus melhores tempos, a discrepância destes para os do outros resultados é

relativamente grande, de 0.5 à 5 segundos para o MDS Stress-1 e de 1 à 50 segundos. E os piores resultados alcançaram a casa de 400 e 1500 vezes mais lento (base Vogal), respectivamente.

Nota-se que entre os dois escalonamentos não clássicos, o MDS Sstress foi o pior, tendo resultados da casa de 500 vezes mais lentos que o MDS Stress-1, do qual só se aproximou na base Sonar, conseguindo resultados ligeiramente melhores em 16 dimensões no cenário $\mathbf{XeD}_{Eucladiana}$, mas mesmo nesta base cenário e base foi para maioria dos casos e na média pior.

Creditamos estes resultados ao fato do MDS para ambas funções de perda serem executados como um problema de otimização, podendo notar que ao comparar os resultados dos próprios métodos, as dimensões que não convergiram são as com tempo mais alto. É devido a grande variação entre o desempenho destes métodos para as diversas dimensões de uma mesma base em um mesmo cenário. Isto indica que se a otimização a ser realizada para uma determinada dimensão de uma base em um determinado cenário e principalmente inicialização da solução \mathbf{Y} foi mais difícil, sendo necessário utilizar mais iterações do método do gradiente para se obter a solução ou até mesmo não convergir. A grande variação dentre dimensões próximas mostra o quanto estes métodos são dependentes de quão boa foi sua inicialização randômica (o que não é uma dependência agradável, que visto serem aleatórias podem tanto ajudar quanto atrapalhar a minimização a cada nova execução).

Os gráficos da base Triagem de Crédito no cenário com a matriz de dados não padronizada são uma boa ilustração destas afirmação. Pois neles os métodos MDS atingem tempos muito altos e tempos que nesta escala até parecem próximos dos demais métodos. Mas recorrendo as tabelas de resultados (A) percebe-se que a diferença deles ainda é relativamente alta, aproximadamente 3s.

Como a escalas impostas pelos resultados discrepantes do MDS Stress-1 e MDS Sstress não permitirem uma visualização adequada dos tempos dos demais métodos, os gráficos acima foram novamente plotados mas apenas para o PCA, SVD, CMDS e FastMap (Fig. B.8 à Fig. B.13).

A primeira informação que obtemos destas plotagens, sem os métodos baseados

em otimização, é que o CMDS foi mais lento que o PCA, SVD e o FastMap em todos as bases e cenários, exceto para a base Espiral no cenário **X** e **D_{Euclidiana}**, onde sua execução foi instantânea ($0s$). Temos então, que os três escalonamentos multidimensionais, foram os piores métodos no critério tempo de execução. O custo da parte do CMDS não se deve a tarefa de otimização ou a dependência de uma matriz de soluções iniciais boa, como os dois MDS, e sim a manipulação das matrizes de produtos internos $B_X = \mathbf{X}\mathbf{X}^T$ e $B_Y = \mathbf{Y}\mathbf{Y}^T$, tanto que os maiores tempos foram para as bases com mais amostras, chegando a aproximadamente $6s$ para a base Vogais (com 990 amostras) e o menor para a Espiral ($n = 92$).

Entretanto, por mais que dos métodos sem otimização o CMDS ter sido o pior, vale lembrar que as dimensões obtidas por ele são aninhadas, ou seja, em uma base de $p = 100$ só é necessário executá-lo uma única vez para gerar as cem representações, e por isto o seu gráfico de tempo de execução é constante. Esta características de obter espaços aninhados só é vantajosa em casos onde a dimensão desejada seja não seja pré-definida, logo, será necessário gerar as representações para diversas dimensões.

O FastMap é um método versátil pois apesar de ter a propriedade de obter dimensões aninhadas, o que é uma vantagem no caso citado de se quererem todas as dimensões²⁰. Ele também permite que se escolher o k desejado, não obrigando, como os demais métodos que obtém dimensões aninhadas (PCA, SVD e CMDS) que o custo de uma representação para uma dimensão apenas seja o custo de obter-se todas.

Para estudar esta características do método que ao invés de aproveitarmos seu aninhamento de dimensões executamos-o para cada uma separadamente. Observado os gráficos tempo de execução destas foram na maior parte $0s$ e o segundo tempo que mais ocorreu foi o tão pequeno quanto $0,01s$. Os únicos casos que o FastMap obteve resultados maiores que $0,01s$ foi para a dimensões $k = 17$ da base Imagens ($p = 18$), com o resultado de $0,02s$, e para a base Sonar ($p = 60$) na dimensão $k = 51$ dos dois cenários com a matriz original (\mathbf{X}), onde o tempo gasto foi de $0,5s$.

Inicialmente pode-se pensar que este crescimento é devido as dimensões, entretanto o fato do tempo de execução do método para as dimensões vizinhas, que obrigatori-

²⁰para se obter todas as dimensões por uma única execução do FastMap deve executá-lo com o k para a maior dimensão

amente tem que executar o as dimensões que apresentaram um tempo maior, serem $0s$ ou $0,01s$ mostra que esta variação de $0s$ à $0,05s$ é algo externo ao algoritmo do método, que só foi identificada porque ele foi executado p vezes. A fim de corroborar esta verificação executamos o PCA diversas vezes e observou-se a mesma variação. Portanto, em síntese, o FastMap obteve tempo ótimo de aproximadamente $0s$ para todos os casos, e ainda tem a possibilidade de ser utilizado exclusivo para uma dimensão ou obtendo todas de forma aninhada.

Igualmente ao FastMap o PCA também executou os experimentos em $0s$ para todos os casos, fora para a base de dados com maior dimensionalidade, a Sonar ($p = 60$). Como vimos, este método tem complexidade cúbica em relação ao número de dimensões, e não possibilita, como o FastMap, ser executado exclusivamente para a desejada. Mas, estes tempos diferentes de $0s$ não foram ruins, mesmo em uma base de alta dimensão, sendo $0,01s$ para os dois cenários **Z** e $0,04s$ para os cenários **X** (já desconsiderando a variação da máquina).

Percebe-se, então que a piora da análise de componentes principais das outras bases com dimensão de quatro à dezoito para a Sonar, com sessenta dimensões, foi pequeno. E que o que mais influenciou em uma piora de seu desempenho foi ter que extrair a média dos dados, e portanto seu pior resultado foi para o cenário onde era realizado a partir da matriz de dados original.

Por sua vez o SVD não tem esta necessidade, mas nos testes apresentou um tempo de $0,01s$ ligeiramente maior que o do PCA, exceto para o caso no qual a remoção das médias fez que o segundo perde-se em desempenho. Do restante a análise do SVD segue a mesma do PCA.

Desta forma, temos que avaliando os métodos de acordo com o seu tempo de execução o PCA, FastMap e o SVD são os vencedores, em ordem. Superando o CMDS e de longe os MDS Stress-1 e o MDS Sstress. Entre os três o FastMap tem a vantagem de ter a possibilidade de ser executado para um k fixo economizando parte do custo compulsório as demais. Mas o PCA apresentou o melhor resultado absoluto mesmo estando implementado em uma linguagem mais lenta que o FastMap. E demonstrou que apesar de sofrer influencia da dimensionalidade, está não foi representativa (para as bases

analisadas). E então ele seria uma opção tão boa quanto ou melhor q o FastMap onde a economia descrita acima não seja vantajosa. Sobrando para o SVD, apenas os casos onde a matriz de dados de entrada não é normalizada ou padronizada e a remoção da média possa ser custosa ao PCA.

Vale lembrar, que se $p \gg n$, o PCA e o SVD iram piorar, mas o FastMap continuara sendo uma opção valiosamente rápida, este caso o CMDS se torna um método que vale ser considerado, principalmente quando a possibilidade de economia do FastMap não for uma vantagem e o n for pequeno, para que a linearidade do último não supere o complexidade cúbica em relação ao número de amostras do CMDS.

E, por fim, os dois escalonamento não clássicos não se mostraram interessantes devido a sua otimização, sendo que a função de perda *Sstress* foi a que gerou maiores dificuldades de custo computacional para a minimização.

4.2 PROPORÇÃO DA VARIÂNCIA ACUMULADA

A primeira métrica de qualidade da representação obtida pelos métodos que analisaremos é a proporção da variância acumulada em uma representação k dimensional ($k \leq p$). Conforme vamos observar a métrica de distância utilizada para os dados de entrada influência de forma drástica os resultados dos métodos que a levam em conta, estes são: CMDS, MDS Stress-1 e MDS Stress-2.

Lembrando que o PCA e o SVD não utilizam distâncias e portanto obtém exatamente o mesmo resultado tanto para os cenários com $D_{Euclidian}$ quanto os com $D_{Manhattan}$, e por sua vez o FastMap, a pesar de utilizar como entrada a matriz de distância, só aceita a métrica Euclidiana. Ou seja, em um cenário onde se saiba que a melhor métrica para a distância das amostras é a Manhattan, o FastMap desconsiderará isto e utilizará (ou calculará a partir da matriz de dados) a matriz de distâncias Euclidianas, e desta forma gera exatamente as mesmas representações independentemente da métrica de distância mais adequada aos dados. Ainda temos que o CMDS quando em cenários onde recebe como entrada a matriz de distâncias Euclidiana obtém representações equivalentes ao PCA e portanto a proporção da variância acumulada nestes casos será a mesmas para os dois métodos.

Analisemos, então, os cenários onde supomos que a distância mais adequada é a Euclidiana. Os dois métodos do MDS conseguem capturar uma proporção da variância acumulada muito maior com dimensões muito menores que os demais métodos, tendo alcançados 100% antes da 1/4 das dimensões para todas as bases de dados. Enquanto o PCA (consequentemente o CMDS), o SVD e o FastMap demoram, na maioria dos casos, pelo menos 1/4 (chegando a necessitar da metade ou até 2/3) das dimensões para se aproximarem da explicação total da variabilidade dos dados. Uma questão vem a tona a partir destas observações: como isto é possível, se o PCA tem exatamente por objetivo otimizar a explicação da variabilidade dos dados?

Dois fatores fazem com que o MDS consiga capturar com menos dimensões mais da variância dos dados. O primeiro é que o escalonamento multidimensional tanto o não clássico quanto o clássico (quanto as distâncias não são Euclidianas) não são métodos ortogonais, diferentemente do PCA, SVD e FastMap. Os quais tem a restrição de que seus eixos coordenados sejam perpendiculares (lembre-se isto é uma restrição do problema de otimização que define o PCA, Secção2). A segunda explicação é que por serem um método não linear conseguem expressar relações não lineares entre as V.A. do sistema.

Comparando as duas funções de perda do MDS nota-se que a *Stress* – 1 domina a *Sstress* para todos os experimentos. Em relação a dominância entre os demais métodos ortogonais não lineares, temos que o FastMap é o que apresentou as piores proporções, exceto para as bases as representações unidimensionais ($k = 1$) das bases Vogal e Iris, e para as três primeiras representações ($k=1,2,3.$) na base Imagens, mas isto apenas quando a matriz de dados utilizada era a original (\mathbf{X}). Pois nos casos onde foram executados para os dados padronizados o SVD se tornou equivalente ao PCA (e ao CMDS). Isto indica que a discrepância entre eles para os dados originais era o fato do PCA e CMDS removerem (mesmo que implicitamente) a média dos dados, e como a média da matriz \mathbf{Z} é zero, pois ela é removida durante a padronização Escore-z, os métodos acabam se equivalendo.

Considerando, agora, os cenários com a distância de Manhattan, vemos que a equivalência do SVD ao PCA para os dados padronizados é mantida. E a dominância do PCA sobre o FastMap e sobre os SVD, também, nos casos onde foram aplicados a matriz de dados original. Tal qual as relações entre o SVD e o FastMap. Isto, como foi dito se

deve, a estes três métodos não levarem em consideração a métrica da distância e portanto terem exatamente o mesmo resultado para os cenário com métrica Euclidiana ou não.

Por sua vez o CMDS é um método de escalonamento multidimensional, e quando não recebe uma matriz de distâncias euclidianas perde sua equivalência ao PCA, e apresenta as mesmas características descritas para os outros escalonamentos multidimensionais, conseguindo explicar mais da variabilidade com menos dimensões. Note que estes três métodos também alcançam proporções da variabilidade acumulada maiores que 1 (100%), o que inicialmente parece um erro se deve ao objetivo destes métodos ser obter uma matriz de dados qualquer que suas distâncias (no caso dos MDS) ou seus produtos internos (para o CMDS) sejam o mais próximo possível das dos dados de entradas. Isto fará com que a representação obtida realce a dissimilaridade entre pontos, ocasionando em um aumento da variância entre eles. Mas até este ponto, é o mesmo motivo deles capturarem em menos dimensões mais variabilidade nos cenários com a distância Euclidiana.

Informalmente, a distância de Manhattan (ou *city block distance*) é a distância entre os pontos quando só se pode andar paralelamente a alguns dos eixos coordenados, da mesma forma que a locomoção pelos quarteirões de uma cidade (*city blocks*) tal qual Manhattan. Como vimos o PCA é considerado uma rotação no espaço dos dados, e é uma transformação linear, isto é, preserva as operações de adição e multiplicação escalar entre o espaço fonte e o final. Agora tomemos a heurística utilizada pelo FastMap, ele pega para ser o eixo de projeção dos pontos o segmento que liga os dois pontos mais distantes, tentando desta forma não exata conseguir acertar qual seria a direção de maior variabilidade. Por sua vez o MDS não clássico tentará encontrar uma representação em menor dimensão onde os pontos tem a mesma distância que na representação de entrada, ou seja, no caso da métrica Manhattan, onde a distância Euclidiana entre eles seja a de Manhattan (diretamente ligada a variabilidade) entre os pontos em todas as direções. E para isto utilizara-se de transformações não lineares e amplificara a variância dos dados. De maneira análoga temos para o CMDS, substituindo a distância pelo produto interno, e lembrando que o produto interno referente aos dados de entrada ($B_{\mathbf{X}}$) é obtida a partir da matriz de distâncias informada ($D_{\mathbf{X}}$).

Comparando os escalonamentos multidimensionais entre si temos da mesma forma

que quando realizados a partir dos dados originais o a função $Stress - 1$ é mais rápida em capturar a variabilidade em relação a função $Sstress$, e ambas sobre o CMDS (o qual otimiza analiticamente a função de $Strain$). Entretanto eles atingem ao final um quantidade de variabilidade maior em ordem oposta, mais pelo CMDS, passando pelo MDS $Sstress$ e menos pelo MDS $Sstress-1$.

Ainda é necessário dizer que a extração do 100% de variabilidade explicada pelo MDS $Sstress$ para a base Triagem de Crédito no cenário **X** e **D_{Euclidiana}** é devida a um ajuste não perfeito durante a otimização e não dos motivos acima descritos.

Concluindo, temos que apesar do PCA otimizar a explicação da variabilidade dos dados ele tem uma restrição de ortogonalidade e é baseado em combinações lineares e portanto os métodos MDS conseguem capturar mais da variabilidade. E portanto são superiores neste critério. E o PCA é o melhor dentre os demais, exceto nos casos em que se igualam. E o CMDS para a distâncias Euclidianas pode ser uma alternativa ao PCA, se os dados não forem favoráveis a este quanto a sua complexidade (isto é, $p >> n$), e vice e versa.

4.3 STRAIN

A primeira observação que percebemos nestes gráficos é que as equivalências do SVD e do PCA para os dados padronizados é mantida, tão qual a esperada equivalência entre o PCA e o CMDS para distâncias Euclidianas. E que desconsiderando as situações onde os métodos do MDS não convergiram eles são aproximadamente equivalentes ao CMDS nos cenários com métrica de distância Euclidiana. Pois eles convergem para o valor de $Strain$ atingido pelo CMDS ante de um décimo das dimensões para todas as bases de dados.

De todos os métodos o FastMap é o que apresenta os piores resultados, com uma curva crescente e valores muito maiores na maioria dos cenários. Exceto para a base de dados Imagens com matriz de dados original, onde inicialmente apresentou um decrescimento do $Strain$, seguido de um crescimento e um a estabilização jundo com o $Strain$ do PCA e, nos cenários onde a distância é Euclidiana, junto aos demais escalonamentos.

Observa-se, também, que o CMDS é o método que converge mais rapidamente (em

menos dimensões) para um valor de *Strain*. Entretanto sabemos que o objetivo do CMDS é otimizar o *Strain*, e faz isto através de sua solução analítica (no caso), não estando sujeito a erros de minimização. Portanto, como na maioria dos casos ele apresenta o pior resultado quando a métrica é a de Manhattan, fora as vezes que supera apenas o FastMap ou se iguala ao PCA, MDS Stress-1 e MDS Sstress.

Conjectura-se que isto é devido a ele tentar minimizar o *Strain* a partir dos produtos internos obtidos da matriz de distâncias de Manhattan. E o PCA, SVD e o FastMap desconsiderarem isto. E os métodos MDS não clássicos por tratarem diretamente com a distância e não extraírem antes dela a matriz de produtos internos. Insto ainda, nos indica um outro ponto que pode ocasionar o resultado não esperado para o CMDS para a distância de Manhattan. Lembre que a teoria de conversão da matriz distâncias na de produto interno é baseada na suposição de que as distâncias de entrada são euclidianas, por mais que não obrigue que a sejam (como o FastMap que exige ou gera números complexos). E portanto no caso de não o ser, esta conversão fica prejudicada.

Mas um ponto a ser investigado em trabalhos futuros é o fato de o SVD conseguir um valor de *Strain* inferior ao CMDS em quase todos os casos, e nos que isto não acontece eles são equivalentes.

Finalmente, temos que o *Strain* não parece ser uma boa métrica de comparação entre os diferentes métodos, mas que para ela o que obteve mais sucesso foi o SVD, seguido do CMDS para cenários onde a distância é Euclidiana, pois converge em menos dimensões, e consequentemente junto com o CMDS nestes caso fica o PCA, pois com a métrica Euclidiana eles são equivalentes. Para os casos onde a distância considerada é a Manhattan o PCA fica em segundo lugar após o SVD seguido pelos MDS não clássicos. Vale indicar que estes resultados podem ter influencia da real inadequação da distância de Manhattan para as bases utilizadas.

Quanto ao FastMap por apresentar grande variação dos seus resultados, os quais não são em maioria bons, fica sendo considerado o pior método para esta métrica, por ser na maior parte ruim e não confiável.

4.4 STRESS-1

A seguir apresentaremos a análises tanto para a métrica do *Stress* – 1 tradicional utilizada na minimização do MDS Stress-1, a qual considera a matriz de distâncias dos dados não transformados na métrica informada, mas fixa a métrica da matriz dos dados transformados para a Euclidiana. E em seguida, faremos as comparações equivalentes para esta métrica na formulação alternativa desenvolvida neste trabalho. Sendo que como quando a métrica do dados de entrada é a Euclidiana, as duas formulações são idênticas, a segunda parte desta secção considerou apenas os cenários onde a matriz de distâncias de entrada são dadas pela métrica de Manhattan.

4.4.1 Métrica Euclidiana Para Matriz De Distância Dos Scores

Os resultados da métrica *Stress* – 1 foram pererecados para todas as bases e cenários. E o método MDS Stress-1 apresentou sempre o melhor resultado, como esperado. Exceto para a dimensão $k = 1$, onde foi superado dez vezes pelo MDS Sstress, destas sendo duas pelo CMDS, também. Entretanto o valor desta diferença que o fez ser pior que os outros é muito pequena, menor que o valor do erro permitido ao sua técnica de minimização. Portanto, considerando a margem de erro de 1^{-4} definida para o método do gradiente o MDS Stress-1 obteve os melhores resultados em todos os experimentos, sendo que o MDS Sstress e o CMDS empata com ele em algumas situações, para $k = 1$. E para dimensões maiores (o quarto mais próximo de p) em certos pares bases-cenários estes métodos voltam a obter resultados equivalentes, juntamente com os que vinham tendo representações piores sempre.

Após o MDS Stress-1 o método que a curva domina os demais em todos os experimentos é o MDS Sstress, mostrando a intima relação entre as duas funções de perda. Seguidos de perto pelo escalonamento multidimensional restante, o CMDS. Sendo que as equivalências dos métodos CMDS e PCA para distâncias Euclidianas e do SVD e PCA para os dados padronizados foram novamente verificadas.

Nas situações onde o SVD não é equivalente ao PCA ele foi, na maior parte ligeiramente inferior ao FastMap. Mas quando gerou o mesmo resultado que a analise de componentes principais foi melhor, mas na mesma grandeza. Logo, o PCA foi o melhor

método para todos os casos, dentre estes três.

Vale ressaltar que os péssimos resultados do MDS Sstress para a base de dados Triagem de Crédito, nos cenários onde a matriz de dados não era padronizada, se deveram ao não convergência para estas dimensões.

Um detalhe interessante quanto a boa parte das curvas do CMDS é que ela após atingir um valor bem pequeno de *Stress – 1*, na metade das dimensões menores, começa a apresentar um leve crescimento de aproximadamente 0,1 no seu valor para a métrica.

Destas observações temos, que referentemente ao *Stress – 1* o melhor método é o MDS Stress-1 sequiado de perto pelo MDS Sstress e do CMDS, o qual deve ser tomar cuidado pois não melhora com o aumento do número de dimensões sempre. E dos demais, o PCA é a opção menos pior.

4.4.2 Mesma Métrica da Matriz de Distância De Entrada Para Matriz De Distância Dos Scores

Quando a métrica da matriz de distância de entrada é Euclidiana, o *Stress – 1* é o mesmo em ambas as formulações, pois a para a métrica na noma formulação para a matriz de scores sendo a mesma da de entrada está será Euclidiana, tal qual já o é na formulação tradicional. Portanto só iremos mostrar os gráficos para os cenários **X** e **D_{Manhattan}** e **Z** e **D_{Manhattan}**, onde a métrica das distância da matriz de entrada é Manhattan (Fig. B.32 à Fig. B.37), visto que para os outros dois cenários os gráficos são os mesmos que os (a) e (c) das Fig. B.26 à Fig. B.31.

Notou-se que a modificação da formulação, não modificou os resultado comparativos, e nem mesmo boa parte dos valores obtidos. Sendo as relações para o *Stress – 1* na formulação alternativa exatamente os mesmos que os obtidos pela análise do na formulação tradicional. E portanto, corrobora-se esta métrica como valida mas que não acrescenta informações que não sejam obtidas da forma anterior.

4.5 SSTRESS

Da mesma forma que na secção anterior, os resultados para a métrica *Sstress* serão apresentados primeiramente para a formulação tradicional e posteriormente para a alternativa, sob as mesmas condições que foram feitas as análises para os duas formulações do *Stress – 1*.

4.5.1 Métrica Euclidiana Para Matriz De Distância Dos Scores

De maneira similar aos resultados para o *Stress – 1* os para o *Sstress* mostraram seguir o mesmo padrão para as diversas bases e cenários. Mas diferentemente, do anterior, aqui, o método que optimiza a função de perda, o MDS *Sstress* obteve resultados melhores em todos os casos, só sendo superado quando não convergiu, e mesmo na maior parte das vezes que não conseguiu obter a solução ótima ainda assim teve o melhor resultado. Sua curva dominou a de todos os demais métodos, os quais só conseguiram alcançar quando k muito próximo de p .

Reforçando ainda mais a inter-relação entre as duas funções de perda do escalonamento multidimensional não clássico, os resultados do MDS Stress-1 foram os segundo melhores para esta métrica e se aproximaram bastante dos valores obtidos pelo MDS *Sstress*, por mais que só chegasse a empatar em dimensões altas.

As demais comparações dentre os outros métodos seguem as mesmas descritas pela a análise do *Stress – 1* e não carecem de serem repetidas. Exceto por alguns resultados do CMDS o qual apresentou uma variação um pouco maior, conseguindo superar o MDS Stress-1 para a base Imagens no cenário **X** e **D_{Manhattan}**. Mas, também, sendo o pior para a base Triagem de Crédito no caso de **Z** e **D_{Euclidiana}**. Outro fato relevante quanto ao desempenho do escalonamento clássico é que o crescimento que apresenta em certas situações após atingir um baixo valor para a métrica *Sstress*, da mesma forma que ocorreu para a *Stress – 1*, foram maiores, chegando a piorar em 0,2 o seu resultado.

Temos, então que em relação ao *Sstress* as diretrizes obtidas são as mesmas das do *Stress – 1*, com o acréscimo da ressalva ao CMDS, por ter perdido em robustez e ser menos confiável, quando avaliado por esta perspectiva.

4.5.2 Mesma Métrica da Matriz de Distância De Entrada Para Matriz De Distância Dos Scores

Analogamente ao *Stress*–1, quando a métrica da matriz de distância de entrada é Euclidiana, o *Sstress* é o mesmo em ambas as formulações, pois a para a métrica na noma formulação para a matriz de scores sendo a mesma da de entrada está será Euclidiana, tal qual já o é na formulação tradicional. Portanto só iremos mostrar os gráficos para os cenários **X** e **D_{Manhattan}** e **Z** e **D_{Manhattan}**, onde a métrica das distância da matriz de entrada é Manhattan (Fig. B.44 à Fig. B.49), visto que para os outros dois cenários os gráficos são os mesmos que os (a) e (c) das Fig. B.38 à Fig. B.43.

Igualmente ao caso da função de perda *Stress* – 1 na formulação tradicional e alternativa, os resultados do *Sstress* alternativo são comparativamente os mesmo que o tradicional, inclusive com a maior parte dos valores iguais. E as relações entre os métodos são as mesmas descritas acima (Secção 4.5.1).

Sendo o MDS *Sstress* o melhor método, seguido de perto pelo MDS *Stress*-1 e fazendo a mesma ponderação sobre o CMDS. Portanto, é evidente que as formulações alternativas do *Stress* – 1 e do *Sstress* não contribuem com informações não antes possíveis de serem conquistadas a través das formulações tradicionais. Talvez isto esteja ligado ao fato da nova formulação só ter sido aplicado a métrica de distância Manhattan, seria valido, em trabalhos futuros, realizar a verificação de sua validade e utilidade para outras distâncias.

4.6 ACURÁCIA DO CLASSIFICADOR

A principal métrica para avaliação dos métodos de redução de dimensionalidade quando empregados na etapa de pré-processamento em um KDD, é o quanto melhorou ou não piorou os resultados das estapas seguintes, sendo a extração (o reconhecimento) de padrões nos dados a mais importantes destas. Portanto a acurácia (ou outras métricas de avaliação) da mineração realizada é o grande indicador da adequação do método de redução como a base, cenário e etapas subsequente do KDD.

Existem diversas tarefas que podem ser realizadas na mineração dos dados, sendo

a classificação uma das principais. Esta trata de obter um modelo durante a fase de treinamento que consiga predizer corretamente qual é a classe (y_i) de uma nova amostra. O algoritmo de Árvore de Decisão uma das estratégias básicas de solução deste problema, com suas vantagens e desvantagens (HAN, 2006). Este se mostrou interessante para comparação por ser um método simples, rápido, não paramétrico (visto que não foi utilizado poda) e sem condições de uso, como classificar apenas entre duas classes por vez, que exigiriam um ajuste da técnica ou utilização de extensões.

O Apêndice B apresenta os resultados expostos da Tab. 3.4 à Tab. 3.4 de forma gráfica (Fig. B.52 à Fig. B.57). Onde, novamente, se verificou a equivalência do CMDS e PCA, quando a métrica de distância é Euclidiana, e do SVD e PCA, ao serem realizados a partir dos dados padronizados (dos quais, a média dos dados é removida pelo Escore-z).

Antes dos comentários quanto as classificações dos escores, observe que em todas as bases estudadas a acurácia da Árvore de Decisão para os dados originais (**X**) e os dados padronizados (**Z**), praticamente, não apresentam diferença. Portanto, temos que a aplicação do Escore-z para as bases Triagem de Crédito, Imagens, Iris, Sonar, Espiral e Vogal, não são diretamente vantajosas, nem prejudiciais. Resta-nos avaliar quando forem combinadas com os métodos analisados.

Como apresentado, existem basicamente dois objetivos ao se utilizar a redução de dimensionalidade como um pré-processamento: explicitar as relações entre os dados de forma a obter um melhor resultado (no caso uma maior acurácia na classificação); e perder o mínimo de informação relevante, ou seja, priorar ao mínimo o resultado do algoritmo de mineração, quando é necessário reduzir os dados (seja por motivos de custo de execução - tanto quanto ao tempo ou memória - ou pelas condições de armazenamento dos dados). É evidente que um método que atinja o primeiro objetivo, de melhora a acurácia do classificador, consequentemente, supre o segundo. Mas que a recíproco não é verdadeira.

Analisemos, então, inicialmente para quando se deseja aumentar a acurácia da classificação quando aplica-se a Árvore de Decisão sobre os dados transformados. Veja que não avaliamos apenas para quando a redução da dimensionalidade foi realizada ($k < p$), mas também, para os casos onde os métodos de redução foram aplicados para $k = p$.

e nenhuma dimensão foi descartada. Isto pode ser uma estratégia interessante, para se destacar as características relevantes, caso o volume de dados não seja um problema. No entanto, observa-se, por exemplo, nos resultados para as bases Sonar e Espiral, que muitas vezes realizar apenas a transformação e não a redução não é uma decisão vantajosa, pois apesar de ao utilizar-se as coordenadas obtidas pelo métodos estudados, possivelmente, se destacou as distinções entre os dados, informações irrelevantes e que estão poluindo o sistema e ocasionando uma piora na mineração não são eliminados.

A base Espiral é um caso muito interessante de ocorrência deste fenômeno. A primeira análise das características destes dados, se tem a impressão que uma redução de dimensão não é vantajosa, quiçá necessária, por ser tratar de um problema com espaço dimensional muito pequeno $p = 2$. Mas ao compararmos os resultados da classificação sem uso dos métodos de redução, com a dos dados transformados mas mantendo ambas as dimensões ($k = p = 2$), e com os reduzidos ($2 = p > k = 1$), observamos que a simples modificação das coordenadas já proporciona uma melhora considerável. Ainda mais considerando que se trata de um problema de classificação com apenas duas classes, e portanto a acurácia dos dados não transformados é a mesma que a de se "lançar uma moeda" (50% a 50%).

Verifica-se que neste caso, a padronização unificada com os métodos, obtém representações bidimensionais sempre melhores que a padronizada e não padronizada. E mesmo, para o caso onde as transformações são feitas sobre \mathbf{X} , apenas três métodos (PCA, SVD e CMDS Euclidian) pioraram os resultados. Indo mais além, confirmamos a afirmação de que nem sempre a simples transformação é melhor que a redução. Observar que apesar da para a maior parte dos métodos ambas ocasionarem em uma melhor acurácia, os melhores valores gerais são os obtidos em pelas diversas configurações do MDS (exceto o Sstress Manhattan, que obteve um péssimo resultado com a representação unidimensional) chegando aproximadamente aos 75% de acerto contra o 70% do MDS Sstress Manhattan para $k = p$.

Outro base onde os métodos atingiram o primeiro objetivo e melhoraram a performance do classificador foi a Sonar. Exatamente a base com o maior número de dimensões ($p = 60$), oposta a Espiral. Devido ao seu p muito grande (comparativamente os demais

conjuntos de dados analisados neste trabalho) este melhora se torna muito satisfatória, por este ser um dos casos típicos da necessidade da aplicação da uma redução de dimensionalidade (por quase todos os motivos levantados).

Note que novamente, a padronização em conjunto com os métodos proporcionou maiores acurárias, exceto para o FastMap. Aliás, este método sempre obteve uma performance consideravelmente pior quando utilizado em conjunto com o Escore-z. Por exemplo, nesta base de dados, a partir dos não padronizados ele apresenta os maiores valores de acurácia, dominando com uma grande vantagem os demais, entretanto ao ser aplicado sobre \mathbf{Z} , torna-se o pior dos métodos, com a maior parte dos valores inferiores a ambas matrizes de dados não transformadas. Isto é devido ao fato da heurística do FastMap de escolher a maior distância esperando que seja a com maior variabilidade ser prejudicada pelo ajuste intencional do escore-z.

Sendo assim, o FastMap vai na contra mão dos demais métodos, os quais ao serem aplicados ao dados padronizados, melhoraram os resultados para o 2/3 de dimensões maiores. Fora o CMDS Manhattan que manteve os mesmos valores aproximadamente. E para as dimensões menores, onde os resultados já eram superiores aos dos dados não transformados, todos os algoritmos, também, apresentaram melhora, exceto o MDS Stress Manhattan, o qual não modificou os resultados, entretanto este que já eram os segundo melhores, se tornaram o melhor para os dados padronizados (visto que se mantiveram e o FastMap piorou), chegando em ambos os casos a 80% de acerto, 10% a mais que os dados não transformados, e isto para uma dimensão muito menor que a original ($60 = p >> k \cong 8$).

Por sua vez os resultados para a base de dados Iris, não atingem para a maioria o primeiro objetivo de superar o 92% de acurácia dos dados não transformados, mas suprem de maneira muito adequada ao segundo, não apresentando acurácia inferior a 88% para nenhum método quando o $k = 2, 3, 4$. E para a dimensão unitária os resultados piores resultados são melhores que 70%, os quais foram obtidos pelo SVD e pelo MDS Stress-1, tanto Euclidian quanto Manhattan. Interessantemente estes são os algoritmos que conseguem melhorar mais o desempenho do classificador nas demais dimensões. Com o MDS Stress-1 Manhattan e o SVD alcançando 95% de acerto para os dados originais, e o

SVD (e consequentemente o PCA) juntamente com o MDS Stress-1 Euclidiano atingem 94% para os dados padronizados.

A última base para a qual os métodos conquistaram qualquer melhoria foi a Vogal, e mesmo assim formam valores mínimos e apenas para algumas das dimensões da metade mais alta e para os métodos SVD e MDS Stress-1 Euclidiano para os dados originais e apenas o MDS Sstress Euclidiano no caso dos padronizados. Sendo que este último é o primeiro a convergir para o valor da classificação a partir dos não transformados em ambos os casos, sendo que isto ocorreu na dimensão $k = 4$ para matriz de dados não padronizada, e em $k = 6$ para o outro. Em relação ao segundo objetivo, todos os demais métodos, exceto o FastMap para os dados padronizados, obtiveram representações com valores de acurácia bem próximas a curva do MDS Sstress Euclidiano e convergindo próximo a metade das dimensões.

Em relação a base Imagens, os resultados só podem ser considerados perante ao segundo objetivo, onde não se visa a melhora, mas uma perda mínima. Pois apenas, para os dados padronizados o PCA-SVD e o MDS Sstress Manhattan se aproximaram da acurácia dos dados não transformados, e isto apenas para $k \geq 14$. Entretanto, desconsiderando o FastMap padronizado, todos métodos obtiveram resultados maiores a 70% a partir da quarta dimensão no caso da matriz de dados original, e 75% da segunda dimensão em diante para o outro caso. Sendo que os melhores resultados unidimensionais não foram bons sendo de 50% e 65% respectivamente.

Por fim temos os resultados para a base Triagem de Crédito os quais só atingiram resultados maiores que 75% (isto é, 10% pior que os não transformados) a partir de $k = 2$ para os dados padronizados. E no caso dos não padronizados, foi pior, onde apenas o SVD, FastMap e CMDS Euclidiano conseguiram ultrapassar a marca de 70% de acerto e mesmo assim na cinco maiores dimensões somente.

Além das observações diretas de cada método perante as bases e a padronização ou não dos dados, podemos retirar desta análise que os métodos MDS apresentam uma melhora mais rápida que os demais, entretanto, principalmente nos casos unidimensionais obtém resultados muito ruins. Sendo destes o com a função Stress-1 o mais estável e robusto, e portanto será considerado o melhor de acordo com esta métrica, seguido de

perto pelo SVD, que por mais que tenha tido resultados medianos, conseguem em certas ocasiões se destacar comparativamente.

4.7 VISUALIZAÇÃO DOS DADOS

Finalmente chegamos ao último critério de análise, a visualização dos dados. O Apêndice C dispõe os *matrix plot* dos dados originais e padronizados para as bases Iris, Espiral e Vogal.²¹ E em seguiam os gráficos de dispersão (*scatter plot*) para todas as bases e combinações métodos, padronização ou não, e distância Euclidiana ou de Manhattan (quando se aplica). Sendo primeiro os unidimensionais e por último os tridimensionais²².

Antes das análise ressaltaremos que a equivalência do PCA e CMDS Euclidiano e do PCA padronizado e SVD foram definitivamente comprovadas. Pois além de apresentarem, nas secções anteriores, os mesmos resultados para as métricas e o classificador, onde suas possíveis diferenças de sinal não geraram resultados distintos, na análise atual verificou-se que os gráficos obtidos quando esperava-se as equivalência eram exatamente os mesmos gráficos, só variando em casos onde um era o espelho do outro.

As únicas bases onde as plotagens em uma única dimensão geraram gráficas adequados foram a Iris não e padronizada, e apenas as Triagem de Crédito e Vogal originais. As demais apenas pode se supor que os dados sejam tenham certo nível de dificuldade de separação, pois apresentaram gráficos onde os pontos ficaram muito misturados e sobrepostos

Nos dois casos da Iris, apenas os métodos PCA, SVD e CMDS Euclidiano conseguiram capturar de forma adequada a configuração dos dados ressaltando a separabilidade linear de uma das classes e a não-linear entre as outras duas, e que apesar disto a maior parte dos elementos destas últimas classes não são concorrentes. Quanto a base Triagem de Crédito foi identificado que os seus dados são polarizados em duas classes distintas mas com um transição conturbada. E finalmente, para o gráfico da Vogal percebe-se que os itens se organizam de uma maneira aos de mesma cor ficarem próximos de forma que a

²¹Os *matrix plot* das bases Triagem de Crédito, Imagens e Sonar, não foram expostas pelo fato de serem de dimensões grandes e isto tornar a sua visualização inviável neste. Entretanto, eles foram gerados e analisados, podendo ser obtidos com o autor.

²²Não foram gerados gráficos tridimensionais para a base Espiral, pior sua dimensão máxima é $p = 2$.

imagem final parece uma *color bar* tradicional em gráficos de terreno.

Em relação a Triagem de Crédito os gráficos do PCA, SVD, ambos CMDS, ambos MDS Stress-1 e FastMap são extremamente similares, aparentando serem praticamente rotações um do outro. Por mais que as plotagens dos dois MDS Sstress sejam mais próximas do que imaginamos de um gráfico de dispersão, ele apresenta a pior visualização para o fim de identificação das classes dos dados (que é foco delas neste trabalho), pois os itens das distintas classes aparecem misturados sem nenhum padrão perceptível. Enquanto, nos demais é possível se notar torno de três eixos sendo dois dominados por amostras da classe azul e o outro tal qual o espaço fora destes eixos é predominantemente dos vermelhos.

No caso, esta dicotomia foi melhor retratada quando os métodos são aplicados sobre os dados padronizados. Onde todas os métodos se aproximaram, sendo que o FastMap se mostrou levemente pior. Enquanto o PCA-SVD ambos CMDS e o MDS Stress-1 Manhattan são minimamente melhores.

Para a base Imagens original o SVD teve uma performance medíocre, e o FastMap praticamente, também. Já o PCA-CMDS Euclidiano conseguiram mostrar que a classe dos azuis é linearmente separável das demais e que os verdes são os que mais se distanciam desta ouras se aproximando dos azuis. Mas as melhores representações foram de ambos MDS Stress-1 e do MDS Sstress Manhattan, onde o grupo das classe não lineares fica mais visível. Sendo que o CMDS Manhattan e o MDS Stress Euclidiano, também isto conseguem, por mais que em menor quantidade.

Novamente a padronização melhorou a visibilidade dos dados, os espalhando mais pelo plano. O Fast Map consegue se tornam mais visível, tal qual o SVD que melhora até ficar equivalente ao PCA, sendo que este e o CMDS Euclidiano distribuíram melhor o grupo dos itens não lineares se aproximando da representação do CMDS Manhattan, o qual praticamente não mudou. Todos os MDS se mantiveram como os melhores, e o MDS Sstress Euclidiano ficou bem próximo dos demais. Através deles, agora é possível se identificar mais uma classe linearmente separável, a dos marrons. E que os itens cianos por mais que se agrupem primordialmente em um ponto, também tem itens espalhados por todo o espaço. Ainda, é possível agora se identificar claramente os demais grupos de dados e até separa-los (por mais que apenas com polinômios de grau alto).

Os dados da base Iris foram primordialmente plotados por todos os métodos, os quais mostraram claramente os dois blocos um separável linearmente o dos azuis e outro não, o dos vermelhos e verdes. Sendo que os elementos que produção esta separabilidade não linear são poucos e podem ser corretamente classificados com polinômios não muito grandes. Sendo que o PCA-CMDS Euclidiano tiverem uma representação levemente melhor, com menos sobreposição.

Diferentemente, das bases anteriores para a Iris a padronização foi prejudicial, mas de maneira igual para todos os métodos. Pois, alguns dos pontos mais distantes do centro do classe foram isolados, isto mesmo para a linearmente separável. Sendo que mais pontos ficam do lado errado da fronteira verde-vermelho, e em locais dificilmente corrigíveis.

As representações bidimensionais da base Sonar originalmente com 60 dimensões não foi satisfatória para nenhum dos métodos. Sendo que o FastMap padronizado foi o que apresentou um resultado ligeiramente mais visível quando a separação das classes, seguido pelo PCA também padronizado. Em relação aos demais métodos não padronizados, principalmente os dois MDS Sstress, percebe-se que os dados se dispões de forma a terem o centro vazio. E em todos parece existir uma sutil tendencia dos vermelhos para um lado e dos azuis para o outro, mas com elementos bem misturados e ocorrendo na extremidade que seria considerada como da outra classe.

Todos os métodos, independentemente da padronização ou não apresentaram exatamente as mesmas plotagens variando apenas levemente quanto a rotações e modificações mínimas de escala. Apesar de terem produzido imagens bonitas, e visíveis isto se deve apenas ao fato das classes estarem destacadas e sabermos que a base se trata de duas espirais concêntricas. Se não soubéssemos disto e as amostras não estivessem coloridas de nada adiantariam. Note que os gráficos obtidos foram exatamente os mesmo da matriz original e padronizada. Desta forma, conjecturamos que a transformação de dados mantendo o número de dimensões ($k = p$) não é de grande valia para a visualização.

Finalmente, comparando as plotagens bidimensionais da base Vogal percebe-se que todos os métodos produziram representações praticamente equivalentes entre si, tanto a partir dos dados originais quanto dos padronizados. Sendo que o PCA, SVD e ambos

CMDS melhoraram levemente quando feitos sobre **Z**, se aproximando mais do obtido em ambos os casos pelos MDS e FastMap. Entretanto esta diferença entre eles é minima e consideramos que eles são aproximadamente tão bom quanto. Apesar de não ser possível separar linearmente qualquer classe, sua concentração e o espalhamento (que não é muito grande) de seus elemento é facilmente observável. Tal qual a interação entre as classes, que se da basicamente com as vizinhas, o que procede ao sabermos que se tratam da pronuncia de vogais fonéticas no inglês britânico. Ou seja a de vogais similares é mais condível, enquanto a de distintas são identificadas sem dificuldades.

Agora, abordaremos os resultados das representações tridimensionais. Estas para a base Iris não foram de grande valia, apresentando os mesmos resultados que as em duas dimensões. Sendo que devido a nem sempre o ajuste do angulo do gráfico este da melhor maneira, a tridimensional pode até mesmo prejudicar levemente a visualização, ao esconder alguns dados encobrindo-os com os demais, devido ao posicionamento da câmera.

Os resultados para a Triagem de Crédito também foram bem próximo do bidualimensional, entretanto, agora os eixos dominados pelas diferentes classes são mais visíveis para todos os métodos que já o tinha apresentado, quando se tratando dos dados originais, e, para estes ambos MDS Sstress se mantiveram péssimos. Em relação aos padronizados, da mesma forma, tem-se praticamente o obtido no caso anterior. Mas com uma melhora no contraste entre as duas classes. Sendo que, agora, é evidente que as representações do MDS Sstress Euclidiano e FastMap não são boas. Mas as dos outros se equiparam de forma positiva.

Por sua vez a uma dimensão a mais para os métodos sobre a base Imagens original foi muito útil. Permitindo-lhes destacar a distribuição dos dados no espaço, de forma que ficassemenos unidos. Tendo cada método pontos onde é mais visível e outros onde é menos. Já para os padronizados, a vantagem foi ainda maior. É possível se analisar a distribuição claramente e conseguimos identificar todos as classes, sendo aparentemente factível a separabilidade linear quase perfeita de no mínimo três delas. Entretanto o FastMap foi o único método que não gerou uma representação comparativamente boa. Sendo que os métodos PCA-SVD, CMDS Euclidiano conseguirem um resultado levemente melhor que o dos demais.

Para a base Sonar as melhorias não foram tão expressivas, mas é possível se identifica um pouco mais facilmente a tendencia de os vermelhos para um lado e os azuis para o outro. Sendo, as representações padronizadas evidentemente superiores que as não. Mais uma vez, com a ressalva de exceto para o FastMap.

Por último, temos que a terceira dimensão não propiciou uma melhora para as representações da base vogal, tornando as visibilidade que eram suficientemente boas, em um pouco confusas, para todos os métodos independentemente da padronização. Talvez isto se deva a relação entre as vogais se darem de uma formas bidimensional, conforme sabemos dos quadros fonéticos da linguística.

Ainda é importante notarmos o aninhamento das dimensões ocorrido no PCA, SVD, CMDS e FastMap. Perceba que sempre a dimensão $(k - 1)$ é a projeção em alguma reta ou plano (em casos não visíveis hiper-plano) contido na representação da dimensão k . E observe que isto não ocorre (obrigatoriamente) para os MDS não métrico.

Perceba ainda o plano onde é feito estas projeções para redução da dimensão. No caso do PCA, SVD e CMDS eles sempre são algum plano definido pelos vetores dos eixos coordenados. Já para o FastMap, isto não é necessariamente válido, no gráfico do Triagem de Crédito, por exemplo, é evidente que o plano da nova projeção é o plano perpendicular ao segmento entre as amostras mais distantes.

5 Considerações Finais

Este trabalho apresentou os principais e tradicionais métodos de transformação de dados para redução de dimensionalidade com o foco nas aplicações de pré-processamento em um processo de descoberta de conhecimento em banco de dados e de visualização. Tendo coberto das importância deste tema para a computação e ciências em geral no contexto atual e das bases teóricas matemáticas, principalmente da álgebra linear, e da estatística multivariada até uma análise comparativa detalhada destes métodos para bases de dados de uso comum na área. Passando por toda explanação da teoria e computação de cada um dos métodos selecionados e suas principais técnicas de resolução.

Desta forma pudemos observar que não existe um método que seja a solução universal para todos os problemas, sendo sempre a escolha dele uma arte baseada em decidir quais as características desejadas perante ao base de dados em questão. Tendo que levar em conta todos o *trade offs*. Sendo estas decisões auxiliadas pela experiência adquirida com aplicações anteriores e estudos comparativos detalhados como este. Lembrando sempre das fundamentações de cada estratégia.

Ainda foi apresentado uma maneira de comparar os métodos não apenas pela análise de suas visualizações e, ou do resultados dos algoritmos de mineração que utilizam as novas representações de dados. Mas sim utilizando as próprias métricas que guiam a busca pela nova configuração. Algo, que do conhecimento do autor nunca tinha sido realizado. Ao se permitir calcular a métrica utilizada por um método para todos os demais pode-se realmente realizar uma análise do relacionamento destes segundo a forma que os regem.

Em acréscimo a estas métricas foram utilizadas duas novas formulações, as quais se mostraram válidas mas com resultados tão similares aos das formulações já existentes, e aqui estendidas para os outros métodos, que não se identificou uma utilidade direta. Mas levantou-se a proposta de verificação de sua aplicação para outras distâncias e a determinação do motivo desta equivalência. Podendo ser isto um conhecimento que revele mais sobre os métodos.

Também como trabalhos futuros sugere-se a extensão desta análise para outras bases de dados, maiores, e para dados de aplicações reais. Tal qual o uso de outros métodos e das extensões mencionadas. Além disto, também, propomos que seja feita a verificação da possibilidade da adaptação da métrica natural ao SVD, a proporção de aproximação acumulada, aos demais métodos, e caso positivo o seu desenvolvimento e análise deles perante esta. E, principalmente, sugere-se que seja feita uma investigação sobre os motivos da decomposição em valor singular ter superado o escalonamento multi-dimensional clássico, em sua métrica nativa. O que foi conjecturado, ser devido a métrica utilizada no caso onde isto ocorreu a Manhattan e não Euclidiana.

A Resultados

Abaixo segue a tabela que relata as métricas de qualidade da representação e o tempo de execução de cada método para cada dimensão possível das seis bases avaliadas nos quatro cenários definidos.

Tabela A.1: Métricas de qualidade da representação e tempo de execução

Cenário: X e DEuclidiana						
PCA						
Base de dados	k	Ac	Strain	Stress-1 (E->E)	Sstress (E->E)	(s)
Triagem de Crédito	15	1	0.29648	1E-015	1E-015	0
	14	1	0.29648	1E-015	1E-015	0
	13	1	0.29648	3.076E-007	6.145E-010	0
	12	1	0.29648	4.6928E-007	1.1273E-009	0
	11	1	0.29648	7.3531E-007	1.8564E-009	0
	10	1	0.29648	1.0092E-006	2.7713E-009	0
	9	1	0.29648	1.3454E-006	3.5596E-009	0
	8	1	0.29648	1.8843E-006	0.000000005	0
	7	1	0.29648	7.1126E-006	1.7167E-008	0
	6	1	0.29648	0.000029752	8.5701E-008	0
	5	1	0.29648	0.000067304	4.4813E-007	0
	4	1	0.29648	0.00026043	1.0433E-006	0
	3	0.99929	0.29648	0.015009	0.00035511	0
	2	0.94873	0.30354	0.1543	0.033546	0
	1	0.69545	0.44548	0.50798	0.13356	0
Imagens	18	1	0.93908	1E-015	1E-015	0
	17	1	0.93908	1E-015	1E-015	0
	16	1	0.93908	1E-015	1E-015	0
	15	1	0.93908	1E-015	1E-015	0
	14	1	0.93908	1E-015	1E-015	0
	13	1	0.93908	6.6863E-008	1.7255E-008	0
	12	1	0.93908	0.000000166	3.7263E-008	0
	11	1	0.93908	0.000001953	3.153E-007	0
	10	0.99999	0.93908	0.000028496	5.7855E-006	0
	9	0.99995	0.93908	0.000081801	0.000019214	0
	8	0.99983	0.93908	0.00066272	0.00014926	0
	7	0.99933	0.93908	0.0022994	0.00065504	0
	6	0.99739	0.93908	0.0042885	0.0011498	0
	5	0.99178	0.93908	0.015977	0.011296	0
	4	0.98424	0.9391	0.020576	0.012487	0
	3	0.90441	0.94043	0.12854	0.038334	0
	2	0.71422	0.94795	0.34631	0.097134	0
	1	0.39547	0.96878	0.651	0.20332	0
Iris	4	1	0.99763	0	0	0
	3	0.99401	0.99763	0.014312	0.0075228	0
	2	0.97707	0.99763	0.044085	0.02375	0
	1	0.92373	0.99764	0.11132	0.076169	0
	60	1	0.99608	1E-015	1E-015	0.04
	59	1	0.99608	3.7417E-006	5.5817E-006	0.04
	58	0.99999	0.99608	7.5876E-006	0.000011659	0.04
	57	0.99998	0.99608	0.000012574	0.000019573	0.04
	56	0.99998	0.99608	0.000017913	0.000027293	0.04
	55	0.99997	0.99608	0.000024326	0.000037771	0.04
	54	0.99995	0.99608	0.000032497	0.000049869	0.04
	53	0.99994	0.99608	0.000043061	0.000065939	0.04
	52	0.99992	0.99608	0.000054997	0.000084643	0.04
	51	0.9999	0.99608	0.000071536	0.00010974	0.04
	50	0.99986	0.99608	0.000092799	0.00014302	0.04
	49	0.99982	0.99608	0.00012322	0.00018747	0.04
	48	0.99975	0.99608	0.0001766	0.00026457	0.04
	47	0.99967	0.99608	0.00022634	0.00034172	0.04
	46	0.99958	0.99608	0.00028692	0.00043573	0.04
	45	0.99943	0.99608	0.00038423	0.00058684	0.04

Cenário: X e DEuclidiana					
SVD					
Base de dados	k	Prop Var Ac	Strain	Stress-1 (E->E)	Sstress (E->E)
Sonar	44	0.99923	0.99608	0.00050918	0.00079044
	43	0.999	0.99608	0.00066193	0.0010403
	42	0.99873	0.99608	0.00083297	0.0013082
	41	0.99836	0.99608	0.0010754	0.0016822
	40	0.99797	0.99608	0.0013219	0.0020684
	39	0.99756	0.99608	0.0015691	0.0024438
	38	0.99708	0.99608	0.0018607	0.0029062
	37	0.99655	0.99608	0.0022036	0.0034542
	36	0.996	0.99608	0.002542	0.0039841
	35	0.99535	0.99608	0.0029352	0.0045967
	34	0.99465	0.99608	0.0033893	0.0053222
	33	0.99393	0.99608	0.0038504	0.0060204
	32	0.99317	0.99608	0.004319	0.0067508
	31	0.99226	0.99608	0.0048795	0.0076628
	30	0.99129	0.99608	0.0054535	0.0085617
	29	0.99011	0.99608	0.0062231	0.0097022
	28	0.98881	0.99608	0.0070461	0.01105
	27	0.9874	0.99608	0.0079453	0.012492
	26	0.98584	0.99608	0.0089115	0.014055
	25	0.98368	0.99608	0.010262	0.016121
	24	0.98136	0.99608	0.01187	0.018738
	23	0.97896	0.99608	0.013695	0.021672
	22	0.97628	0.99608	0.015569	0.024525
	21	0.97335	0.99608	0.017513	0.027385
	20	0.96969	0.99608	0.0201	0.031284
	19	0.96481	0.99608	0.023243	0.036141
	18	0.95953	0.99608	0.026663	0.041479
	17	0.95388	0.99608	0.030519	0.047515
	16	0.94668	0.99608	0.035867	0.055684
	15	0.93902	0.99608	0.040819	0.063507
	14	0.93029	0.99608	0.046904	0.072104
	13	0.92072	0.99609	0.053213	0.081367
	12	0.90902	0.99609	0.061404	0.093446
	11	0.89542	0.99609	0.070747	0.10708
	10	0.87977	0.9961	0.080781	0.1219
	9	0.86055	0.99611	0.093823	0.14071
	8	0.83825	0.99612	0.10876	0.16262
	7	0.81193	0.99614	0.12687	0.18815
	6	0.76985	0.99618	0.157	0.23209
	5	0.72534	0.99623	0.19168	0.27653
	4	0.67369	0.99629	0.23159	0.32651
	3	0.6091	0.99639	0.28664	0.3876
	2	0.52354	0.99657	0.36531	0.46875
	1	0.31971	0.99756	0.55858	0.67691
Espiral	2	1	1E-015	0	1E-015
	1	0.53759	0.6521	0.44902	0.56296
Vogal	10	1	0.98922	1E-015	1E-015
	9	0.99063	0.98922	0.0097372	0.01369
	8	0.96806	0.98925	0.027693	0.039718
	7	0.94084	0.98929	0.046284	0.065969
	6	0.90305	0.98937	0.073589	0.10332
	5	0.86286	0.98945	0.10131	0.14011
	4	0.79339	0.9897	0.15519	0.20851
	3	0.70016	0.99016	0.2303	0.29585
	2	0.56474	0.99113	0.34138	0.42747
	1	0.35295	0.99348	0.53502	0.64009

Triagem de Crédito	15	1	1E-015	0	1E-015	0.07	1
	14	1	1E-015	0	1E-015	0.07	1
	13	1	3.4966E-009	2.5389E-007	6.5941E-010	0.07	0.9999685
	12	1	5.6841E-009	0.000000509	1.2736E-009	0.07	0.9999328
	11	1	9.2184E-009	7.8645E-007	2.096E-009	0.07	0.9998874
	10	1	0.000000012	1.1053E-006	0.000000003	0.07	0.9998409
	9	1	1.4721E-008	1.4904E-006	3.9013E-009	0.07	0.9997915
	8	1	2.5408E-008	0.000002623	7.4353E-009	0.07	0.9997148
	7	1	1.6229E-007	0.000010047	3.1863E-008	0.07	0.9995014
	6	1	7.4233E-007	0.000052098	3.7922E-007	0.07	0.9990478
	5	1	1.1891E-006	0.000076207	4.8041E-007	0.07	0.9985341
	4	0.99999	0.000020704	0.00029789	3.8784E-006	0.07	0.9961109
	3	0.99917	0.001788	0.014852	0.00037969	0.07	0.9735739
	2	0.94836	0.067297	0.14997	0.032997	0.07	0.8353309
	1	0.69359	0.37031	0.50559	0.13405	0.07	0.5136991
Imagens	18	1	0	0	0	0.01	1
	17	1	0	0	0	0.01	1
	16	1	0	0	0	0.01	1
	15	1	0	1E-015	0	0.01	1
	14	1	0	1E-015	0	0.01	1
	13	1	1.3668E-008	6.7408E-008	1.7559E-008	0.01	0.9999513
	12	1	2.3587E-008	1.6624E-007	3.7379E-008	0.01	0.9998935
	11	1	5.5685E-007	0.000002202	0.000000418	0.01	0.9995825
	10	0.99999	9.0235E-006	0.000035122	8.2943E-006	0.01	0.9983316
	9	0.99994	0.000029218	0.000089106	0.000022197	0.01	0.9961341
	8	0.99982	0.00008347	0.00066459	0.00015013	0.01	0.9924481
	7	0.99931	0.00032723	0.0023168	0.00066253	0.01	0.9850333
	6	0.99734	0.0012822	0.0043722	0.0011794	0.01	0.9703557
	5	0.99171	0.0037528	0.015874	0.011202	0.01	0.9456
	4	0.98361	0.0063384	0.021319	0.012883	0.01	0.9158077
	3	0.85705	0.081002	0.17843	0.057273	0.01	0.7973509
	2	0.54493	0.21221	0.45308	0.17182	0.01	0.6127393
	1	0.14994	0.32614	0.7756	0.98863	0.01	0.405293
Iris	4	1	0	0	1E-015	0.01	1
	3	0.99401	0.00044483	0.014314	0.0075229	0.01	0.9830224
	2	0.97658	0.00137	0.044141	0.024296	0.01	0.9540463
	1	0.52819	0.034434	0.32199	0.47849	0.01	0.8047312
Santos et al. [19]	60	1	0	0	1E-015	0.03	1
	59	1	0.000000726	3.7491E-006	5.5864E-006	0.03	0.9997127
	58	0.99999	1.3828E-006	0.000007621	0.000011701	0.03	0.999347
	57	0.99998	2.0218E-006	0.000012611	0.000019605	0.03	0.9989376
	56	0.99998	2.6542E-006	0.000017958	0.000027352	0.03	0.9984955
	55	0.99997	3.5324E-006	0.000024376	0.000037822	0.03	0.9979808
	54	0.99995	4.4344E-006	0.000032559	0.000049939	0.03	0.9974289
	53	0.99994	5.5353E-006	0.000043141	0.000066013	0.03	0.9968153
	52	0.99992	6.8757E-006	0.000055143	0.000084848	0.03	0.9961344
	51	0.99989	8.8094E-006	0.000071839	0.00011036	0.03	0.9953433
	50	0.99986	0.000011242	0.000093198	0.0001438	0.03	0.9944523
	49	0.99982	0.000015247	0.00012473	0.00018994	0.03	0.9933704
	48	0.99974	0.000021788	0.00017833	0.00026735	0.03	0.9920404
	47	0.99967	0.000027761	0.000228	0.00034433	0.03	0.9906421
	46	0.99957	0.000034266	0.00028847	0.00043848	0.03	0.9891312
	45	0.99942	0.000047197	0.00038577	0.00058936	0.03	0.9872106
	44	0.99923	0.000063216	0.00051044	0.00079259	0.03	0.9850244
	43	0.999	0.00008123	0.00066348	0.0010429	0.03	0.9826166
	42	0.99872	0.00010109	0.00083422	0.001311	0.03	0.9800014
	41	0.99836	0.00012906	0.0010801	0.0016881	0.03	0.9769816
	40	0.99796	0.00015524	0.001325	0.0020737	0.03	0.9738504
	39	0.99756	0.00017859	0.0015727	0.0024499	0.03	0.9706827
	38	0.99707	0.00020823	0.0018667	0.002917	0.03	0.9671944

Base de dados	k	Ac	Cenário: X e DEuclidianas			
			CMDS	Stress-1 (E->E)	Sstress (E->E)	(s)
Sonar	15	1	0.29648	4E-014	1E-015	1.55
	14	1	0.29648	4E-014	1E-015	1.55
	13	1	0.29648	3.076E-007	6.145E-010	1.55
	12	1	0.29648	4.6928E-007	1.1273E-009	1.55
	11	1	0.29648	7.3531E-007	1.8564E-009	1.55
	10	1	0.29648	1.0092E-006	2.7713E-009	1.55
	9	1	0.29648	1.3454E-006	3.5596E-009	1.55
	8	1	0.29648	1.8843E-006	0.000000005	1.55
	7	1	0.29648	2.6187E-006	0.000000005	1.55
	6	1	0.29648	3.5937E-006	0.000000005	1.55
Espiral	2	1	0	0	0	1
	1	0.53759	0.6521	0.44902	0.56296	0
Vogal	10	1	1E-015	0	1E-015	0.14
	9	0.98933	0.003461	0.010811	0.015618	0.14
	8	0.96558	0.0084293	0.02877	0.042071	0.14
	7	0.93815	0.012228	0.047405	0.068318	0.14
	6	0.90026	0.017299	0.074754	0.10586	0.14
	5	0.85881	0.021881	0.10379	0.14468	0.14
	4	0.78657	0.031996	0.16043	0.21724	0.14
	3	0.66214	0.051443	0.25576	0.34145	0.14
	2	0.48507	0.077302	0.38774	0.52315	0.14
	1	0.21995	0.11611	0.6342	0.7852	0.14
Cenário: X e DEuclidianas						
CMDS						
Base de dados	k	Ac	Strain	Stress-1 (E->E)	Sstress (E->E)	(s)
Triagem de Crédito	15	1	0.29648	4E-014	1E-015	1.55
	14	1	0.29648	4E-014	1E-015	1.55
	13	1	0.29648	3.076E-007	6.145E-010	1.55
	12	1	0.29648	4.6928E-007	1.1273E-009	1.55
	11	1	0.29648	7.3531E-007	1.8564E-009	1.55
	10	1	0.29648	1.0092E-006	2.7713E-009	1.55
	9	1	0.29648	1.3454E-006	3.5596E-009	1.55
	8	1	0.29648	1.8843E-006	0.000000005	1.55
	7	1	0.29648	2.6187E-006	0.000000005	1.55
	6	1	0.29648	3.5937E-006	0.000000005	1.55

Breast						
7	1	0.29648	7.1126E-006	1.7167E-008	1.55	
6	1	0.29648	0.000029752	8.5701E-008	1.55	
5	1	0.29648	0.000067304	4.4813E-007	1.55	
4	1	0.29648	0.00026043	1.0433E-006	1.55	
3	0.99929	0.29648	0.015009	0.00035511	1.55	
2	0.94873	0.30354	0.1543	0.033546	1.55	
1	0.69545	0.44548	0.50798	0.13356	1.55	
Imagens						
18	1	0.93908	1E-015	1E-015	0.07	
17	1	0.93908	1E-015	1E-015	0.07	
16	1	0.93908	1E-015	1E-015	0.07	
15	1	0.93908	1E-015	1E-015	0.07	
14	1	0.93908	1E-015	1E-015	0.07	
13	1	0.93908	6.6863E-008	1.7255E-008	0.07	
12	1	0.93908	0.000000166	3.7263E-008	0.07	
11	1	0.93908	0.0000001953	3.153E-007	0.07	
10	0.99999	0.93908	0.000028496	5.7855E-006	0.07	
9	0.99995	0.93908	0.000081801	0.000019214	0.07	
8	0.99983	0.93908	0.00066272	0.00014926	0.07	
7	0.99933	0.93908	0.0022994	0.00065504	0.07	
6	0.99739	0.93908	0.0042885	0.0011498	0.07	
5	0.99178	0.93908	0.015977	0.011296	0.07	
4	0.98424	0.9391	0.020576	0.012487	0.07	
3	0.90441	0.94043	0.12854	0.038334	0.07	
2	0.71422	0.94795	0.34631	0.097134	0.07	
1	0.39547	0.96878	0.651	0.20332	0.07	
Iris						
4	1	0.99763	0	1E-015	0.03	
3	0.99401	0.99763	0.014312	0.0075228	0.03	
2	0.97707	0.99763	0.044085	0.02375	0.03	
1	0.92373	0.99764	0.11132	0.076169	0.03	
Sonar						
60	1	0.99608	1E-015	1E-015	0.07	
59	1	0.99608	3.7417E-006	5.5817E-006	0.07	
58	0.99999	0.99608	7.5876E-006	0.000011659	0.07	
57	0.99998	0.99608	0.000012574	0.000019573	0.07	
56	0.99998	0.99608	0.000017913	0.000027293	0.07	
55	0.99997	0.99608	0.000024326	0.000037771	0.07	
54	0.99995	0.99608	0.000032497	0.000049869	0.07	
53	0.99994	0.99608	0.000043061	0.000065939	0.07	
52	0.99992	0.99608	0.000054997	0.000084643	0.07	
51	0.9999	0.99608	0.000071536	0.00010974	0.07	
50	0.99986	0.99608	0.000092799	0.00014302	0.07	
49	0.99982	0.99608	0.00012322	0.00018747	0.07	
48	0.99975	0.99608	0.0001766	0.00026457	0.07	
47	0.99967	0.99608	0.00022634	0.00034172	0.07	
46	0.99958	0.99608	0.00028692	0.00043573	0.07	
45	0.99943	0.99608	0.00038423	0.00058684	0.07	
44	0.99923	0.99608	0.00050918	0.00079044	0.07	
43	0.999	0.99608	0.00066193	0.0010403	0.07	
42	0.99873	0.99608	0.00083297	0.0013082	0.07	
41	0.99836	0.99608	0.0010754	0.0016822	0.07	
40	0.99797	0.99608	0.0013219	0.0020684	0.07	
39	0.99756	0.99608	0.0015691	0.0024438	0.07	
38	0.99708	0.99608	0.0018607	0.0029062	0.07	
37	0.99655	0.99608	0.0022036	0.0034542	0.07	
36	0.996	0.99608	0.002542	0.0039841	0.07	
35	0.99535	0.99608	0.0029352	0.0045967	0.07	
34	0.99465	0.99608	0.0033893	0.0053222	0.07	
33	0.99393	0.99608	0.0038504	0.0060204	0.07	
32	0.99317	0.99608	0.004319	0.0067508	0.07	
31	0.99226	0.99608	0.0048795	0.0076628	0.07	
30	0.99129	0.99608	0.0054535	0.0085617	0.07	

29	0.99011	0.99608	0.0062231	0.0097022	0.07
28	0.98881	0.99608	0.0070461	0.01105	0.07
27	0.9874	0.99608	0.0079453	0.012492	0.07
26	0.98584	0.99608	0.0089115	0.014055	0.07
25	0.98368	0.99608	0.010262	0.016121	0.07
24	0.98136	0.99608	0.01187	0.018738	0.07
23	0.97896	0.99608	0.013695	0.021672	0.07
22	0.97628	0.99608	0.015569	0.024525	0.07
21	0.97335	0.99608	0.017513	0.027385	0.07
20	0.96969	0.99608	0.0201	0.031284	0.07
19	0.96481	0.99608	0.023243	0.036141	0.07
18	0.95953	0.99608	0.026663	0.041479	0.07
17	0.95388	0.99608	0.030519	0.047515	0.07
16	0.94668	0.99608	0.035867	0.055684	0.07
15	0.93902	0.99608	0.040819	0.063507	0.07
14	0.93029	0.99608	0.046904	0.072104	0.07
13	0.92072	0.99609	0.053213	0.081367	0.07
12	0.90902	0.99609	0.061404	0.093446	0.07
11	0.89542	0.99609	0.070747	0.10708	0.07
10	0.87977	0.9961	0.080781	0.1219	0.07
9	0.86055	0.99611	0.093823	0.14071	0.07
8	0.83825	0.99612	0.10876	0.16262	0.07
7	0.81193	0.99614	0.12687	0.18815	0.07
6	0.76985	0.99618	0.157	0.23209	0.07
5	0.72534	0.99623	0.19168	0.27653	0.07
4	0.67369	0.99629	0.23159	0.32651	0.07
3	0.6091	0.99639	0.28664	0.3876	0.07
2	0.52354	0.99657	0.36531	0.46875	0.07
1	0.31971	0.99756	0.55858	0.67691	0.07
Espiral	2	1	1E-015	0	1E-015
	1	0.53759	0.6521	0.44902	0.56296
Vogal	10	1	0.98922	6.4198e-16	1.2211e-15
	9	0.99063	0.98922	0.0097372	0.01369
	8	0.96806	0.98925	0.027693	0.039718
	7	0.94084	0.98929	0.046284	0.065969
	6	0.90305	0.98937	0.073589	0.10332
	5	0.86286	0.98945	0.10131	0.14011
	4	0.79339	0.9897	0.15519	0.20851
	3	0.70016	0.99016	0.2303	0.29585
	2	0.56474	0.99113	0.34138	0.42747
	1	0.35295	0.99348	0.53502	0.64009
Cenário: X e DEuclidiana					
MDS Stress-1					
Base de dados	k	Ac	Strain	Stress-1 (E->E)	Sstress (E->E)
Triagem de Crédito	15	1	0.2965	0.00018863	0.00039061
	14	1	0.2965	0.00018791	0.00039576
	13	1	0.2965	0.00016548	0.00034461
	12	0.99998	0.2965	0.00019095	0.0004026
	11	1	0.2965	0.00018229	0.00038775
	10	1	0.2965	0.0001683	0.00035467
	9	1	0.2965	0.00017833	0.00037175
	8	1	0.2965	0.00016098	0.00034408
	7	1	0.2965	0.00017403	0.00037038
	6	0.99999	0.2965	0.00015789	0.00033612
	5	1	0.2965	0.00019995	0.0003552
	4	1	0.29649	0.0016018	0.00029563
	3	0.99997	0.29665	0.0056757	0.0010624
	2	0.99759	0.41723	0.047465	0.03093
	1	0.97058	1.4246	0.17196	0.14597
	18	1	0.93908	0.000077902	0.00012511
					21.22

Imagenes	17	0.99999	0.93908	0.000075532	0.00012221	20.52
	16	0.99999	0.93908	0.000079979	0.00013443	27.37
	15	1	0.93908	0.000074234	0.00012151	42.58
	14	1	0.93908	0.000072271	0.00011553	28.58
	13	1	0.93908	0.000065095	0.00010206	17.12
	12	1	0.93908	0.000083398	0.00013199	20.01
	11	1	0.93908	0.000086712	0.00012338	17.88
	10	1	0.93908	0.000078101	0.00011862	33
	9	0.99999	0.93908	0.00018288	0.0002122	32.91
	8	0.99999	0.93908	0.00016262	0.00014835	16.22
	7	1	0.93908	0.00068927	0.0006608	14.32
	6	0.9999	0.93909	0.0021538	0.0020385	14.57
	5	0.99999	0.93912	0.0039218	0.003916	14.06
	4	0.99984	0.94038	0.013229	0.020171	19.42
	3	0.99894	0.9532	0.032811	0.066585	14.49
	2	0.98737	0.99133	0.11254	0.12969	8.53
	1	0.8693	1.0771	0.36152	0.22177	3.86
Iris	4	1	0.99763	0.0021842	0.0039836	3.91
	3	0.9999	0.99763	0.0096615	0.0067011	1.84
	2	0.99871	0.99764	0.035867	0.026769	1.45
	1	0.99	0.99766	0.099992	0.08922	0.48
Sonar	60	0.99997	0.99608	0.00023911	0.000434	24.74
	59	0.99999	0.99608	0.00021703	0.00040167	35.07
	58	1	0.99608	0.00025494	0.0004683	47.33
	57	0.99997	0.99608	0.00022534	0.00041121	55.38
	56	0.99999	0.99608	0.00023006	0.00042232	60.89
	55	1	0.99608	0.00023439	0.00042593	33.93
	54	0.99999	0.99608	0.00022785	0.00041017	55.79
	53	0.99997	0.99608	0.00022675	0.00041103	57.99
	52	0.99998	0.99608	0.00028314	0.00050908	63.75
	51	1	0.99608	0.00025859	0.00046674	54.85
	50	1	0.99608	0.00031469	0.00056657	47.11
	49	0.99998	0.99608	0.00035001	0.00064586	71.03
	48	0.99999	0.99608	0.00031275	0.00055748	46.52
	47	1	0.99608	0.00037785	0.00067964	55.42
	46	1	0.99608	0.0002925	0.00052597	56.38
	45	1	0.99608	0.00041495	0.00074254	47.02
	44	1	0.99608	0.00041127	0.00074363	62
	43	1	0.99608	0.00046268	0.00082214	46.49
	42	0.99998	0.99608	0.00050827	0.00091899	45.22
	41	1	0.99608	0.00049046	0.00086712	54.48
	40	1	0.99608	0.00059628	0.0010647	39.98
	39	1	0.99608	0.00063982	0.0011396	60.2
	38	0.99998	0.99608	0.00070173	0.001245	38.73
	37	0.99999	0.99608	0.00074991	0.0013175	49.86
	36	1.0001	0.99608	0.00092819	0.0016273	43.14
	35	0.99996	0.99608	0.001002	0.0017582	42.4
	34	1	0.99608	0.0011033	0.0019237	46.19
	33	0.99999	0.99608	0.0011801	0.0020371	40.41
	32	0.99997	0.99608	0.0012828	0.0022076	44.1
	31	0.99997	0.99608	0.0014215	0.0024472	43
	30	0.99998	0.99608	0.0016111	0.0028095	37.58
	29	1	0.99608	0.0018013	0.003082	48.39
	28	0.99998	0.99608	0.0020423	0.0034743	24.98
	27	1	0.99608	0.0022906	0.0038972	49.34
	26	1	0.99608	0.002563	0.0043467	33.49
	25	1	0.99608	0.0030228	0.0051346	40.62
	24	1	0.99608	0.003462	0.0058908	41.13
	23	0.99997	0.99608	0.0039049	0.0066151	33.05
	22	0.99998	0.99608	0.0044896	0.0075449	45.78

21	0.99991	0.99608	0.0051608	0.0086061	22.12	
20	1	0.99608	0.0059835	0.009965	43.91	
19	0.99998	0.99608	0.0070566	0.011764	29.66	
18	0.99984	0.99608	0.008194	0.013659	30.81	
17	0.99992	0.99608	0.0094035	0.015662	36.35	
16	0.99982	0.99609	0.011159	0.018788	25.29	
15	0.99985	0.99609	0.012919	0.021517	32.09	
14	0.99979	0.99609	0.015273	0.025102	37.78	
13	0.99969	0.9961	0.017757	0.029118	15.87	
12	0.99959	0.99611	0.020832	0.034576	10.23	
11	0.99937	0.99613	0.024572	0.040392	22.44	
10	0.99912	0.99615	0.029236	0.047969	16.35	
9	0.99868	0.99618	0.035452	0.057557	31.27	
8	0.99811	0.99622	0.042948	0.069288	13.6	
7	0.99719	0.99629	0.052447	0.084389	9.94	
6	0.99514	0.99645	0.069196	0.11128	7.46	
5	0.99198	0.99668	0.089029	0.14142	6.97	
4	0.98612	0.99705	0.11805	0.18196	3.57	
3	0.97237	0.99782	0.16642	0.24931	6.27	
2	0.93841	0.99937	0.24852	0.36427	8.78	
1	0.73204	1.0121	0.51765	0.80911	4.6	
Espiral	2	1	2.7509E-006	9.7338E-007	1.5739E-006	
	1	0.83235	0.99178	0.40945	0.56577	
Vogal	10	1	0.98922	9.2613E-007	1.4075E-006	
	9	0.99996	0.98923	0.0062732	0.0091944	
	8	0.99974	0.98926	0.016517	0.025122	
	7	0.99932	0.98933	0.02635	0.039636	
	6	0.99831	0.98949	0.041047	0.061104	
	5	0.99664	0.98972	0.057566	0.084467	
	4	0.99164	0.99044	0.091893	0.13263	
	3	0.97874	0.99206	0.14534	0.20664	
	2	0.93922	0.99841	0.24617	0.37717	
	1	0.72428	1.0263	0.52509	0.78555	
Cenário: X e DEuclidiana						
MDS Sstress						
Base de dados	k	Ac	Strain	Stress-1 (E->E)	Sstress (E->E)	(s)
Triagem de Crédito	15	1.0029	0.29668	0.02682	0.0012477	1018.7
	14	1.0025	0.29667	0.023423	0.0011923	905.77
	13	1.0054	1.0608	1.055	0.9923	0.5
	12	1.002	0.29665	0.020628	0.0011338	946.55
	11	1.0015	0.29667	0.016736	0.00118	977.02
	10	1.0016	0.29657	0.018005	0.00083621	968.24
	9	1.0011	0.29662	0.01478	0.00099835	975.17
	8	1.0013	0.29663	0.016223	0.0010347	946.38
	7	1.0113	1.1135	1.0737	0.99531	0.34
	6	1.0006	0.29659	0.010916	0.00091686	984.65
	5	1.0004	0.29665	0.011998	0.0010954	958.41
	4	1.0001	0.29668	0.013163	0.0011789	941.41
	3	0.99859	1.2429	1.0775	0.99707	0.23
	2	0.99695	1.3497	1.1135	1.0051	0.19
	1	0.84789	0.76123	0.26075	0.083301	328.52
Imagens	18	1.0005	0.93908	0.0039657	0.00066833	86
	17	1.0001	0.93908	0.0030933	0.00058524	83.63
	16	1.0003	0.93908	0.0033952	0.00060608	86.15
	15	1.0001	0.93908	0.0053177	0.001089	65.55
	14	1.0003	0.93908	0.0038128	0.00068055	77.2
	13	1.0003	0.93908	0.0039857	0.00072554	79.26
	12	1	0.93908	0.0028532	0.00059865	74.6
	11	1.0002	0.93908	0.0035564	0.00068823	63.19
	10	0.99998	0.93908	0.0031926	0.00064551	75.93

Iris						
9	0.99992	0.93908	0.0036874	0.00080168	77.67	
8	0.99982	0.93908	0.0031598	0.00069298	54.92	
7	0.99966	0.93908	0.0029248	0.0006929	73.08	
6	0.9994	0.93908	0.0056591	0.0015494	77.07	
5	0.99789	0.9391	0.0086089	0.0025448	65.26	
4	0.99673	0.93914	0.010857	0.0045161	47.83	
3	0.96395	0.94114	0.10613	0.028375	17.32	
2	0.85526	0.95373	0.22878	0.080097	47.67	
1	0.68293	0.991	0.38131	0.15409	29.08	
Sonar						
4	0.99999	0.99763	0.00044822	0.00025279	20.29	
3	0.99836	0.99763	0.011174	0.0052845	14.38	
2	0.99135	0.99763	0.038176	0.017767	11.16	
1	0.95768	0.99765	0.10688	0.062302	5.12	
60	1	0.99608	0.0004536	0.00069813	55.49	
59	0.99997	0.99608	0.00041758	0.00064696	60.81	
58	1	0.99608	0.00045739	0.00070253	59.97	
57	1	0.99608	0.00046981	0.00072503	58.74	
56	0.99996	0.99608	0.00054618	0.0008472	45.33	
55	0.99995	0.99608	0.00051132	0.00079764	57.27	
54	0.99999	0.99608	0.00049006	0.00075942	57.32	
53	0.99999	0.99608	0.0004836	0.00074798	62.09	
52	0.99996	0.99608	0.00049687	0.000767	51.17	
51	0.99996	0.99608	0.00061937	0.00095582	47.22	
50	0.99997	0.99608	0.00070857	0.0010947	56.11	
49	0.99993	0.99608	0.00066909	0.0010374	58.71	
48	1	0.99608	0.00063257	0.00098172	54.45	
47	0.99995	0.99608	0.00062048	0.00096111	38.46	
46	0.99998	0.99608	0.0006771	0.0010492	52.98	
45	0.99995	0.99608	0.00063736	0.00098873	54.21	
44	0.99998	0.99608	0.00075524	0.0011618	36.69	
43	0.99992	0.99608	0.00075146	0.0011654	52.94	
42	0.99996	0.99608	0.00078682	0.0012002	54.71	
41	0.99995	0.99608	0.00084453	0.0012782	44.92	
40	0.99991	0.99608	0.00092776	0.0013968	43.58	
39	0.99989	0.99608	0.00093857	0.0014331	51.83	
38	0.99988	0.99608	0.00098787	0.0014867	47.02	
37	0.99994	0.99608	0.0010794	0.0016351	39.47	
36	0.99989	0.99608	0.0011649	0.0017656	52.86	
35	0.99982	0.99608	0.0012051	0.0018106	39.31	
34	0.99986	0.99608	0.0013077	0.0019461	43.92	
33	0.99977	0.99608	0.0013948	0.0020711	52.55	
32	0.99976	0.99608	0.0015456	0.002325	33.43	
31	0.99975	0.99608	0.001633	0.0023747	49.6	
30	0.99967	0.99608	0.0019061	0.0027791	43.9	
29	0.99962	0.99608	0.0020892	0.003023	38.5	
28	0.99958	0.99608	0.0024094	0.0035101	52.49	
27	0.99947	0.99608	0.0026254	0.003809	31.96	
26	0.99946	0.99608	0.0030674	0.0045077	50.85	
25	0.99938	0.99608	0.0033983	0.0049573	34.37	
24	0.99923	0.99608	0.0038307	0.0055367	51.75	
23	0.99915	0.99608	0.0043204	0.0062959	46.33	
22	0.99895	0.99608	0.0049549	0.007086	38.92	
21	0.9988	0.99608	0.005692	0.0081269	51.62	
20	0.99851	0.99608	0.0065317	0.0093004	42.04	
19	0.99824	0.99608	0.0076653	0.010937	42.49	
18	0.99787	0.99608	0.0088666	0.012762	56.18	
17	0.99745	0.99608	0.010187	0.014753	36.23	
16	0.99694	0.99608	0.012067	0.017537	59.84	
15	0.99618	0.99609	0.013923	0.020264	38.35	
14	0.99532	0.99609	0.016405	0.023637	57.36	

13	0.9943	0.9961	0.019122	0.027322	44.94
12	0.99291	0.99611	0.022635	0.03227	41.38
11	0.99121	0.99612	0.02623	0.038106	31.93
10	0.98885	0.99613	0.031523	0.044933	57.28
9	0.98577	0.99616	0.038059	0.053834	34.97
8	0.98175	0.99619	0.046053	0.064657	40.44
7	0.97589	0.99625	0.056357	0.07843	21.06
6	0.96672	0.99636	0.072164	0.10182	31.3
5	0.95309	0.99654	0.095201	0.13123	54.94
4	0.93223	0.9968	0.12524	0.16684	18.35
3	0.89531	0.99726	0.1766	0.22134	28.92
2	0.82369	0.99809	0.26675	0.30871	28.47
1	0.61984	1.0012	0.47119	0.54757	26.45
Espiral	2	1	1.6293E-006	7.0632E-007	1.0898E-006
	1	0.75047	0.78841	0.39294	0.47864
Vogal	10	1	0.98922	0.000001938	2.8147E-006
	9	0.999	0.98922	0.0065208	0.0087986
	8	0.99623	0.98926	0.017203	0.023902
	7	0.99218	0.98932	0.027466	0.037678
	6	0.98523	0.98945	0.042826	0.057841
	5	0.97519	0.98963	0.060446	0.078725
	4	0.95436	0.99016	0.097014	0.12194
	3	0.91374	0.99128	0.15332	0.18755
	2	0.83488	0.99375	0.246	0.2937
	1	0.62778	1.0027	0.46149	0.55521

Cenário: X e DEuclidiana

FastMap

Base de dados	k	Ac	Strain	Stress-1 (E->E)	Sstress (E->E)	(s)
Triagem de Crédito	15	1	4.1277	1.5173E-006	1.6047E-006	0.04
	14	1	4.1277	1.5173E-006	1.6047E-006	0
	13	1	4.1277	1.5586E-006	1.6047E-006	0.01
	12	1	4.1277	1.6321E-006	1.6046E-006	0
	11	1	4.1277	1.7673E-006	1.6046E-006	0
	10	1	4.1277	1.9316E-006	1.6045E-006	0
	9	1	4.1277	2.3026E-006	1.6045E-006	0
	8	1	4.1277	3.0287E-006	1.6043E-006	0
	7	1	4.1277	7.6229E-006	1.6042E-006	0
	6	1	4.1277	0.000031752	1.6037E-006	0
	5	1	4.1277	0.00016564	1.7458E-006	0
	4	1	4.1277	0.00025917	1.9036E-006	0
	3	0.99752	4.1245	0.013882	0.00090189	0
	2	0.8828	4.07	0.16844	0.051691	0
	1	0.67276	2.3421	0.42393	0.13957	0
Imagens	18	1	0.94257	2.0758E-006	5.1056E-006	0.02
	17	1	0.94257	2.0758E-006	5.1056E-006	0.02
	16	1	0.94257	2.0758E-006	5.1056E-006	0
	15	1	0.94257	2.0758E-006	5.1056E-006	0.01
	14	1	0.94257	2.0758E-006	5.1056E-006	0.01
	13	1	0.94257	0.000002078	0.000005106	0
	12	1	0.94257	2.0844E-006	5.1067E-006	0
	11	1	0.94256	3.7038E-006	0.000005149	0.01
	10	0.99995	0.94249	0.000077162	0.000018823	0
	9	0.99993	0.94237	0.00011411	0.000027891	0
	8	0.9995	0.94151	0.00086113	0.00026391	0
	7	0.99768	0.93356	0.003057	0.00095123	0.01
	6	0.99611	0.93088	0.0052582	0.0015961	0
	5	0.98694	0.90795	0.014658	0.0048214	0
	4	0.93432	0.78431	0.065497	0.023771	0
	3	0.88264	0.68263	0.13704	0.044456	0
	2	0.67279	0.71948	0.36213	0.11221	0

	1	0.39363	0.94804	0.69259	0.20308	<i>0</i>
Iris	4	1	0.7438	1.0794E-006	1.4181E-006	<i>0</i>
	3	0.99284	0.74872	0.017697	0.0093979	<i>0</i>
	2	0.97261	0.75822	0.046293	0.029106	<i>0</i>
	1	0.90714	0.79685	0.11171	0.091752	<i>0</i>
Sonar	60	1	1.603	1.5665E-006	2.8114E-006	<i>0.05</i>
	59	0.99999	1.603	9.6745E-006	0.000015768	<i>0</i>
	58	0.99997	1.603	0.000026443	0.000041434	<i>0</i>
	57	0.99995	1.6029	0.000039455	0.000061319	<i>0</i>
	56	0.99992	1.6029	0.00006169	0.000093878	<i>0</i>
	55	0.99988	1.6028	0.00008657	0.00013223	<i>0</i>
	54	0.99984	1.6027	0.00011111	0.00017225	<i>0</i>
	53	0.99979	1.6026	0.00014157	0.00021837	<i>0</i>
	52	0.99972	1.6024	0.00018909	0.00029284	<i>0.05</i>
	51	0.99964	1.6022	0.0002368	0.0003665	<i>0.01</i>
	50	0.99954	1.6019	0.0003018	0.00046892	<i>0</i>
	49	0.9994	1.6016	0.00039742	0.00061048	<i>0</i>
	48	0.99914	1.601	0.0005799	0.00088785	<i>0</i>
	47	0.99883	1.6005	0.00078945	0.0012211	<i>0</i>
	46	0.99836	1.5996	0.0011094	0.0016995	<i>0</i>
	45	0.99798	1.5981	0.0013592	0.0020829	<i>0.01</i>
	44	0.99755	1.5965	0.0016367	0.002507	<i>0</i>
	43	0.99707	1.5955	0.0019245	0.0029796	<i>0</i>
	42	0.99641	1.5937	0.0023276	0.0036112	<i>0</i>
	41	0.99585	1.5915	0.0026694	0.0041539	<i>0</i>
	40	0.99479	1.5882	0.0033303	0.0052078	<i>0</i>
	39	0.99358	1.5846	0.0040887	0.0064507	<i>0.01</i>
	38	0.99295	1.5826	0.0044305	0.0069977	<i>0</i>
	37	0.99198	1.578	0.0049904	0.0079423	<i>0</i>
	36	0.99088	1.5729	0.0056225	0.0090281	<i>0</i>
	35	0.98971	1.5697	0.0063654	0.010122	<i>0</i>
	34	0.98818	1.563	0.0072745	0.011652	<i>0</i>
	33	0.98661	1.5568	0.0082541	0.013148	<i>0</i>
	32	0.98471	1.5488	0.0095773	0.015175	<i>0</i>
	31	0.98362	1.5415	0.01016	0.016149	<i>0</i>
	30	0.98179	1.5369	0.011277	0.01794	<i>0</i>
	29	0.9792	1.5231	0.01317	0.020695	<i>0</i>
	28	0.97614	1.5065	0.015068	0.023678	<i>0.01</i>
	27	0.97286	1.4915	0.017038	0.026854	<i>0</i>
	26	0.96864	1.4791	0.019517	0.030942	<i>0</i>
	25	0.96593	1.4623	0.021154	0.033604	<i>0</i>
	24	0.96186	1.4513	0.023603	0.037476	<i>0</i>
	23	0.95622	1.4331	0.026789	0.043344	<i>0</i>
	22	0.95184	1.418	0.029798	0.047902	<i>0</i>
	21	0.94646	1.3993	0.033237	0.053214	<i>0</i>
	20	0.93948	1.3722	0.038046	0.060191	<i>0</i>
	19	0.93105	1.3455	0.043774	0.069081	<i>0.01</i>
	18	0.92107	1.3112	0.049258	0.078542	<i>0</i>
	17	0.91247	1.2894	0.055299	0.087423	<i>0</i>
	16	0.90428	1.2426	0.060642	0.095843	<i>0.02</i>
	15	0.88521	1.2027	0.071508	0.11365	<i>0.02</i>
	14	0.87426	1.1348	0.079339	0.12446	<i>0.01</i>
	13	0.8632	1.0897	0.085744	0.13524	<i>0.02</i>
	12	0.8489	1.0427	0.095369	0.14975	<i>0</i>
	11	0.83349	0.9877	0.10577	0.16472	<i>0</i>
	10	0.81154	0.93533	0.12142	0.1866	<i>0</i>
	9	0.7884	0.85055	0.13621	0.2082	<i>0</i>
	8	0.76239	0.768	0.15394	0.23319	<i>0</i>
	7	0.71443	0.62321	0.18537	0.27823	<i>0</i>
	6	0.67785	0.53462	0.21333	0.31418	<i>0</i>

	5	0.63431	0.42837	0.24932	0.35665	0
	4	0.58814	0.35014	0.28373	0.39875	0
	3	0.5236	0.32678	0.34229	0.45991	0
	2	0.43216	0.38289	0.42553	0.55254	0
	1	0.247	0.46924	0.62179	0.74525	0
Espiral	2	1	7.378	2.3572E-006	4.3144E-006	0
	1	0.51548	4.3637	0.46741	0.59299	0
Vogal	10	1	2.1981	1.1856E-006	0.000002007	0.04
	9	0.98348	2.1667	0.016109	0.024294	0
	8	0.95521	2.0729	0.037927	0.054788	0
	7	0.91848	1.9636	0.06425	0.092374	0
	6	0.8721	1.7805	0.094492	0.13599	0
	5	0.83088	1.676	0.12216	0.17199	0
	4	0.7515	1.4616	0.17929	0.25299	0
	3	0.63999	1.1725	0.25985	0.35551	0.01
	2	0.4881	0.65241	0.38827	0.51519	0
	1	0.31317	0.23037	0.56686	0.6828	0

Cenário: X e DManhattan								
PCA								
Base de dados	k	Prop Var Ac	Strain	Stress-1 (M->E)	Stress-1 (M->M)	Sstress (M->E)	Sstress (M->M)	Tempo (s)
Triagem de Crédito	15	1	0.29648	0.14242	0.046804	0.11484	0.059317	0
	14	1	0.29648	0.14242	0.046804	0.11484	0.059317	0
	13	1	0.29648	0.14242	0.046793	0.11484	0.059304	0
	12	1	0.29648	0.14242	0.046777	0.11484	0.059297	0
	11	1	0.29648	0.14242	0.046755	0.11484	0.059287	0
	10	1	0.29648	0.14242	0.046733	0.11484	0.059274	0
	9	1	0.29648	0.14242	0.046709	0.11484	0.059263	0
	8	1	0.29648	0.14242	0.046682	0.11484	0.059239	0
	7	1	0.29648	0.14242	0.046632	0.11484	0.059204	0
	6	1	0.29648	0.14242	0.046459	0.11484	0.059116	0
	5	1	0.29648	0.14242	0.046284	0.11484	0.059	0
	4	1	0.29648	0.14242	0.045789	0.11484	0.058635	0
	3	0.99929	0.29648	0.14367	0.04627	0.11486	0.056089	0
	2	0.94873	0.30354	0.23763	0.20628	0.12899	0.095975	0
	1	0.69545	0.44548	0.56078	0.56078	0.20043	0.20043	0
Imagens	18	1	0.93908	0.57948	0.32887	0.69137	0.46062	0
	17	1	0.93908	0.57948	0.32887	0.69137	0.46062	0
	16	1	0.93908	0.57948	0.32887	0.69137	0.46062	0
	15	1	0.93908	0.57948	0.32887	0.69137	0.46062	0
	14	1	0.93908	0.57948	0.32887	0.69137	0.46062	0
	13	1	0.93908	0.57948	0.3289	0.69137	0.46063	0
	12	1	0.93908	0.57948	0.32893	0.69137	0.46065	0
	11	1	0.93908	0.57948	0.32911	0.69137	0.46077	0
	10	0.99999	0.93908	0.57949	0.32963	0.69137	0.46113	0
	9	0.99995	0.93908	0.5795	0.33126	0.69137	0.46215	0
	8	0.99983	0.93908	0.57953	0.33289	0.69138	0.46351	0
	7	0.99933	0.93908	0.57968	0.33671	0.6914	0.46612	0
	6	0.99739	0.93908	0.58019	0.34914	0.69149	0.47349	0
	5	0.99178	0.93908	0.58147	0.35969	0.69209	0.48751	0
	4	0.98424	0.9391	0.58356	0.38635	0.69251	0.50249	0
	3	0.90441	0.94043	0.60751	0.47035	0.6957	0.54191	0
	2	0.71422	0.94795	0.67469	0.61966	0.70588	0.61609	0
	1	0.39547	0.96878	0.85041	0.85041	0.74326	0.74326	0
Iris	4	1	0.99763	0.40426	0.26144	0.6426	0.47383	0
	3	0.99401	0.99763	0.40618	0.28635	0.64361	0.50633	0
	2	0.97707	0.99763	0.41257	0.33561	0.64659	0.56278	0
	1	0.92373	0.99764	0.43568	0.43568	0.6578	0.6578	0
iris	60	1	0.99608	0.81119	0.34504	0.96515	0.60605	0.04
	59	1	0.99608	0.81119	0.34528	0.96515	0.60632	0.04
	58	0.99999	0.99608	0.81119	0.34558	0.96515	0.60667	0.04
	57	0.99998	0.99608	0.81119	0.34592	0.96515	0.60708	0.04
	56	0.99998	0.99608	0.81119	0.34628	0.96515	0.6075	0.04
	55	0.99997	0.99608	0.81119	0.34672	0.96515	0.60802	0.04
	54	0.99995	0.99608	0.81119	0.34718	0.96515	0.60856	0.04
	53	0.99994	0.99608	0.81119	0.34768	0.96515	0.60915	0.04
	52	0.99992	0.99608	0.81119	0.34823	0.96515	0.60979	0.04
	51	0.9999	0.99608	0.8112	0.34888	0.96515	0.61055	0.04
	50	0.99986	0.99608	0.8112	0.34958	0.96515	0.61136	0.04
	49	0.99982	0.99608	0.8112	0.35045	0.96515	0.61234	0.04
	48	0.99975	0.99608	0.81121	0.35154	0.96515	0.61358	0.04
	47	0.99967	0.99608	0.81122	0.35268	0.96516	0.6149	0.04
	46	0.99958	0.99608	0.81123	0.35396	0.96516	0.61639	0.04
	45	0.99943	0.99608	0.81124	0.35551	0.96516	0.61817	0.04
	44	0.99923	0.99608	0.81126	0.35736	0.96517	0.62035	0.04
	43	0.999	0.99608	0.81128	0.35932	0.96518	0.62263	0.04
	42	0.99873	0.99608	0.81131	0.36151	0.96518	0.62512	0.04

	41	0.99836	0.99608	0.81134	0.36396	0.96519	0.62791	0.04
	40	0.99797	0.99608	0.81138	0.3666	0.96521	0.63095	0.04
	39	0.99756	0.99608	0.81142	0.36927	0.96522	0.63398	0.04
	38	0.99708	0.99608	0.81146	0.37221	0.96523	0.63732	0.04
	37	0.99655	0.99608	0.81151	0.37516	0.96525	0.64069	0.04
	36	0.996	0.99608	0.81157	0.37823	0.96526	0.64414	0.04
	35	0.99535	0.99608	0.81163	0.38159	0.96528	0.64785	0.04
	34	0.99465	0.99608	0.81169	0.38504	0.9653	0.65166	0.04
	33	0.99393	0.99608	0.81176	0.38846	0.96532	0.65534	0.04
	32	0.99317	0.99608	0.81184	0.39206	0.96534	0.65925	0.04
Sonar	31	0.99226	0.99608	0.81192	0.39606	0.96537	0.66367	0.04
	30	0.99129	0.99608	0.81201	0.40018	0.96539	0.66811	0.04
	29	0.99011	0.99608	0.81213	0.40471	0.96543	0.67282	0.04
	28	0.98881	0.99608	0.81225	0.40955	0.96546	0.67807	0.04
	27	0.9874	0.99608	0.81239	0.41429	0.9655	0.68306	0.04
	26	0.98584	0.99608	0.81254	0.41964	0.96555	0.68871	0.04
	25	0.98368	0.99608	0.81274	0.42588	0.96561	0.69507	0.04
	24	0.98136	0.99608	0.81297	0.43215	0.96568	0.70157	0.04
	23	0.97896	0.99608	0.8132	0.43865	0.96575	0.7082	0.04
	22	0.97628	0.99608	0.81347	0.44552	0.96582	0.7151	0.04
	21	0.97335	0.99608	0.81375	0.45277	0.9659	0.7222	0.04
	20	0.96969	0.99608	0.81411	0.46107	0.96601	0.73027	0.04
	19	0.96481	0.99608	0.81458	0.47081	0.96615	0.73955	0.04
	18	0.95953	0.99608	0.8151	0.48063	0.96629	0.74851	0.04
	17	0.95388	0.99608	0.81565	0.49074	0.96645	0.75775	0.04
	16	0.94668	0.99608	0.81637	0.50212	0.96665	0.76782	0.04
	15	0.93902	0.99608	0.81713	0.51417	0.96687	0.77832	0.04
	14	0.93029	0.99608	0.818	0.52702	0.96709	0.78885	0.04
	13	0.92072	0.99609	0.81897	0.54088	0.96736	0.80011	0.04
	12	0.90902	0.99609	0.82016	0.55609	0.96769	0.81222	0.04
	11	0.89542	0.99609	0.82156	0.57259	0.96807	0.82495	0.04
	10	0.87977	0.9961	0.8232	0.59086	0.96854	0.83876	0.04
	9	0.86055	0.99611	0.82523	0.61056	0.96908	0.85263	0.04
	8	0.83825	0.99612	0.82766	0.63225	0.96974	0.86754	0.04
	7	0.81193	0.99614	0.83058	0.65592	0.97052	0.88277	0.04
	6	0.76985	0.99618	0.8354	0.68609	0.97179	0.90086	0.04
	5	0.72534	0.99623	0.84083	0.71759	0.97312	0.91789	0.04
	4	0.67369	0.99629	0.84736	0.7518	0.97466	0.93434	0.04
	3	0.6091	0.99639	0.85649	0.79019	0.97665	0.95065	0.04
	2	0.52354	0.99657	0.86953	0.83377	0.97922	0.96592	0.04
	1	0.31971	0.99756	0.90584	0.90584	0.98667	0.98667	0.04
Espiral	2	1	1E-015	0.23402	0.17744	0.423	0.34135	0
	1	0.53759	0.6521	0.54059	0.54059	0.70011	0.70011	0
Vogal	10	1	0.98922	0.61266	0.15285	0.85036	0.30494	0
	9	0.99063	0.98922	0.61452	0.17017	0.85139	0.33223	0
	8	0.96806	0.98925	0.61908	0.20448	0.854	0.38301	0
	7	0.94084	0.98929	0.62478	0.24787	0.85725	0.44314	0
	6	0.90305	0.98937	0.63304	0.30251	0.86171	0.51207	0
	5	0.86286	0.98945	0.64208	0.36116	0.86657	0.58166	0
	4	0.79339	0.9897	0.65899	0.44052	0.87494	0.66483	0
	3	0.70016	0.99016	0.68403	0.5342	0.88624	0.74902	0
	2	0.56474	0.99113	0.72543	0.6534	0.90602	0.84648	0
	1	0.35295	0.99348	0.80369	0.80369	0.94185	0.94185	0

Cenário: X e DManhattan

SVD

Base de dados	k	Prop Var Ac	Strain	Stress-1 (M->E)	Stress-1 (M->M)	Sstress (M->E)	Sstress (M->M)	Tempo (s)	Prop Aprox Ac
	15	1	1E-015	0.14242	0.066707	0.11484	0.089791	0.07	1
	14	1	1E-015	0.14242	0.066707	0.11484	0.089791	0.07	1

Triagem de Crédito	13	1	3.4966E-009	0.14242	0.06669	0.11484	0.089782	0.07	1
	12	1	5.6841E-009	0.14242	0.066672	0.11484	0.089774	0.07	0.99993
	11	1	9.2184E-009	0.14242	0.066647	0.11484	0.089758	0.07	0.99989
	10	1	0.000000012	0.14242	0.066622	0.11484	0.089744	0.07	0.99984
	9	1	1.4721E-008	0.14242	0.066596	0.11484	0.089728	0.07	0.99979
	8	1	2.5408E-008	0.14242	0.06656	0.11484	0.08971	0.07	0.99971
	7	1	1.6229E-007	0.14242	0.066447	0.11484	0.089657	0.07	0.9995
	6	1	7.4233E-007	0.14242	0.066262	0.11484	0.089544	0.07	0.99905
	5	1	1.1891E-006	0.14242	0.066006	0.11484	0.089377	0.07	0.99853
	4	0.99999	0.000020704	0.14243	0.064936	0.11484	0.08783	0.07	0.99611
	3	0.99917	0.001788	0.14374	0.05951	0.11489	0.076673	0.07	0.97357
	2	0.94836	0.067297	0.23423	0.2001	0.12873	0.093082	0.07	0.83533
	1	0.69359	0.37031	0.55904	0.55904	0.20307	0.20307	0.07	0.5137
Imagens	18	1	0	0.57948	0.33092	0.69137	0.45144	0.01	1
	17	1	0	0.57948	0.33092	0.69137	0.45144	0.01	1
	16	1	0	0.57948	0.33092	0.69137	0.45144	0.01	1
	15	1	0	0.57948	0.33092	0.69137	0.45144	0.01	1
	14	1	0	0.57948	0.33092	0.69137	0.45144	0.01	1
	13	1	1.3668E-008	0.57948	0.33094	0.69137	0.45145	0.01	1
	12	1	2.3587E-008	0.57948	0.33097	0.69137	0.45147	0.01	0.99989
	11	1	5.5685E-007	0.57948	0.33119	0.69137	0.45161	0.01	0.99958
	10	0.99999	9.0235E-006	0.57949	0.33199	0.69137	0.45215	0.01	0.99833
	9	0.99994	0.000029218	0.5795	0.33368	0.69137	0.45324	0.01	0.99613
	8	0.99982	0.00008347	0.57953	0.33525	0.69138	0.45459	0.01	0.99245
	7	0.99931	0.00032723	0.57968	0.3389	0.6914	0.45722	0.01	0.98503
	6	0.99734	0.0012822	0.5802	0.35088	0.69149	0.46461	0.01	0.97036
	5	0.99171	0.0037528	0.58149	0.36138	0.69209	0.47873	0.01	0.9456
	4	0.98361	0.0063384	0.58365	0.38745	0.69255	0.49548	0.01	0.91581
	3	0.85705	0.081002	0.61981	0.49499	0.69867	0.55635	0.01	0.79735
	2	0.54493	0.21221	0.76339	0.72733	0.73483	0.7013	0.01	0.61274
	1	0.14994	0.32614	0.8811	0.8811	0.9932	0.9932	0.01	0.40529
Iris	4	1	0	0.40426	0.12511	0.6426	0.21879	0.01	1
	3	0.99401	0.00044483	0.40618	0.14972	0.64361	0.25626	0.01	0.98302
	2	0.97658	0.00137	0.41284	0.20429	0.64674	0.3245	0.01	0.95405
	1	0.52819	0.034434	0.58066	0.58066	0.8055	0.8055	0.01	0.80473
iris	60	1	0	0.81119	0.34171	0.96515	0.60148	0.03	1
	59	1	0.000000726	0.81119	0.34194	0.96515	0.60175	0.03	0.99971
	58	0.99999	1.3828E-006	0.81119	0.34224	0.96515	0.60211	0.03	0.99935
	57	0.99998	2.0218E-006	0.81119	0.34258	0.96515	0.60252	0.03	0.99894
	56	0.99998	2.6542E-006	0.81119	0.34294	0.96515	0.60294	0.03	0.9985
	55	0.99997	3.5324E-006	0.81119	0.34338	0.96515	0.60347	0.03	0.99798
	54	0.99995	4.4344E-006	0.81119	0.34384	0.96515	0.604	0.03	0.99743
	53	0.99994	5.5353E-006	0.81119	0.34435	0.96515	0.60461	0.03	0.99682
	52	0.99992	6.8757E-006	0.81119	0.3449	0.96515	0.60525	0.03	0.99613
	51	0.99989	8.8094E-006	0.8112	0.34555	0.96515	0.60602	0.03	0.99534
	50	0.99986	0.000011242	0.8112	0.34625	0.96515	0.60683	0.03	0.99445
	49	0.99982	0.000015247	0.8112	0.34713	0.96515	0.60783	0.03	0.99337
	48	0.99974	0.000021788	0.81121	0.34822	0.96515	0.60908	0.03	0.99204
	47	0.99967	0.000027761	0.81122	0.34936	0.96516	0.6104	0.03	0.99064
	46	0.99957	0.000034266	0.81123	0.35064	0.96516	0.6119	0.03	0.98913
	45	0.99942	0.000047197	0.81124	0.35219	0.96516	0.6137	0.03	0.98721
	44	0.99923	0.000063216	0.81126	0.35404	0.96517	0.61589	0.03	0.98502
	43	0.999	0.00008123	0.81128	0.356	0.96518	0.61818	0.03	0.98262
	42	0.99872	0.00010109	0.81131	0.35819	0.96518	0.6207	0.03	0.98
	41	0.99836	0.00012906	0.81134	0.36066	0.96519	0.62351	0.03	0.97698
	40	0.99796	0.00015524	0.81138	0.36331	0.96521	0.62658	0.03	0.97385
	39	0.99756	0.00017859	0.81142	0.36598	0.96522	0.62963	0.03	0.97068
	38	0.99707	0.00020823	0.81146	0.36891	0.96523	0.63298	0.03	0.96719
	37	0.99653	0.00023849	0.81152	0.37191	0.96525	0.63645	0.03	0.96356
	36	0.99599	0.00026687	0.81157	0.37497	0.96526	0.63991	0.03	0.95987

	35	0.99533	0.00030269	0.81163	0.37832	0.96528	0.6436	0.03	0.95584
Sonar	34	0.99463	0.00033916	0.8117	0.38176	0.9653	0.64744	0.03	0.95167
	33	0.99389	0.00037546	0.81177	0.38525	0.96532	0.65119	0.03	0.94739
	32	0.99313	0.00041112	0.81184	0.38885	0.96534	0.65514	0.03	0.94303
	31	0.99221	0.00045743	0.81193	0.39284	0.96537	0.65954	0.03	0.93826
	30	0.99124	0.00050409	0.81202	0.39696	0.9654	0.66401	0.03	0.93335
	29	0.99001	0.00057097	0.81214	0.40152	0.96543	0.6688	0.03	0.92783
	28	0.98869	0.00063953	0.81226	0.40644	0.96547	0.67414	0.03	0.92211
	27	0.98727	0.00071149	0.8124	0.41108	0.96551	0.67905	0.03	0.91615
	26	0.9857	0.00078957	0.81255	0.41644	0.96555	0.68473	0.03	0.90992
	25	0.98352	0.00092147	0.81276	0.42272	0.96561	0.69121	0.03	0.90257
	24	0.9812	0.0010517	0.81298	0.42899	0.96568	0.69774	0.03	0.89498
	23	0.9788	0.0011755	0.81322	0.43548	0.96575	0.70444	0.03	0.88725
	22	0.97607	0.001318	0.81349	0.44245	0.96583	0.71145	0.03	0.87902
	21	0.97307	0.0014722	0.81378	0.44981	0.96591	0.71871	0.03	0.87039
	20	0.9694	0.0016762	0.81414	0.45812	0.96602	0.72685	0.03	0.86084
	19	0.96451	0.0019868	0.81461	0.46786	0.96616	0.73619	0.03	0.84983
	18	0.95921	0.0022996	0.81513	0.47767	0.9663	0.7452	0.03	0.83836
	17	0.9535	0.0026161	0.81569	0.48786	0.96646	0.75458	0.03	0.82646
	16	0.94631	0.0030519	0.81641	0.49926	0.96666	0.76473	0.03	0.81309
	15	0.93854	0.0034916	0.81718	0.51137	0.96688	0.77529	0.03	0.79921
	14	0.92975	0.0039848	0.81806	0.52427	0.96711	0.78593	0.03	0.78443
	13	0.92018	0.0044996	0.81902	0.53815	0.96737	0.79726	0.03	0.76902
	12	0.90829	0.0051954	0.82024	0.55343	0.96771	0.80959	0.03	0.75184
	11	0.89394	0.0060688	0.82171	0.57061	0.96813	0.82309	0.03	0.73296
	10	0.87741	0.0070632	0.82344	0.58948	0.96861	0.83726	0.03	0.7127
	9	0.85818	0.0082176	0.82548	0.60921	0.96915	0.85118	0.03	0.69085
	8	0.83574	0.0095673	0.82792	0.63097	0.96982	0.86624	0.03	0.66725
	7	0.80939	0.011165	0.83085	0.65469	0.9706	0.88162	0.03	0.64167
	6	0.76616	0.014623	0.83585	0.68483	0.97187	0.8993	0.03	0.60891
	5	0.72154	0.017574	0.84134	0.71675	0.97325	0.91679	0.03	0.57563
	4	0.66848	0.021051	0.84812	0.75149	0.97486	0.93381	0.03	0.53934
	3	0.59312	0.026726	0.85866	0.79324	0.97711	0.9511	0.03	0.49608
	2	0.39072	0.051665	0.89023	0.8651	0.98447	0.97694	0.03	0.42519
	1	0.0739	0.086383	0.95919	0.9977	0.9977	0.03	0.33649	
Espiral	2	1	0	0.23402	0.17744	0.423	0.34135	0	1
	1	0.53759	0.6521	0.54059	0.54059	0.70011	0.70011	0	0.51882
Vogal	10	1	1E-015	0.61266	0.13842	0.85036	0.27315	0.14	1
	9	0.98933	0.003461	0.61477	0.15582	0.85154	0.30032	0.14	0.97498
	8	0.96558	0.0084293	0.61956	0.19019	0.85433	0.35106	0.14	0.9377
	7	0.93815	0.012228	0.62531	0.23386	0.8576	0.4117	0.14	0.89768
	6	0.90026	0.017299	0.63361	0.28891	0.8621	0.48254	0.14	0.85064
	5	0.85881	0.021881	0.64296	0.34814	0.86714	0.55424	0.14	0.80142
	4	0.78657	0.031996	0.66073	0.42916	0.87596	0.64069	0.14	0.73645
	3	0.66214	0.051443	0.69465	0.54359	0.89349	0.75367	0.14	0.6511
	2	0.48507	0.077302	0.74827	0.6806	0.92381	0.87328	0.14	0.54896
	1	0.21995	0.11611	0.84926	0.84926	0.96676	0.96676	0.14	0.42379
Cenário: X e D Manhattan									
CMDS									
Base de dados	k	Prop Var Ac	Strain	Stress-1 (M->E)	Stress-1 (M->M)	Sstress (M->E)	Sstress (M->M)	Tempo (s)	
Triagem de Crédito	15	1.4441	0.88873	0.17501	1.2788	0.087105	1.3544	2.01	
	14	1.4418	0.88873	0.17386	1.2447	0.086741	1.318	2.01	
	13	1.4393	0.88872	0.17243	1.2102	0.086358	1.2798	2.01	
	12	1.4361	0.88871	0.17055	1.1684	0.085879	1.2326	2.01	
	11	1.4322	0.8887	0.16817	1.1242	0.085291	1.1845	2.01	
	10	1.427	0.88867	0.16505	1.0718	0.084471	1.1301	2.01	
	9	1.4215	0.88865	0.16202	1.0204	0.083701	1.0829	2.01	
	8	1.4141	0.8886	0.15801	0.9578	0.081911	1.0115	2.01	
	7	1.4024	0.88849	0.15164	0.88112	0.080379	0.94338	2.01	

	6	1.3901	0.88842	0.14597	0.80063	0.07859	0.86565	2.01
	5	1.3726	0.88814	0.13529	0.70447	0.075391	0.76558	2.01
	4	1.3468	0.88777	0.12241	0.59432	0.069338	0.63427	2.01
	3	1.3141	0.88735	0.1104	0.45149	0.061235	0.44422	2.01
	2	1.2291	0.88718	0.12568	0.29124	0.059025	0.31364	2.01
	1	1.0381	0.90721	0.19702	0.19702	0.069624	0.069624	2.01
Imagens	18	6.1755	2.0468	0.15211	1.6886	0.11912	3.4747	0.07
	17	6.1623	2.0468	0.15044	1.6501	0.11828	3.3971	0.07
	16	6.148	2.0468	0.14884	1.609	0.11728	3.3087	0.07
	15	6.1302	2.0467	0.14682	1.5632	0.11606	3.2148	0.07
	14	6.1073	2.0467	0.14416	1.5125	0.11455	3.1137	0.07
	13	6.0839	2.0466	0.14151	1.46	0.11293	3.0096	0.07
	12	6.06	2.0466	0.13889	1.4058	0.11104	2.8966	0.07
	11	6.0315	2.0465	0.1358	1.349	0.10913	2.7908	0.07
	10	5.9953	2.0464	0.13197	1.2847	0.10672	2.6727	0.07
	9	5.9525	2.0462	0.1272	1.2129	0.10369	2.5361	0.07
	8	5.9008	2.046	0.12222	1.1337	0.099789	2.3909	0.07
	7	5.8316	2.0456	0.11508	1.0441	0.094475	2.2382	0.07
	6	5.751	2.045	0.10729	0.94772	0.08896	2.0795	0.07
	5	5.6178	2.0433	0.096872	0.82323	0.081028	1.8829	0.07
	4	5.4503	2.0407	0.085433	0.67983	0.070996	1.6279	0.07
	3	5.0087	2.0294	0.11186	0.47289	0.07281	1.2238	0.07
	2	4.223	1.9948	0.20382	0.27286	0.12289	0.68266	0.07
	1	2.9368	1.9085	0.39624	0.39624	0.7307	0.7307	0.07
Iris	4	2.8997	1.0051	0.057181	0.42498	0.052935	0.85488	0.05
	3	2.8529	1.0051	0.060329	0.33087	0.050404	0.66043	0.05
	2	2.7832	1.005	0.059065	0.20747	0.048215	0.40405	0.05
	1	2.5485	1.005	0.1247	0.1247	0.10706	0.10706	0.05
Sonar	60	34.422	2.7574	0.12207	4.0829	0.21141	22.387	0.09
	59	34.387	2.7574	0.12147	4.0554	0.2104	22.146	0.09
	58	34.351	2.7574	0.12085	4.027	0.20935	21.899	0.09
	57	34.312	2.7574	0.12016	3.9981	0.20822	21.651	0.09
	56	34.269	2.7574	0.11939	3.9674	0.20698	21.391	0.09
	55	34.225	2.7574	0.11863	3.9361	0.20568	21.123	0.09
	54	34.179	2.7574	0.11782	3.904	0.20436	20.852	0.09
	53	34.132	2.7573	0.117	3.8724	0.203	20.586	0.09
	52	34.083	2.7573	0.11616	3.8392	0.2015	20.303	0.09
	51	34.032	2.7573	0.11525	3.8057	0.20001	20.025	0.09
	50	33.976	2.7573	0.11431	3.7705	0.1984	19.73	0.09
	49	33.918	2.7573	0.1133	3.7345	0.19673	19.435	0.09
	48	33.858	2.7572	0.11227	3.6977	0.19499	19.135	0.09
	47	33.796	2.7572	0.11119	3.6602	0.19317	18.83	0.09
	46	33.732	2.7572	0.11007	3.6224	0.19124	18.523	0.09
	45	33.665	2.7571	0.10891	3.5841	0.18934	18.218	0.09
	44	33.596	2.7571	0.10772	3.5449	0.18738	17.911	0.09
	43	33.524	2.757	0.10649	3.5058	0.18539	17.606	0.09
	42	33.448	2.757	0.10518	3.4648	0.18326	17.289	0.09
	41	33.364	2.7569	0.10376	3.4209	0.18083	16.946	0.09
	40	33.277	2.7569	0.10228	3.3767	0.17822	16.599	0.09
	39	33.187	2.7568	0.10073	3.331	0.17553	16.247	0.09
	38	33.095	2.7567	0.099198	3.2853	0.17293	15.901	0.09
	37	32.997	2.7566	0.09755	3.238	0.17001	15.537	0.09
	36	32.897	2.7565	0.095921	3.1907	0.1672	15.185	0.09
	35	32.79	2.7565	0.094184	3.142	0.1642	14.825	0.09
	34	32.68	2.7564	0.092393	3.0919	0.16106	14.453	0.09
	33	32.566	2.7562	0.090479	3.0413	0.15778	14.09	0.09
	32	32.447	2.7561	0.088472	2.9905	0.15447	13.735	0.09
	31	32.319	2.756	0.086288	2.9366	0.15064	13.349	0.09
	30	32.177	2.7558	0.084053	2.8802	0.14672	12.955	0.09
	29	32.031	2.7556	0.081751	2.8224	0.14276	12.556	0.09

28	31.876	2.7554	0.079239	2.7641	0.13833	12.161	0.09	
27	31.71	2.7552	0.076547	2.7024	0.13356	11.745	0.09	
26	31.533	2.7549	0.073911	2.6398	0.12894	11.332	0.09	
25	31.348	2.7547	0.071235	2.5774	0.12419	10.926	0.09	
24	31.154	2.7543	0.068825	2.5133	0.11997	10.52	0.09	
23	30.936	2.7539	0.06564	2.4436	0.11459	10.091	0.09	
22	30.709	2.7535	0.062477	2.3723	0.10901	9.6489	0.09	
21	30.459	2.753	0.059536	2.2983	0.10361	9.2028	0.09	
20	30.2	2.7524	0.056478	2.2224	0.09811	8.7602	0.09	
19	29.925	2.7518	0.053576	2.1443	0.092788	8.3078	0.09	
18	29.646	2.7511	0.052039	2.0685	0.089355	7.8907	0.09	
17	29.34	2.7503	0.049706	1.9851	0.084591	7.4417	0.09	
16	29.017	2.7494	0.048752	1.9001	0.081699	6.9956	0.09	
15	28.668	2.7484	0.049391	1.813	0.080948	6.5465	0.09	
14	28.247	2.7469	0.050439	1.7151	0.080901	6.064	0.09	
13	27.814	2.7453	0.052869	1.6163	0.082608	5.5834	0.09	
12	27.32	2.7433	0.058367	1.5108	0.089073	5.0923	0.09	
11	26.79	2.741	0.064621	1.4018	0.09673	4.6007	0.09	
10	26.22	2.7382	0.072606	1.2894	0.10728	4.1158	0.09	
9	25.577	2.7348	0.084006	1.1698	0.12218	3.6358	0.09	
8	24.834	2.7302	0.097682	1.0422	0.14199	3.1303	0.09	
7	24.027	2.7248	0.11568	0.91396	0.16524	2.6727	0.09	
6	22.908	2.7146	0.14487	0.76827	0.20353	2.1851	0.09	
5	21.569	2.6999	0.17899	0.61165	0.24879	1.6877	0.09	
4	20.211	2.6846	0.21337	0.46116	0.29327	1.2343	0.09	
3	17.889	2.6394	0.27761	0.30939	0.36596	0.74842	0.09	
2	15.3	2.5816	0.35858	0.29753	0.4489	0.44377	0.09	
1	9.8992	2.3185	0.53959	0.53959	0.64311	0.64311	0.09	
Espiral	2	1.6518	0.71251	0.11799	0.3346	0.18131	0.6915	0.01
	1	0.90493	0.85681	0.47939	0.47939	0.5879	0.5879	0.01
Vogal	10	7.2223	1.2861	0.075847	1.5331	0.12955	5.1496	6.06
	9	7.0415	1.2849	0.069438	1.4013	0.11722	4.5553	6.06
	8	6.8069	1.283	0.067717	1.2533	0.11136	3.9403	6.06
	7	6.5362	1.2803	0.071435	1.0921	0.1135	3.2921	6.06
	6	6.2136	1.277	0.088193	0.91847	0.13484	2.6515	6.06
	5	5.8593	1.2728	0.11326	0.74279	0.16725	2.0632	6.06
	4	5.3976	1.2666	0.15918	0.55614	0.22246	1.5026	6.06
	3	4.7157	1.2524	0.2374	0.37025	0.31315	0.95648	6.06
	2	3.5902	1.2117	0.36699	0.3113	0.47329	0.48385	6.06
	1	1.9217	1.1195	0.58499	0.58499	0.72827	0.72827	6.06

Cenário: X e DManhattan

MDS Stress-1

Base de dados	k	Prop Var Ac	Strain	Stress-1 (M->E)	Stress-1 (M->M)	Sstress (M->E)	Sstress (M->M)	Tempo (s)
Triagem de Crédito	15	1.1755	0.7881	0.032892	0.47049	0.02557	0.63694	133.28
	14	1.1755	0.78609	0.032895	0.47247	0.025663	0.63451	114.3
	13	1.1754	0.78565	0.032895	0.47368	0.025681	0.6543	125.28
	12	1.1755	0.787	0.032891	0.4607	0.025601	0.61634	130.55
	11	1.1754	0.78884	0.032892	0.46902	0.025622	0.64009	125.69
	10	1.1753	0.78706	0.032891	0.45761	0.025603	0.61059	143.67
	9	1.1755	0.78509	0.032895	0.45571	0.025663	0.64627	140.82
	8	1.1754	0.78558	0.032895	0.46618	0.025639	0.64961	121.38
	7	1.1753	0.78748	0.032891	0.44374	0.025565	0.58129	152.07
	6	1.1754	0.79011	0.032897	0.42872	0.025567	0.57391	121.75
	5	1.1754	0.79258	0.032896	0.41678	0.02556	0.53288	138.56
	4	1.1752	0.79295	0.03293	0.39501	0.025682	0.51668	130.67
	3	1.1756	0.8077	0.033357	0.34384	0.025814	0.40809	99.98
	2	1.1735	0.84494	0.050809	0.23126	0.033557	0.34154	59.64
	1	1.1286	1.7157	0.20223	0.20223	0.19427	0.19427	7.25
	18	5.0303	1.9341	0.047383	0.80404	0.054806	1.7318	21.38

Imagens	17	5.0299	1.934	0.047383	0.80497	0.054811	1.7347	10.09
	16	5.0303	1.9341	0.047383	0.80399	0.054806	1.7338	10.4
	15	5.0302	1.9342	0.047383	0.80431	0.054801	1.7351	10.32
	14	5.03	1.934	0.047383	0.8029	0.054815	1.7332	11.17
	13	5.0302	1.9342	0.047383	0.80307	0.054805	1.7329	10.1
	12	5.0302	1.9341	0.047383	0.80132	0.054808	1.7291	13.04
	11	5.0302	1.9342	0.047383	0.79448	0.054804	1.714	16.07
	10	5.0302	1.9341	0.047383	0.79877	0.054813	1.7279	18.02
	9	5.0299	1.9339	0.047383	0.78637	0.054802	1.7028	17.82
	8	5.0299	1.9343	0.047398	0.76969	0.054841	1.6793	10.16
	7	5.03	1.9356	0.047459	0.7385	0.054722	1.6346	6.85
	6	5.0301	1.9381	0.047619	0.70079	0.054647	1.5673	7.73
	5	5.0297	1.9453	0.048257	0.63995	0.055165	1.433	6.01
	4	5.0287	1.9625	0.049985	0.5597	0.056715	1.2854	4.6
	3	5.0234	2.0522	0.059411	0.44663	0.08427	1.0628	8.94
	2	4.9644	2.2765	0.12357	0.3101	0.16267	0.86026	3.63
	1	4.186	2.7206	0.41197	0.41197	0.42539	0.42539	1.64
Iris	4	2.7984	1.0049	0.026972	0.33827	0.038986	0.63416	2.57
	3	2.798	1.005	0.029816	0.27777	0.04075	0.51755	2.71
	2	2.7935	1.0053	0.050039	0.19788	0.05099	0.36702	1
	1	1.9484	1.009	0.55159	0.55159	0.77776	0.77776	0.41
Sonar	60	27.906	2.6363	0.025292	2.4723	0.048842	10.174	27.5
	59	27.906	2.6363	0.025292	2.4728	0.048842	10.177	36.6
	58	27.905	2.6361	0.025292	2.4724	0.048849	10.174	47.57
	57	27.906	2.6363	0.025292	2.4731	0.048843	10.178	47.12
	56	27.905	2.6361	0.025292	2.4722	0.04885	10.174	50.51
	55	27.906	2.6363	0.025292	2.4707	0.048842	10.164	32.77
	54	27.906	2.6363	0.025292	2.4711	0.048841	10.167	53.76
	53	27.906	2.6363	0.025292	2.47	0.048842	10.161	36.1
	52	27.905	2.6362	0.025292	2.4703	0.048848	10.161	56.98
	51	27.905	2.6361	0.025292	2.4707	0.048849	10.164	34.82
	50	27.906	2.6363	0.025292	2.4699	0.048841	10.159	52.82
	49	27.905	2.6362	0.025292	2.4662	0.048848	10.137	41.65
	48	27.906	2.6363	0.025292	2.4692	0.048841	10.155	53.58
	47	27.905	2.6362	0.025292	2.4658	0.04885	10.134	38.3
	46	27.905	2.6362	0.025292	2.4625	0.048847	10.114	56.33
	45	27.906	2.6363	0.025292	2.4611	0.04884	10.105	37.4
	44	27.906	2.6363	0.025292	2.4622	0.048841	10.112	53.05
	43	27.905	2.6362	0.025292	2.4643	0.048847	10.124	25.7
	42	27.906	2.6363	0.025292	2.4612	0.048841	10.106	53.2
	41	27.905	2.6362	0.025292	2.4562	0.048847	10.074	40.13
	40	27.905	2.6362	0.025292	2.4512	0.048846	10.044	55.46
	39	27.906	2.6362	0.025292	2.4545	0.048845	10.064	34.26
	38	27.906	2.6363	0.025292	2.4476	0.048839	10.022	49.73
	37	27.906	2.6364	0.025292	2.44	0.048839	9.9752	32.03
	36	27.906	2.6364	0.025293	2.4355	0.048836	9.948	53.73
	35	27.905	2.6363	0.025292	2.4347	0.048845	9.9424	40.09
	34	27.906	2.6364	0.025293	2.4256	0.048837	9.888	49.54
	33	27.906	2.6366	0.025295	2.4105	0.048835	9.7962	48.36
	32	27.906	2.6367	0.025298	2.3962	0.048835	9.7083	50.31
	31	27.905	2.6369	0.025304	2.3775	0.048829	9.5979	59.71
	30	27.906	2.6373	0.025314	2.356	0.048827	9.4687	28.26
	29	27.906	2.6377	0.025325	2.3337	0.048824	9.3379	57.97
	28	27.905	2.6383	0.025348	2.307	0.048836	9.1783	28.91
	27	27.906	2.6391	0.025382	2.2795	0.04885	9.0177	50.33
	26	27.906	2.64	0.025433	2.2506	0.048893	8.8485	43.53
	25	27.904	2.6411	0.025506	2.2175	0.048955	8.6545	43.38
	24	27.905	2.6428	0.025613	2.1838	0.04905	8.4603	44.3
	23	27.905	2.6447	0.025749	2.1471	0.049178	8.2547	35.45
	22	27.904	2.6473	0.025943	2.1083	0.049434	8.0403	39.72

21	27.904	2.65	0.02615	2.0681	0.049604	7.8162	20.86	
20	27.905	2.6534	0.0264	2.0257	0.049857	7.5871	38.92	
19	27.902	2.6571	0.026738	1.9776	0.050222	7.3299	35.99	
18	27.903	2.6619	0.027223	1.9302	0.050739	7.0841	19.78	
17	27.901	2.6673	0.02777	1.8779	0.051354	6.8086	35.01	
16	27.9	2.6737	0.028461	1.8241	0.052172	6.5359	37.04	
15	27.9	2.6817	0.02946	1.7676	0.053303	6.2553	13.24	
14	27.896	2.6933	0.030721	1.7017	0.055032	5.9297	33.1	
13	27.893	2.7069	0.032253	1.6346	0.057156	5.6121	27	
12	27.891	2.7254	0.034396	1.5624	0.06035	5.2752	25.55	
11	27.884	2.747	0.037084	1.4826	0.063836	4.9082	12.65	
10	27.879	2.7739	0.04026	1.4016	0.068231	4.557	16.14	
9	27.866	2.81	0.044856	1.3093	0.074994	4.1604	24.17	
8	27.85	2.854	0.050736	1.2086	0.08323	3.7512	18.18	
7	27.824	2.9145	0.059074	1.1013	0.09442	3.3382	15.98	
6	27.771	3.0092	0.07321	0.97827	0.11603	2.8867	6.02	
5	27.699	3.1388	0.090023	0.84005	0.13948	2.4002	5.08	
4	27.543	3.3313	0.11661	0.69554	0.17585	1.9441	3.19	
3	27.153	3.6461	0.1659	0.53007	0.24383	1.4407	7.19	
2	26.168	4.1549	0.25089	0.38507	0.36409	0.95513	13.08	
1	20.634	4.611	0.51091	0.51091	0.79454	0.79454	6.92	
Espiral	2	1.6168	0.6272	0.092745	0.33266	0.1637	0.72464	0.34
	1	1.3593	1.4357	0.40798	0.40798	0.54481	0.54481	0.51
Vogal	10	6.5778	1.2502	0.046866	1.378	0.082891	4.3904	262.9
	9	6.5768	1.2556	0.048315	1.2987	0.084721	4.0642	267.04
	8	6.5745	1.2657	0.051877	1.2044	0.089381	3.6878	234.62
	7	6.5709	1.2789	0.057073	1.1003	0.096485	3.2908	242.05
	6	6.5635	1.2999	0.066166	0.98268	0.10925	2.8651	233.68
	5	6.5499	1.3298	0.07975	0.8541	0.12754	2.4327	134.74
	4	6.5162	1.38	0.10744	0.70701	0.1617	1.9721	163.64
	3	6.4257	1.468	0.15914	0.5447	0.23454	1.5009	123.78
	2	6.1805	1.6314	0.24982	0.38357	0.37494	0.95508	105.29
	1	4.9154	1.8597	0.50433	0.50433	0.77033	0.77033	165.55

Cenário: X e DManhattan

MDS Sstress

Base de dados	k	Prop Var Ac	Strain	Stress-1 (M->E)	Stress-1 (M->M)	Sstress (M->E)	Sstress (M->M)	Tempo (s)
Triagem de Crédito	15	1.2003	0.85165	0.048438	0.43235	0.02011	0.39872	1029.3
	14	1.2002	0.852	0.04832	0.41083	0.020102	0.37249	1018.4
	13	1.183	1.0839	1.0285	2.6231	0.99024	1.6131	1.08
	12	1.2001	0.85194	0.048327	0.41008	0.020104	0.37241	1046.8
	11	1.2004	0.85146	0.048462	0.40913	0.020099	0.37815	1076.8
	10	1.2003	0.85206	0.048376	0.40267	0.020102	0.36473	1027.3
	9	1.2002	0.85227	0.048375	0.39901	0.020105	0.36311	1041
	8	1.2002	0.85191	0.048403	0.39941	0.020101	0.37013	1045.2
	7	1.19	1.1545	1.0469	1.9669	0.9938	1.2272	0.33
	6	1.2	0.85207	0.04834	0.375	0.020105	0.33563	1037.4
	5	1.1992	0.85301	0.047908	0.36201	0.020141	0.33214	1007.3
	4	1.1999	0.8523	0.048348	0.369	0.020108	0.33714	1023.2
	3	1.175	1.3252	1.0525	1.3868	0.99602	1.0466	0.22
	2	1.1731	1.4622	1.0898	1.2655	1.0056	1.0343	0.2
	1	1.0845	1.0712	0.15863	0.15863	0.098934	0.098934	227.71
	18	5.1491	1.9486	0.0608	0.77446	0.042626	1.703	37.04
	17	5.1492	1.9488	0.060804	0.77319	0.042626	1.7014	72.79
	16	5.1487	1.9488	0.060814	0.76563	0.042625	1.6908	39.93
	15	5.1493	1.9487	0.060808	0.77163	0.042626	1.6993	76.84
	14	5.1494	1.9486	0.060823	0.7685	0.042625	1.6963	41.92
	13	5.1494	1.9488	0.060774	0.76946	0.042626	1.6961	65.82
	12	5.1489	1.9488	0.06082	0.76647	0.042625	1.6925	45
	11	5.1488	1.9489	0.06082	0.76279	0.042625	1.6876	61.24

Imagens	10	5.1492	1.9488	0.060818	0.76007	0.042625	1.6843	46.73
	9	5.1488	1.9488	0.06081	0.7614	0.042626	1.6853	45.83
	8	5.1486	1.9488	0.060807	0.76417	0.042626	1.6896	62.8
	7	5.1493	1.9487	0.060826	0.75591	0.042625	1.6793	49.61
	6	5.1486	1.9488	0.060834	0.74591	0.042625	1.6642	77.6
	5	5.1409	1.9546	0.061552	0.69159	0.042861	1.5917	76.05
	4	5.121	1.9668	0.062342	0.60787	0.043716	1.4436	36.29
	3	4.9987	2.0228	0.096458	0.46981	0.059182	1.1996	24.31
	2	4.6696	2.166	0.15151	0.27748	0.10583	0.72336	61.39
	1	4.0671	2.5311	0.27541	0.27541	0.22686	0.22686	10.69
Iris	4	2.8173	1.0048	0.037623	0.35973	0.032337	0.68417	47.77
	3	2.8084	1.0049	0.043156	0.29714	0.03395	0.56349	17.92
	2	2.7886	1.0051	0.050761	0.20406	0.038173	0.38268	9.11
	1	2.6582	1.006	0.11825	0.11825	0.09194	0.09194	2.46
Sonar	60	28.08	2.6685	0.02697	2.3745	0.046215	9.6027	92.47
	59	28.079	2.6685	0.02697	2.3755	0.046215	9.6086	88.68
	58	28.079	2.6685	0.02697	2.3757	0.046215	9.6106	82.22
	57	28.079	2.6684	0.02697	2.3769	0.046216	9.6168	80.03
	56	28.079	2.6684	0.026969	2.376	0.046216	9.6111	77.84
	55	28.08	2.6685	0.02697	2.3746	0.046215	9.6027	78.07
	54	28.079	2.6684	0.026969	2.3748	0.046215	9.6046	75.64
	53	28.08	2.6685	0.02697	2.3752	0.046215	9.6075	78.95
	52	28.079	2.6685	0.026968	2.3746	0.046215	9.6038	74.8
	51	28.079	2.6684	0.02697	2.373	0.046215	9.5929	73.35
	50	28.079	2.6684	0.02697	2.3731	0.046215	9.595	75.63
	49	28.079	2.6685	0.026969	2.3713	0.046215	9.5838	71.9
	48	28.079	2.6684	0.026969	2.3737	0.046215	9.5981	71.03
	47	28.079	2.6684	0.026969	2.3732	0.046215	9.5958	72.71
	46	28.079	2.6685	0.026968	2.3708	0.046215	9.5813	73.03
	45	28.079	2.6685	0.02697	2.368	0.046215	9.5635	65.18
	44	28.08	2.6685	0.026971	2.3714	0.046215	9.5858	79.49
	43	28.079	2.6685	0.026971	2.3641	0.046215	9.5394	84.4
	42	28.079	2.6684	0.026969	2.3674	0.046215	9.5606	61.16
	41	28.079	2.6685	0.026969	2.3655	0.046215	9.5492	63.28
	40	28.079	2.6685	0.02697	2.3583	0.046215	9.5033	82.63
	39	28.08	2.6685	0.02697	2.363	0.046215	9.5335	84.23
	38	28.08	2.6686	0.026971	2.3601	0.046215	9.5158	77.5
	37	28.079	2.6686	0.026971	2.3596	0.046216	9.5124	73.05
	36	28.079	2.6685	0.026972	2.3536	0.046215	9.4764	74.79
	35	28.079	2.6685	0.026971	2.3548	0.046215	9.482	67.15
	34	28.079	2.6686	0.026973	2.3446	0.046215	9.4204	82.4
	33	28.079	2.6686	0.026972	2.3424	0.046215	9.4094	82.95
	32	28.079	2.6688	0.026978	2.3319	0.046218	9.3466	86.01
	31	28.079	2.6687	0.026975	2.3284	0.046217	9.3258	80.79
	30	28.079	2.6688	0.026979	2.3195	0.046219	9.2736	74.29
	29	28.077	2.669	0.026987	2.3044	0.046226	9.1848	53.34
	28	28.077	2.6697	0.027016	2.2779	0.04624	9.0243	70.19
	27	28.076	2.6701	0.027036	2.2575	0.04625	8.9086	52.26
	26	28.074	2.6709	0.027108	2.2317	0.046309	8.7599	52.26
	25	28.072	2.6719	0.027166	2.2018	0.046362	8.583	53.78
	24	28.067	2.6737	0.027263	2.1677	0.046448	8.3871	64.96
	23	28.063	2.6753	0.027394	2.1337	0.046563	8.1962	48.85
	22	28.059	2.6773	0.027577	2.0964	0.04673	7.9854	47.77
	21	28.052	2.6803	0.027813	2.0562	0.046989	7.759	49.37
	20	28.045	2.6835	0.028087	2.0116	0.047303	7.5151	48.21
	19	28.036	2.6871	0.028385	1.9662	0.047617	7.2735	49.09
	18	28.024	2.6918	0.028884	1.9188	0.048123	7.0273	48.46
	17	28.011	2.6968	0.029449	1.8691	0.048731	6.77	48.15
	16	27.997	2.7033	0.030147	1.814	0.049548	6.4867	46.58
	15	27.977	2.7111	0.031147	1.7555	0.05067	6.1934	40.01

14	27.951	2.7218	0.032518	1.6892	0.05232	5.8693	45.67	
13	27.92	2.7342	0.034123	1.6237	0.054314	5.5564	44.83	
12	27.879	2.7504	0.036433	1.5474	0.057212	5.2024	42.93	
11	27.827	2.7699	0.039287	1.4691	0.06066	4.849	43.74	
10	27.766	2.7943	0.042478	1.3849	0.064826	4.4825	43.23	
9	27.676	2.8241	0.047739	1.2912	0.07081	4.0927	33.64	
8	27.566	2.8624	0.053905	1.1878	0.078591	3.6714	24.48	
7	27.409	2.9108	0.06285	1.0769	0.08887	3.2452	19.51	
6	27.172	2.9822	0.075691	0.94999	0.10668	2.7805	43.63	
5	26.815	3.0817	0.09514	0.80551	0.13095	2.273	30.6	
4	26.259	3.2143	0.12361	0.65388	0.16251	1.7935	19.03	
3	25.178	3.4197	0.17659	0.47875	0.22118	1.2437	15.63	
2	23.188	3.7031	0.26427	0.34357	0.30522	0.73315	17.46	
1	17.869	3.9672	0.4423	0.4423	0.50518	0.50518	12.15	
Espiral	2	1.6352	0.67638	0.10019	0.33143	0.15362	0.69786	1.47
	1	1.2242	1.3033	0.40471	0.40471	0.50485	0.50485	3.48
Vogal	10	6.6465	1.2559	0.048504	1.3897	0.080188	4.4501	1485.7
	9	6.6337	1.2611	0.049825	1.3076	0.082177	4.1045	802.57
	8	6.6104	1.2696	0.053409	1.21	0.086846	3.7142	1060.8
	7	6.5787	1.2813	0.058685	1.0989	0.093887	3.2839	1100.8
	6	6.5283	1.2989	0.068163	0.97315	0.10604	2.8234	1108.1
	5	6.4528	1.3225	0.082541	0.83495	0.12272	2.3533	727.73
	4	6.3173	1.3571	0.11192	0.67992	0.15328	1.8683	559.98
	3	6.0226	1.4122	0.16662	0.50347	0.21789	1.3512	646.14
	2	5.431	1.4964	0.26337	0.34236	0.33717	0.7626	485.49
	1	3.9479	1.5651	0.48807	0.48807	0.6065	0.6065	1205.1

Cenário: X e D Manhattan

FastMap

Base de dados	k	Prop Var Ac	Strain	Stress-1 (M->E)	Stress-1 (M->M)	Sstress (M->E)	Sstress (M->M)	Tempo (s)
Triagem de Crédito	15	1	4.1277	0.14242	0.33679	0.11484	0.68949	0.04
	14	1	4.1277	0.14242	0.33679	0.11484	0.68949	0
	13	1	4.1277	0.14242	0.33677	0.11484	0.68948	0.01
	12	1	4.1277	0.14242	0.33675	0.11484	0.68946	0
	11	1	4.1277	0.14242	0.33672	0.11484	0.68945	0
	10	1	4.1277	0.14242	0.3367	0.11484	0.68943	0
	9	1	4.1277	0.14242	0.33666	0.11484	0.6894	0
	8	1	4.1277	0.14242	0.33662	0.11484	0.68937	0
	7	1	4.1277	0.14242	0.33656	0.11484	0.68933	0
	6	1	4.1277	0.14242	0.33637	0.11484	0.68921	0
	5	1	4.1277	0.14242	0.3357	0.11484	0.68849	0
	4	1	4.1277	0.14242	0.33521	0.11484	0.68786	0
	3	0.99752	4.1245	0.1446	0.30628	0.11517	0.64781	0
	2	0.8828	4.07	0.27321	0.25817	0.14384	0.37476	0
	1	0.67276	2.3421	0.48953	0.48953	0.20668	0.20668	0
Imagens	18	1	0.94257	0.57948	0.29639	0.69137	0.47208	0.02
	17	1	0.94257	0.57948	0.29639	0.69137	0.47208	0.02
	16	1	0.94257	0.57948	0.29639	0.69137	0.47208	0
	15	1	0.94257	0.57948	0.29639	0.69137	0.47208	0.01
	14	1	0.94257	0.57948	0.29639	0.69137	0.47208	0.01
	13	1	0.94257	0.57948	0.29642	0.69137	0.47209	0
	12	1	0.94257	0.57948	0.29644	0.69137	0.47211	0
	11	1	0.94256	0.57948	0.29662	0.69137	0.47224	0.01
	10	0.99995	0.94249	0.57949	0.29795	0.69137	0.47335	0
	9	0.99993	0.94237	0.5795	0.29905	0.69137	0.47428	0
	8	0.9995	0.94151	0.57961	0.30358	0.69139	0.47718	0
	7	0.99768	0.93356	0.58008	0.31419	0.69147	0.48395	0.01
	6	0.99611	0.93088	0.5805	0.32368	0.69154	0.48984	0
	5	0.98694	0.90795	0.58312	0.35214	0.69206	0.50796	0
	4	0.93432	0.78431	0.59809	0.42226	0.69512	0.55217	0

	3	0.88264	0.68263	0.61643	0.50043	0.69838	0.59436	0
	2	0.67279	0.71948	0.69809	0.66177	0.71135	0.65938	0
	1	0.39363	0.94804	0.86954	0.86954	0.73917	0.73917	0
Iris	4	1	0.7438	0.40426	0.24124	0.6426	0.44877	0
	3	0.99284	0.74872	0.40665	0.26828	0.64381	0.48453	0
	2	0.97261	0.75822	0.4142	0.32451	0.64802	0.5497	0
	1	0.90714	0.79685	0.43991	0.43991	0.66288	0.66288	0
Sonar	60	1	1.603	0.81119	0.28238	0.96515	0.52688	0.05
	59	0.99999	1.603	0.81119	0.28275	0.96515	0.52737	0
	58	0.99997	1.603	0.81119	0.28332	0.96515	0.52811	0
	57	0.99995	1.6029	0.81119	0.28385	0.96515	0.52879	0
	56	0.99992	1.6029	0.81119	0.28449	0.96515	0.52962	0
	55	0.99988	1.6028	0.8112	0.28515	0.96515	0.53047	0
	54	0.99984	1.6027	0.8112	0.28595	0.96515	0.53151	0
	53	0.99979	1.6026	0.81121	0.28676	0.96515	0.53253	0
	52	0.99972	1.6024	0.81121	0.28774	0.96516	0.53378	0.05
	51	0.99964	1.6022	0.81122	0.28878	0.96516	0.53511	0.01
	50	0.99954	1.6019	0.81123	0.28995	0.96516	0.53659	0
	49	0.9994	1.6016	0.81124	0.29146	0.96516	0.53851	0
	48	0.99914	1.601	0.81127	0.29333	0.96517	0.54086	0
	47	0.99883	1.6005	0.8113	0.29541	0.96518	0.54346	0
	46	0.99836	1.5996	0.81134	0.29804	0.96519	0.54675	0
	45	0.99798	1.5981	0.81138	0.30035	0.9652	0.54963	0.01
	44	0.99755	1.5965	0.81142	0.30284	0.96521	0.55273	0
	43	0.99707	1.5955	0.81146	0.3056	0.96523	0.55619	0
	42	0.99641	1.5937	0.81152	0.30869	0.96524	0.55998	0
	41	0.99585	1.5915	0.81158	0.31165	0.96526	0.56358	0
	40	0.99479	1.5882	0.81168	0.31576	0.96529	0.56867	0
	39	0.99358	1.5846	0.81179	0.32009	0.96532	0.57396	0.01
	38	0.99295	1.5826	0.81185	0.3232	0.96534	0.57775	0
	37	0.99198	1.578	0.81194	0.32721	0.96537	0.58269	0
	36	0.99088	1.5729	0.81205	0.33144	0.9654	0.58782	0
	35	0.98971	1.5697	0.81216	0.33556	0.96543	0.59263	0
	34	0.98818	1.563	0.8123	0.34053	0.96548	0.5985	0
	33	0.98661	1.5568	0.81246	0.34535	0.96552	0.60403	0
	32	0.98471	1.5488	0.81264	0.35051	0.96557	0.60994	0
	31	0.98362	1.5415	0.81274	0.35472	0.9656	0.61489	0
	30	0.98179	1.5369	0.81291	0.35988	0.96565	0.6208	0
	29	0.9792	1.5231	0.81317	0.36651	0.96573	0.62837	0
	28	0.97614	1.5065	0.81346	0.37372	0.96581	0.63643	0.01
	27	0.97286	1.4915	0.81377	0.38126	0.96591	0.6448	0
	26	0.96864	1.4791	0.81418	0.38978	0.96603	0.65402	0
	25	0.96593	1.4623	0.81445	0.39673	0.9661	0.66157	0
	24	0.96186	1.4513	0.81484	0.40548	0.96622	0.67086	0
	23	0.95622	1.4331	0.81539	0.41599	0.9664	0.68245	0
	22	0.95184	1.418	0.81583	0.42474	0.96653	0.69135	0
	21	0.94646	1.3993	0.81636	0.43473	0.96668	0.70139	0
	20	0.93948	1.3722	0.81705	0.4458	0.96687	0.71212	0
	19	0.93105	1.3455	0.81788	0.45749	0.9671	0.72341	0.01
	18	0.92107	1.3112	0.81887	0.47147	0.96742	0.73725	0
	17	0.91247	1.2894	0.81975	0.484	0.96766	0.74884	0
	16	0.90428	1.2426	0.82057	0.4961	0.96789	0.75981	0.02
	15	0.88521	1.2027	0.82251	0.51637	0.96849	0.77842	0.02
	14	0.87426	1.1348	0.82365	0.53053	0.96879	0.79023	0.01
	13	0.8632	1.0897	0.82481	0.5457	0.96917	0.80374	0.02
	12	0.8489	1.0427	0.82633	0.56214	0.96958	0.81668	0
	11	0.83349	0.9877	0.82799	0.5797	0.97003	0.82999	0
	10	0.81154	0.93533	0.83045	0.60062	0.97066	0.84482	0
	9	0.7884	0.85055	0.833	0.62213	0.97134	0.86005	0
	8	0.76239	0.768	0.83592	0.64547	0.97208	0.87529	0

	7	0.71443	0.62321	0.84149	0.67757	0.97358	0.89532	0
	6	0.67785	0.53462	0.84601	0.70553	0.97466	0.91099	0
	5	0.63431	0.42837	0.85179	0.73654	0.97595	0.92651	0
	4	0.58814	0.35014	0.85789	0.76813	0.97738	0.94152	0
	3	0.5236	0.32678	0.86761	0.80673	0.97929	0.9565	0
	2	0.43216	0.38289	0.88238	0.85151	0.98245	0.97192	0
	1	0.247	0.46924	0.91954	0.91954	0.98952	0.98952	0
Espiral	2	1	7.378	0.23402	2.4096E-006	0.423	4.31E-006	0
	1	0.51548	4.3637	0.54508	0.54508	0.72659	0.72659	0
Vogal	10	1	2.1981	0.61266	0.12886	0.85036	0.25267	0.04
	9	0.98348	2.1667	0.61598	0.15183	0.85236	0.29065	0
	8	0.95521	2.0729	0.62184	0.1902	0.85569	0.34735	0
	7	0.91848	1.9636	0.6297	0.24119	0.86011	0.4174	0
	6	0.8721	1.7805	0.63989	0.30257	0.866	0.49897	0
	5	0.83088	1.676	0.64938	0.36118	0.87094	0.56863	0
	4	0.7515	1.4616	0.66943	0.4504	0.88232	0.6705	0
	3	0.63999	1.1725	0.69921	0.55557	0.89753	0.77265	0.01
	2	0.4881	0.65241	0.74759	0.68317	0.92186	0.87532	0
	1	0.31317	0.23037	0.81805	0.81805	0.94858	0.94858	0

Cenário: Z e DEuclidiana						
PCA						
Base de dados	k	Prop Var Ac	Strain	Stress-1 (E->E)	Sstress (E->E)	Tempo (s)
Triagem de Crédito	15	1	3E-015	1E-015	1E-015	0
	14	1	3E-015	1E-015	1E-015	0
	13	0.97399	0.086609	0.029129	0.069079	0
	12	0.93676	0.1512	0.057742	0.10382	0
	11	0.89765	0.19955	0.081884	0.1295	0
	10	0.85021	0.25451	0.1111	0.18411	0
	9	0.79967	0.3051	0.14108	0.22173	0
	8	0.7465	0.35274	0.16955	0.26702	0
	7	0.68711	0.40437	0.21026	0.30999	0
	6	0.62105	0.46033	0.24999	0.40897	0
	5	0.55322	0.51274	0.30451	0.4591	0
	4	0.47573	0.57398	0.37608	0.56698	0
	3	0.39198	0.63814	0.45525	0.71421	0
	2	0.28573	0.72963	0.54887	0.76862	0
	1	0.16898	0.82671	0.6831	0.86623	0
Imagens	18	1	2E-015	1E-015	1E-015	0
	17	1	2E-015	1E-015	1E-015	0
	16	1	2E-015	1E-015	1E-015	0
	15	1	2E-015	1E-015	1E-015	0
	14	1	2E-015	1E-015	1E-015	0
	13	0.99999	0.000029036	0.000033658	0.000021623	0
	12	0.99779	0.0045206	0.005685	0.0085447	0
	11	0.99238	0.012037	0.011394	0.010983	0
	10	0.97959	0.029011	0.021128	0.018543	0
	9	0.96154	0.047183	0.046138	0.04506	0
	8	0.93462	0.07286	0.070415	0.077833	0
	7	0.90434	0.095962	0.091576	0.09255	0
	6	0.87005	0.11921	0.11287	0.11166	0
	5	0.82771	0.14778	0.14045	0.1409	0
	4	0.77082	0.18869	0.18214	0.17991	0
	3	0.70546	0.23189	0.25326	0.23952	0
	2	0.60157	0.31573	0.3699	0.29137	0
	1	0.42472	0.4824	0.53107	0.83004	0
Iris	4	1	1E-015	1E-015	1E-015	0
	3	0.99398	0.0078782	0.011416	0.0082435	0
	2	0.95691	0.049148	0.064468	0.05137	0
	1	0.72816	0.30334	0.29071	0.32342	0
iris	60	1	2E-015	1E-015	1E-015	0.01
	59	0.99989	0.00035184	0.00010443	0.00017789	0.01
	58	0.9997	0.00069386	0.00024556	0.00039657	0.01
	57	0.99951	0.00092886	0.00035455	0.00057339	0.01
	56	0.99926	0.0012178	0.00050062	0.000808	0.01
	55	0.99899	0.0014928	0.00067192	0.0010717	0.01
	54	0.99867	0.0018128	0.00088094	0.0013765	0.01
	53	0.9983	0.0021586	0.0011267	0.0017455	0.01
	52	0.99792	0.0024825	0.001356	0.0021015	0.01
	51	0.99745	0.0029084	0.0016503	0.0025687	0.01
	50	0.99696	0.003299	0.0019344	0.0030041	0.01
	49	0.99644	0.0036957	0.0022423	0.0035222	0.01
	48	0.99584	0.0041541	0.0026013	0.0040601	0.01
	47	0.99514	0.0047296	0.0030737	0.0047749	0.01
	46	0.9942	0.0055997	0.0036291	0.0056878	0.01
	45	0.99325	0.0063683	0.0042045	0.006593	0.01
	44	0.99222	0.0071646	0.0048441	0.0075423	0.01
	43	0.99096	0.0082141	0.0056093	0.0087446	0.01
	42	0.9896	0.0093041	0.0064069	0.010003	0.01

Sonar	41	0.98801	0.010589	0.007432	0.011562	0.01
	40	0.98617	0.012121	0.0086754	0.013404	0.01
	39	0.98407	0.013856	0.0099666	0.015437	0.01
	38	0.98187	0.01553	0.011443	0.017689	0.01
	37	0.97943	0.017385	0.01292	0.019999	0.01
	36	0.97655	0.019666	0.014786	0.022845	0.01
	35	0.97352	0.021923	0.016631	0.025759	0.01
	34	0.97041	0.024063	0.018458	0.028781	0.01
	33	0.96705	0.026359	0.020582	0.032129	0.01
	32	0.96332	0.028921	0.022818	0.035681	0.01
	31	0.95896	0.032104	0.025369	0.039714	0.01
	30	0.9542	0.035528	0.028322	0.044106	0.01
	29	0.94924	0.038895	0.031369	0.048887	0.01
	28	0.944	0.042355	0.034471	0.053877	0.01
	27	0.93868	0.045634	0.037744	0.058875	0.01
	26	0.93253	0.04969	0.041462	0.064564	0.01
	25	0.92556	0.054445	0.045854	0.071097	0.01
	24	0.91834	0.05913	0.050236	0.078049	0.01
	23	0.91064	0.064055	0.054923	0.085796	0.01
	22	0.90224	0.06945	0.05997	0.093568	0.01
	21	0.89336	0.075017	0.065733	0.10237	0.01
	20	0.88392	0.080856	0.071288	0.111118	0.01
	19	0.87413	0.086696	0.077428	0.1206	0.01
	18	0.86382	0.092742	0.083797	0.13114	0.01
	17	0.85214	0.099977	0.091407	0.14236	0.01
	16	0.83979	0.10748	0.09939	0.15391	0.01
	15	0.82618	0.11594	0.10795	0.16702	0.01
	14	0.8119	0.1246	0.11644	0.18068	0.01
	13	0.79652	0.13394	0.12616	0.19632	0.01
	12	0.77905	0.14511	0.13629	0.21418	0.01
	11	0.76003	0.15732	0.14923	0.23149	0.01
	10	0.73928	0.17073	0.16375	0.25021	0.01
	9	0.71436	0.18837	0.18061	0.27533	0.01
	8	0.6887	0.20544	0.20005	0.30138	0.01
	7	0.65824	0.22734	0.21972	0.33489	0.01
	6	0.62545	0.25032	0.24523	0.36591	0.01
	5	0.5848	0.28201	0.27503	0.40507	0.01
	4	0.53473	0.32424	0.31808	0.45516	0.01
	3	0.47794	0.37157	0.36603	0.51631	0.01
	2	0.39244	0.4612	0.44339	0.59369	0.01
	1	0.20347	0.75981	0.65722	0.7814	0.01
Espiral	2	1	0	0	0	0
	1	0.53428	0.65709	0.44891	0.56575	0
Vogal	10	1	3E-015	1E-015	1E-015	0
	9	0.9915	0.021169	0.0087736	0.012963	0
	8	0.96194	0.076647	0.032973	0.049904	0
	7	0.92078	0.12802	0.061809	0.092916	0
	6	0.86921	0.18137	0.09703	0.14518	0
	5	0.80424	0.24311	0.14185	0.21068	0
	4	0.73109	0.30383	0.19632	0.27805	0
	3	0.63073	0.39348	0.27846	0.37552	0
	2	0.46294	0.5741	0.42089	0.54477	0
	1	0.25204	0.77826	0.63475	0.75568	0

Cenário: Z e DEuclidiana

SVD

Base de dados	k	Prop Var Ac	Strain	Stress-1 (E->E)	Sstress (E->E)	Tempo (s)	Prop Aprox Ac
	15	1	1E-015	0	1E-015	0.08	1
	14	1	1E-015	0	1E-015	0.08	1
	13	0.97399	0.086609	0.029129	0.069079	0.08	0.9556037

Triagem de Crédito	12	0.93676	0.1512	0.057742	0.10382	0.08	0.9024953
	11	0.89765	0.19955	0.081884	0.1295	0.08	0.8480561
	10	0.85021	0.25451	0.1111	0.18411	0.08	0.7880975
	9	0.79967	0.3051	0.14108	0.22173	0.08	0.7262158
	8	0.7465	0.35274	0.16955	0.26702	0.08	0.6627447
	7	0.68711	0.40437	0.21026	0.30999	0.08	0.5956634
	6	0.62105	0.46033	0.24999	0.40897	0.08	0.5249113
	5	0.55322	0.51274	0.30451	0.4591	0.08	0.4532222
	4	0.47573	0.57398	0.37608	0.56698	0.08	0.376598
	3	0.39198	0.63814	0.45525	0.71421	0.08	0.296934
	2	0.28573	0.72963	0.54887	0.76862	0.08	0.2072092
	1	0.16898	0.82671	0.6831	0.86623	0.08	0.1131558
Imagens	18	1	1E-015	0	1E-015	0.01	1
	17	1	1E-015	0	1E-015	0.01	1
	16	1	1E-015	0	1E-015	0.01	1
	15	1	1E-015	1E-015	1E-015	0.01	1
	14	1	1E-015	1E-015	1E-015	0.01	1
	13	0.99999	0.000029036	0.000033658	0.000021623	0.01	0.9987448
	12	0.99779	0.0045206	0.005685	0.0085447	0.01	0.9830834
	11	0.99238	0.012037	0.011394	0.010983	0.01	0.9584804
	10	0.97959	0.029011	0.021128	0.018543	0.01	0.9206353
	9	0.96154	0.047183	0.046138	0.04506	0.01	0.8757026
	8	0.93462	0.07286	0.070415	0.077833	0.01	0.8208168
	7	0.90434	0.095962	0.091576	0.09255	0.01	0.7626058
	6	0.87005	0.11921	0.11287	0.11166	0.01	0.700659
	5	0.82771	0.14778	0.14045	0.1409	0.01	0.6318194
	4	0.77082	0.18869	0.18214	0.17991	0.01	0.5520336
	3	0.70546	0.23189	0.25326	0.23952	0.01	0.4665096
	2	0.60157	0.31573	0.3699	0.29137	0.01	0.3586864
	1	0.42472	0.4824	0.53107	0.83004	0.01	0.2180097
Iris	4	1	0	0	0	0	1
	3	0.99398	0.0078782	0.011416	0.0082435	0	0.9515581
	2	0.95691	0.049148	0.064468	0.05137	0	0.8313504
	1	0.72816	0.30334	0.29071	0.32342	0	0.5327534
Spectra	60	1	1E-015	0	1E-015	0.01	1
	59	0.99989	0.00035184	0.00010443	0.00017789	0.01	0.9981251
	58	0.9997	0.00069386	0.00024556	0.00039657	0.01	0.9956808
	57	0.99951	0.00092886	0.00035455	0.00057339	0.01	0.9931969
	56	0.99926	0.0012178	0.00050062	0.000808	0.01	0.990392
	55	0.99899	0.0014928	0.00067192	0.0010717	0.01	0.9874548
	54	0.99867	0.0018128	0.00088094	0.0013765	0.01	0.9842492
	53	0.9983	0.0021586	0.0011267	0.0017455	0.01	0.9808277
	52	0.99792	0.0024825	0.001356	0.0021015	0.01	0.9773277
	51	0.99745	0.0029084	0.0016503	0.0025687	0.01	0.9734368
	50	0.99696	0.003299	0.0019344	0.0030041	0.01	0.9694928
	49	0.99644	0.0036957	0.0022423	0.0035222	0.01	0.9654133
	48	0.99584	0.0041541	0.0026013	0.0040601	0.01	0.96106
	47	0.99514	0.0047296	0.0030737	0.0047749	0.01	0.9563069
	46	0.9942	0.0055997	0.0036291	0.0056878	0.01	0.9508342
	45	0.99325	0.0063683	0.0042045	0.006593	0.01	0.9453296
	44	0.99222	0.0071646	0.0048441	0.0075423	0.01	0.9396028
	43	0.99096	0.0082141	0.0056093	0.0087446	0.01	0.9332672
	42	0.9896	0.0093041	0.0064069	0.010003	0.01	0.9266599
	41	0.98801	0.010589	0.007432	0.011562	0.01	0.9195525
	40	0.98617	0.012121	0.0086754	0.013404	0.01	0.9118767
	39	0.98407	0.013856	0.0099666	0.015437	0.01	0.9036866
	38	0.98187	0.01553	0.011443	0.017689	0.01	0.8953158
	37	0.97943	0.017385	0.01292	0.019999	0.01	0.8864796
	36	0.97655	0.019666	0.014786	0.022845	0.01	0.8768961
	35	0.97352	0.021923	0.016631	0.025759	0.01	0.8670583

Sonar	34	0.97041	0.024063	0.018458	0.028781	0.01	0.857102
	33	0.96705	0.026359	0.020582	0.032129	0.01	0.8467343
	32	0.96332	0.028921	0.022818	0.035681	0.01	0.8358301
	31	0.95896	0.032104	0.025369	0.039714	0.01	0.8240301
	30	0.9542	0.035528	0.028322	0.044106	0.01	0.8117003
	29	0.94924	0.038895	0.031369	0.048887	0.01	0.7991245
	28	0.944	0.042355	0.034471	0.053877	0.01	0.786182
	27	0.93868	0.045634	0.037744	0.058875	0.01	0.7731548
	26	0.93253	0.04969	0.041462	0.064564	0.01	0.7591387
	25	0.92556	0.054445	0.045854	0.071097	0.01	0.7442287
	24	0.91834	0.05913	0.050236	0.078049	0.01	0.7290475
	23	0.91064	0.064055	0.054923	0.085796	0.01	0.7133605
	22	0.90224	0.06945	0.05997	0.093568	0.01	0.6969861
	21	0.89336	0.075017	0.065733	0.10237	0.01	0.6801533
	20	0.88392	0.080856	0.071288	0.11118	0.01	0.6627928
	19	0.87413	0.086696	0.077428	0.1206	0.01	0.6451141
	18	0.86382	0.092742	0.083797	0.13114	0.01	0.626974
	17	0.85214	0.099977	0.091407	0.14236	0.01	0.6076588
	16	0.83979	0.10748	0.09939	0.15391	0.01	0.587806
	15	0.82618	0.11594	0.10795	0.16702	0.01	0.566962
	14	0.8119	0.1246	0.11644	0.18068	0.01	0.5456123
	13	0.79652	0.13394	0.12616	0.19632	0.01	0.5234505
	12	0.77905	0.14511	0.13629	0.21418	0.01	0.4998348
	11	0.76003	0.15732	0.14923	0.23149	0.01	0.4751935
	10	0.73928	0.17073	0.16375	0.25021	0.01	0.4494543
	9	0.71436	0.18837	0.18061	0.27533	0.01	0.4212539
	8	0.6887	0.20544	0.20005	0.30138	0.01	0.3926331
	7	0.65824	0.22734	0.21972	0.33489	0.01	0.3614473
	6	0.62545	0.25032	0.24523	0.36591	0.01	0.3290933
	5	0.5848	0.28201	0.27503	0.40507	0.01	0.2930696
	4	0.53473	0.32424	0.31808	0.45516	0.01	0.2530889
	3	0.47794	0.37157	0.36603	0.51631	0.01	0.2105093
	2	0.39244	0.4612	0.44339	0.59369	0.01	0.1582648
	1	0.20347	0.75981	0.65722	0.7814	0.01	0.0805942
Espiral	2	1	0	0	0	0	1
	1	0.53428	0.65709	0.44891	0.56575	0	0.5171579
Vogal	10	1	1E-015	0	1E-015	0.13	1
	9	0.9915	0.021169	0.0087736	0.012963	0.13	0.968293
	8	0.96194	0.076647	0.032973	0.049904	0.13	0.9091448
	7	0.92078	0.12802	0.061809	0.092916	0.13	0.8393624
	6	0.86921	0.18137	0.09703	0.14518	0.13	0.761248
	5	0.80424	0.24311	0.14185	0.21068	0.13	0.6735654
	4	0.73109	0.30383	0.19632	0.27805	0.13	0.5805341
	3	0.63073	0.39348	0.27846	0.37552	0.13	0.4715646
	2	0.46294	0.5741	0.42089	0.54477	0.13	0.3306606
	1	0.25204	0.77826	0.63475	0.75568	0.13	0.1726908

Cenário: Z e DEuclidiana

CMDS

Base de dados	k	Prop Var Ac	Strain	Stress-1 (E->E)	Sstress (E->E)	Tempo (s)
Triagem de Crédito	15	1	2E-015	0	1E-015	1.4
	14	1	2E-015	0	1E-015	1.4
	13	0.97399	0.086609	0.029129	0.069079	1.4
	12	0.93676	0.1512	0.057742	0.10382	1.4
	11	0.89765	0.19955	0.081884	0.1295	1.4
	10	0.85021	0.25451	0.1111	0.18411	1.4
	9	0.79967	0.3051	0.14108	0.22173	1.4
	8	0.7465	0.35274	0.16955	0.26702	1.4
	7	0.68711	0.40437	0.21026	0.30999	1.4
	6	0.62105	0.46033	0.24999	0.40897	1.4

5	0.55322	0.51274	0.30451	0.4591	1.4
4	0.47573	0.57398	0.37608	0.56698	1.4
3	0.39198	0.63814	0.45525	0.71421	1.4
2	0.28573	0.72963	0.54887	0.76862	1.4
1	0.16898	0.82671	0.6831	0.86623	1.4
Imagens	18	1	2E-015	1E-015	0.08
	17	1	2E-015	1E-015	0.08
	16	1	2E-015	1E-015	0.08
	15	1	2E-015	1E-015	0.08
	14	1	2E-015	1E-015	0.08
	13	0.99999	0.000029036	0.000033658	0.000021623
	12	0.99779	0.0045206	0.005685	0.0085447
	11	0.99238	0.012037	0.011394	0.010983
	10	0.97959	0.029011	0.021128	0.018543
	9	0.96154	0.047183	0.046138	0.04506
	8	0.93462	0.07286	0.070415	0.077833
	7	0.90434	0.095962	0.091576	0.09255
	6	0.87005	0.11921	0.11287	0.11166
	5	0.82771	0.14778	0.14045	0.1409
	4	0.77082	0.18869	0.18214	0.17991
	3	0.70546	0.23189	0.25326	0.23952
	2	0.60157	0.31573	0.3699	0.29137
	1	0.42472	0.4824	0.53107	0.83004
Iris	4	1	1E-015	1E-015	0.03
	3	0.99398	0.0078782	0.011416	0.0082435
	2	0.95691	0.049148	0.064468	0.05137
	1	0.72816	0.30334	0.29071	0.32342
Sonar	60	1	2E-015	0	1E-015
	59	0.99989	0.00035184	0.00010443	0.00017789
	58	0.9997	0.00069386	0.00024556	0.00039657
	57	0.99951	0.00092886	0.00035455	0.00057339
	56	0.99926	0.0012178	0.00050062	0.000808
	55	0.99899	0.0014928	0.00067192	0.0010717
	54	0.99867	0.0018128	0.00088094	0.0013765
	53	0.9983	0.0021586	0.0011267	0.0017455
	52	0.99792	0.0024825	0.001356	0.0021015
	51	0.99745	0.0029084	0.0016503	0.0025687
	50	0.99696	0.003299	0.0019344	0.0030041
	49	0.99644	0.0036957	0.0022423	0.0035222
	48	0.99584	0.0041541	0.0026013	0.0040601
	47	0.99514	0.0047296	0.0030737	0.0047749
	46	0.9942	0.0055997	0.0036291	0.0056878
	45	0.99325	0.0063683	0.0042045	0.006593
	44	0.99222	0.0071646	0.0048441	0.0075423
	43	0.99096	0.0082141	0.0056093	0.0087446
	42	0.9896	0.0093041	0.0064069	0.010003
	41	0.98801	0.010589	0.007432	0.011562
	40	0.98617	0.012121	0.0086754	0.013404
	39	0.98407	0.013856	0.0099666	0.015437
	38	0.98187	0.01553	0.011443	0.017689
	37	0.97943	0.017385	0.01292	0.019999
	36	0.97655	0.019666	0.014786	0.022845
	35	0.97352	0.021923	0.016631	0.025759
	34	0.97041	0.024063	0.018458	0.028781
	33	0.96705	0.026359	0.020582	0.032129
	32	0.96332	0.028921	0.022818	0.035681
	31	0.95896	0.032104	0.025369	0.039714
	30	0.9542	0.035528	0.028322	0.044106
	29	0.94924	0.038895	0.031369	0.048887
	28	0.944	0.042355	0.034471	0.053877

27	0.93868	0.045634	0.037744	0.058875	0.07
26	0.93253	0.04969	0.041462	0.064564	0.07
25	0.92556	0.054445	0.045854	0.071097	0.07
24	0.91834	0.05913	0.050236	0.078049	0.07
23	0.91064	0.064055	0.054923	0.085796	0.07
22	0.90224	0.06945	0.05997	0.093568	0.07
21	0.89336	0.075017	0.065733	0.10237	0.07
20	0.88392	0.080856	0.071288	0.11118	0.07
19	0.87413	0.086696	0.077428	0.1206	0.07
18	0.86382	0.092742	0.083797	0.13114	0.07
17	0.85214	0.099977	0.091407	0.14236	0.07
16	0.83979	0.10748	0.09939	0.15391	0.07
15	0.82618	0.11594	0.10795	0.16702	0.07
14	0.8119	0.1246	0.11644	0.18068	0.07
13	0.79652	0.13394	0.12616	0.19632	0.07
12	0.77905	0.14511	0.13629	0.21418	0.07
11	0.76003	0.15732	0.14923	0.23149	0.07
10	0.73928	0.17073	0.16375	0.25021	0.07
9	0.71436	0.18837	0.18061	0.27533	0.07
8	0.6887	0.20544	0.20005	0.30138	0.07
7	0.65824	0.22734	0.21972	0.33489	0.07
6	0.62545	0.25032	0.24523	0.36591	0.07
5	0.5848	0.28201	0.27503	0.40507	0.07
4	0.53473	0.32424	0.31808	0.45516	0.07
3	0.47794	0.37157	0.36603	0.51631	0.07
2	0.39244	0.4612	0.44339	0.59369	0.07
1	0.20347	0.75981	0.65722	0.7814	0.07

Espiral	2	1	1E-015	1E-015	1E-015	0.01
	1	0.53428	0.65709	0.44891	0.56575	0.01
Vogal	10	1	2E-015	0	1E-015	4.91
	9	0.9915	0.021169	0.0087736	0.012963	4.91
	8	0.96194	0.076647	0.032973	0.049904	4.91
	7	0.92078	0.12802	0.061809	0.092916	4.91
	6	0.86921	0.18137	0.09703	0.14518	4.91
	5	0.80424	0.24311	0.14185	0.21068	4.91
	4	0.73109	0.30383	0.19632	0.27805	4.91
	3	0.63073	0.39348	0.27846	0.37552	4.91
	2	0.46294	0.5741	0.42089	0.54477	4.91
	1	0.25204	0.77826	0.63475	0.75568	4.91

Cenário: Z e DEuclidiana

MDS Stress-1

Base de dados	k	Prop Var Ac	Strain	Stress-1 (E->E)	Sstress (E->E)	Tempo (s)
Triagem de Crédito	15	1	0.0010085	0.00004839	0.00020531	185.6
	14	1	0.000020196	9.9479E-007	4.1196E-006	153.48
	13	0.99987	0.14294	0.011338	0.02915	209.74
	12	0.99986	0.25259	0.01884	0.051474	254.72
	11	0.99921	0.29652	0.028294	0.060407	155.95
	10	0.99852	0.37532	0.038059	0.076641	235.74
	9	0.99759	0.46264	0.048966	0.094384	199.91
	8	0.9962	0.53983	0.061319	0.11044	130.43
	7	0.99408	0.64908	0.076794	0.13262	151.95
	6	0.99109	0.76681	0.094161	0.15729	87.52
	5	0.98597	0.92219	0.1184	0.18975	128.72
	4	0.97693	1.1228	0.15149	0.23222	159.46
	3	0.96122	1.3811	0.19729	0.28816	112.24
	2	0.91922	1.8488	0.28435	0.39224	127.53
	1	0.77167	2.7304	0.47784	0.59758	29.93
	18	0.99999	0.00047023	0.000066441	0.00012422	15.39
	17	0.99999	0.00054932	0.000068384	0.00014511	14.71

Imagens	16	1	0.00034908	0.000055388	0.000092383	14.09
	15	1	0.00048532	0.000058771	0.00012852	13.73
	14	1	0.00036848	0.000047158	0.0000975	16.17
	13	1	0.00011541	0.000019118	0.000030529	14.82
	12	1	0.0083991	0.00092656	0.0022183	20.81
	11	0.99998	0.039969	0.0041725	0.010548	23.62
	10	0.99992	0.072341	0.0078515	0.019118	24.14
	9	0.99983	0.096412	0.01268	0.025485	7.54
	8	0.99963	0.1318	0.01966	0.034833	15.01
	7	0.9992	0.18146	0.027912	0.048041	9.26
	6	0.99859	0.23386	0.037214	0.062297	15.27
	5	0.99723	0.3144	0.052856	0.083844	5.9
	4	0.99512	0.39427	0.070299	0.10558	7.74
	3	0.98955	0.60387	0.10186	0.16137	18.4
	2	0.97242	0.90369	0.166	0.24292	7.1
	1	0.80671	1.8706	0.43965	0.51771	3.45
Iris	4	1	2.2765E-006	0.000000993	1.3319E-006	1.86
	3	0.99991	0.011713	0.0092375	0.0071604	0.94
	2	0.99722	0.079852	0.052483	0.048661	0.48
	1	0.71418	1.3143	0.53462	0.8004	0.39
Sonar	60	1	0.001708	0.00025257	0.00046776	37.53
	59	0.99999	0.0023713	0.00034912	0.00064888	50.42
	58	0.99999	0.0023798	0.00035253	0.00065083	58.91
	57	1	0.0023345	0.00034968	0.00064062	60.72
	56	0.99998	0.0027527	0.00040552	0.00075216	53.04
	55	1	0.0025857	0.00038512	0.00070754	40.53
	54	0.99995	0.0031707	0.00047455	0.00086738	66.95
	53	1	0.0035207	0.00053354	0.00096173	65.69
	52	1	0.0040762	0.00062446	0.00111141	54.95
	51	1	0.003979	0.00059492	0.0010877	55.17
	50	1	0.0041557	0.00063448	0.0011346	73.5
	49	0.99999	0.0049436	0.000758	0.0013508	58.23
	48	0.99998	0.005365	0.00083173	0.0014647	52.63
	47	1	0.0058701	0.00090577	0.001603	72.59
	46	1	0.007177	0.0011109	0.0019599	54.43
	45	0.99997	0.0080915	0.0012602	0.0022088	48.55
	44	0.99999	0.0088264	0.0013771	0.0024097	69.96
	43	1	0.010277	0.0016026	0.002806	47.83
	42	0.99998	0.011714	0.0018242	0.0031978	52.84
	41	0.99997	0.0132	0.0020629	0.0036035	62.86
	40	1	0.01543	0.0024263	0.0042124	42.38
	39	0.99997	0.017322	0.0027303	0.0047285	59.25
	38	1	0.019752	0.0030901	0.0053936	53.06
	37	0.99997	0.022075	0.0034749	0.0060259	46.55
	36	0.99997	0.025473	0.0039615	0.0069543	66.04
	35	0.99997	0.028275	0.0043813	0.0077192	41.29
	34	0.99999	0.031744	0.0049406	0.008666	57.61
	33	0.99998	0.034726	0.0054265	0.00948	49.62
	32	0.9999	0.038487	0.0060249	0.010507	43.58
	31	0.99997	0.042712	0.006662	0.01166	56.53
	30	0.99994	0.047853	0.0074631	0.013064	40.28
	29	0.99998	0.053082	0.0083009	0.014492	56.65
	28	0.99988	0.058239	0.0091382	0.0159	34.78
	27	0.9999	0.063606	0.010015	0.017366	48.62
	26	0.99985	0.070169	0.011052	0.019158	40.18
	25	0.9998	0.078134	0.012203	0.021334	44.97
	24	0.9998	0.086241	0.013447	0.023549	43.93
	23	0.99979	0.094142	0.014726	0.025707	35.11
	22	0.99975	0.10393	0.016197	0.028379	49.2
	21	0.99968	0.11444	0.017887	0.031252	24.82

20	0.9997	0.12576	0.019597	0.034345	46.56
19	0.99958	0.1364	0.021402	0.037253	33.95
18	0.9996	0.14949	0.023484	0.040834	34.08
17	0.9993	0.16406	0.025831	0.044822	32.01
16	0.99922	0.1811	0.028398	0.049482	29.7
15	0.999	0.19964	0.031344	0.054558	17.26
14	0.99883	0.21929	0.034549	0.059941	26.65
13	0.99859	0.24243	0.038255	0.066277	33.35
12	0.99826	0.26833	0.042306	0.073374	18.71
11	0.99778	0.30146	0.047512	0.082455	18.49
10	0.99712	0.33798	0.053593	0.092477	20.58
9	0.99619	0.38265	0.061196	0.10477	9.17
8	0.99502	0.43833	0.070185	0.12011	6.53
7	0.99351	0.50003	0.081066	0.1372	8.77
6	0.99073	0.58192	0.095866	0.16002	5.5
5	0.98653	0.68862	0.11565	0.19012	9.13
4	0.97865	0.84829	0.14574	0.23637	5.27
3	0.96384	1.0501	0.18975	0.29957	2.45
2	0.92662	1.4086	0.27076	0.41649	2.14
1	0.76267	2.5839	0.48717	0.77488	1.85
Espiral	2	1	0.000002124	8.4325E-007	1.3097E-006
	1	0.82536	1.0037	0.4179	0.58133
Vogal	10	1	4.1159E-006	9.867E-007	1.4813E-006
	9	0.99997	0.024264	0.0056219	0.0085751
	8	0.9996	0.090369	0.020205	0.031958
	7	0.99872	0.16025	0.035782	0.056721
	6	0.99708	0.24329	0.05457	0.086455
	5	0.9939	0.3523	0.078475	0.12552
	4	0.98707	0.48387	0.11376	0.17402
	3	0.97112	0.69784	0.1699	0.25777
	2	0.92785	1.114	0.26838	0.42035
	1	0.73598	2.0147	0.51384	0.78398

Cenário: Z e DEuclidiana

MDS Sstress

Base de dados	k	Prop Var Ac	Strain	Stress-1 (E->E)	Sstress (E->E)	Tempo (s)
Triagem de Crédito	15	1	0.0032175	0.00052941	0.00066609	483.62
	14	1	0.00026858	0.00003952	0.000056376	493.73
	13	0.99805	0.09509	0.014085	0.019427	612.01
	12	0.99469	0.17175	0.02599	0.035205	582.09
	11	0.99023	0.23342	0.039215	0.048089	562.06
	10	0.98527	0.31037	0.048777	0.064103	539.57
	9	0.97929	0.38471	0.058986	0.079684	572.17
	8	0.97123	0.46063	0.072005	0.095967	591.51
	7	0.95921	0.55496	0.090649	0.11658	597.3
	6	0.94488	0.6655	0.11027	0.14097	601.25
	5	0.92333	0.78575	0.1399	0.16883	589.27
	4	0.89114	0.94967	0.17894	0.2079	640.63
	3	0.84363	1.1541	0.22328	0.2589	610.64
	2	0.75716	1.4149	0.30703	0.33339	592.87
	1	0.54325	1.8568	0.5015	0.49261	611.15
Imagens	18	1.0002	0.003569	0.0016567	0.00095602	56.25
	17	1.0001	0.0034799	0.0015437	0.00093295	39.46
	16	1.0002	0.0039781	0.0016663	0.0010683	54.8
	15	1.0001	0.0044881	0.001532	0.0012025	33.76
	14	1	0.0035808	0.0013363	0.00096508	50.36
	13	0.99993	0.0058098	0.0013101	0.0015403	35.08
	12	0.99989	0.01099	0.0038598	0.0029355	57.31
	11	0.99897	0.016448	0.0070675	0.0044298	34.36
	10	0.99757	0.033603	0.012741	0.009023	42.3

Wine						
9	0.99539	0.061635	0.020226	0.016606	55.1	
8	0.99228	0.095409	0.028655	0.025662	42.02	
7	0.98611	0.13009	0.043846	0.035391	56.69	
6	0.97911	0.16727	0.055172	0.045933	51.21	
5	0.96836	0.21439	0.074107	0.05965	24.37	
4	0.95043	0.29195	0.097557	0.082595	32.17	
3	0.91854	0.3987	0.14651	0.11603	33.11	
2	0.84777	0.6044	0.2209	0.18351	52.28	
1	0.69113	1.1396	0.34776	0.35264	14.46	
Iris						
4	0.99996	0.0017586	0.001615	0.0010565	27.15	
3	0.99825	0.0096576	0.0098111	0.0059808	9.33	
2	0.98318	0.060581	0.056203	0.039101	3.21	
1	0.8529	0.37906	0.25608	0.2625	2.14	
Sonar						
60	0.99998	0.004312	0.00076031	0.0011813	67.94	
59	0.99993	0.0051665	0.00089989	0.0014163	82.18	
58	0.99995	0.0041866	0.00072941	0.0011472	68.55	
57	0.99993	0.0046571	0.00081308	0.0012765	69.32	
56	0.99994	0.0046792	0.00082373	0.0012823	71.5	
55	0.99995	0.0050071	0.00086988	0.0013713	74.34	
54	0.99995	0.0052492	0.00091828	0.0014368	73.48	
53	0.99994	0.0053589	0.00092932	0.0014685	71.25	
52	0.99993	0.0057932	0.0010179	0.0015864	73.85	
51	0.9999	0.0060172	0.0010496	0.0016493	70.17	
50	0.99988	0.0065607	0.0011539	0.0017969	70.15	
49	0.99989	0.0064105	0.0011175	0.0017561	70.85	
48	0.99992	0.0076128	0.0013371	0.002083	70.74	
47	0.99987	0.0074052	0.0012975	0.0020288	69.89	
46	0.99986	0.0086241	0.001512	0.002361	68.39	
45	0.99974	0.010241	0.0017916	0.0028076	70.2	
44	0.99982	0.0090479	0.0015971	0.0024765	69.82	
43	0.99974	0.012466	0.0021785	0.0034149	70.74	
42	0.99975	0.011715	0.0020653	0.0032079	70.94	
41	0.99969	0.013713	0.0024282	0.0037552	70.77	
40	0.99963	0.015707	0.0027733	0.0043024	70.18	
39	0.99954	0.018323	0.0032306	0.0050204	70.36	
38	0.99948	0.019909	0.0034761	0.0054549	69.54	
37	0.99938	0.022456	0.0039784	0.0061564	69.29	
36	0.99932	0.024484	0.0043423	0.0067115	71.1	
35	0.99921	0.027347	0.0048284	0.0074999	69.88	
34	0.99912	0.030003	0.0053036	0.0082293	70.48	
33	0.99899	0.032871	0.0058222	0.0090201	69.58	
32	0.99881	0.036939	0.0065236	0.01014	67.66	
31	0.99865	0.040475	0.0071728	0.011115	58.72	
30	0.99844	0.045177	0.0080674	0.012411	64.74	
29	0.99821	0.049988	0.0089289	0.01374	76.05	
28	0.99798	0.054743	0.0097523	0.015052	73.04	
27	0.99771	0.059763	0.01069	0.016441	50.58	
26	0.99738	0.065787	0.011816	0.01811	71.57	
25	0.99694	0.074081	0.013188	0.020406	76.18	
24	0.99656	0.080201	0.014446	0.022108	63.92	
23	0.9961	0.088496	0.015818	0.024407	56.86	
22	0.99556	0.096727	0.017432	0.0267	72.81	
21	0.99487	0.10666	0.019291	0.029475	66.64	
20	0.99426	0.11692	0.0212	0.032327	74.18	
19	0.99345	0.1268	0.02304	0.035103	73.15	
18	0.99257	0.1392	0.025287	0.038573	77.33	
17	0.99145	0.15216	0.027858	0.042237	74.61	
16	0.99013	0.16866	0.030891	0.046879	72.1	
15	0.98866	0.18351	0.033939	0.051115	68.8	
14	0.98706	0.20297	0.037608	0.056604	70.05	

13	0.98499	0.22154	0.041445	0.061949	64.58
12	0.98264	0.24499	0.045899	0.068655	65.56
11	0.97944	0.2719	0.051877	0.076469	64.33
10	0.9756	0.30314	0.058726	0.085588	48.02
9	0.97075	0.34266	0.067	0.097135	60.91
8	0.96462	0.38924	0.076761	0.11091	57.88
7	0.95662	0.44234	0.088404	0.12685	68.4
6	0.9453	0.50936	0.10446	0.14749	31.34
5	0.92927	0.59277	0.1265	0.17394	48.26
4	0.90342	0.70961	0.15801	0.2125	50.03
3	0.8642	0.86268	0.20671	0.26606	29.86
2	0.78747	1.0717	0.2953	0.35187	13.69
1	0.55387	1.7549	0.50442	0.62389	37.06
Espiral	2	1	1.2607E-006	5.3781E-007	1.54
	1	0.73116	0.85908	0.41623	0.52322
Vogal	10	1	0.00009111	0.000022014	0.000032644
	9	0.99908	0.023214	0.0058057	0.008277
	8	0.99549	0.085936	0.020901	0.030749
	7	0.98942	0.15053	0.037157	0.054253
	6	0.97998	0.22526	0.056908	0.082013
	5	0.96528	0.3191	0.082486	0.11758
	4	0.94136	0.42611	0.12017	0.16074
	3	0.89458	0.59628	0.18007	0.23452
	2	0.79719	0.90245	0.28511	0.3719
	1	0.54281	1.5108	0.54514	0.67945

Cenário: Z e DEuclidiana

FastMap

Base de dados	k	Prop Var Ac	Strain	Stress-1 (E->E)	Sstress (E->E)	Tempo (s)
Triagem de Crédito	15	0.99999	74.717	7.9592E-006	0.000020655	0.03
	14	0.99999	74.717	7.9592E-006	0.000020655	0
	13	0.94905	73.545	0.049356	0.058768	0
	12	0.8869	71.186	0.094984	0.1079	0.01
	11	0.82026	69.44	0.13897	0.15202	0
	10	0.76748	68.527	0.1736	0.1869	0
	9	0.7151	66.384	0.20721	0.22362	0
	8	0.64591	62.024	0.25689	0.27277	0
	7	0.57534	57.926	0.3112	0.3237	0
	6	0.50294	53.149	0.36851	0.38484	0
	5	0.402	47.898	0.45022	0.45909	0
	4	0.30582	40.358	0.54051	0.53068	0
	3	0.2363	32.612	0.61131	0.59576	0.01
	2	0.16442	27.539	0.69477	0.68279	0.01
	1	0.092402	12.745	0.79741	0.76163	0
Imagens	18	1	10.935	5.8044E-006	0.000017682	0.03
	17	1	10.935	5.8044E-006	0.000017682	0
	16	1	10.935	5.8044E-006	0.000017682	0
	15	1	10.935	5.8044E-006	0.000017682	0
	14	1	10.935	5.8044E-006	0.000017682	0.01
	13	0.99997	10.935	0.000039454	0.000034304	0
	12	0.99331	10.869	0.010403	0.0061646	0.01
	11	0.98735	10.822	0.014098	0.011479	0
	10	0.96603	10.661	0.033508	0.026357	0
	9	0.92813	9.8844	0.061921	0.052503	0
	8	0.88573	9.5059	0.087639	0.079744	0
	7	0.82517	8.5829	0.13796	0.11766	0
	6	0.73326	7.8699	0.22085	0.17734	0
	5	0.69362	6.7156	0.25043	0.20169	0
	4	0.60382	5.09	0.31372	0.26063	0
	3	0.36324	3.8062	0.50199	0.42912	0

	2	0.31842	2.5785	0.55205	0.46274	<i>0.01</i>
	1	0.18953	1.1179	0.73553	0.56532	<i>0</i>
Iris	4	1	6.7013	0.000010708	0.000017709	<i>0.01</i>
	3	0.98755	6.5942	0.020072	0.017366	<i>0.01</i>
	2	0.94509	6.3036	0.067281	0.063338	<i>0</i>
	1	0.67739	3.7666	0.33329	0.3787	<i>0</i>
Sonar	60	1	61.566	3.1412E-006	5.8756E-006	<i>0.05</i>
	59	0.99956	61.553	0.00040277	0.00066973	<i>0.05</i>
	58	0.99899	61.544	0.00078928	0.0012228	<i>0.03</i>
	57	0.99841	61.524	0.0011606	0.0018157	<i>0.01</i>
	56	0.9976	61.499	0.0016731	0.0026354	<i>0</i>
	55	0.9968	61.456	0.0021556	0.0033822	<i>0.01</i>
	54	0.99578	61.407	0.0027948	0.0043715	<i>0</i>
	53	0.99472	61.37	0.0034881	0.0054247	<i>0</i>
	52	0.99359	61.314	0.0041938	0.006508	<i>0</i>
	51	0.99245	61.255	0.0048662	0.0075781	<i>0</i>
	50	0.99067	61.189	0.0060227	0.0093362	<i>0</i>
	49	0.9893	61.069	0.0068629	0.010693	<i>0</i>
	48	0.98789	60.985	0.0076361	0.011944	<i>0</i>
	47	0.98513	60.892	0.0094697	0.014653	<i>0</i>
	46	0.98266	60.804	0.01093	0.01696	<i>0</i>
	45	0.98041	60.713	0.012188	0.018966	<i>0</i>
	44	0.97815	60.596	0.013495	0.020985	<i>0</i>
	43	0.97518	60.501	0.015062	0.02364	<i>0.01</i>
	42	0.97179	60.269	0.017302	0.026859	<i>0</i>
	41	0.96829	60.139	0.019235	0.03011	<i>0</i>
	40	0.96454	59.898	0.021577	0.033722	<i>0</i>
	39	0.96173	59.723	0.023368	0.036357	<i>0</i>
	38	0.95538	59.503	0.027482	0.042709	<i>0</i>
	37	0.95128	59.268	0.029883	0.046429	<i>0</i>
	36	0.94511	59.002	0.033719	0.052041	<i>0</i>
	35	0.94023	58.755	0.036874	0.056578	<i>0</i>
	34	0.93567	58.318	0.039908	0.060903	<i>0</i>
	33	0.93196	57.98	0.042351	0.064401	<i>0</i>
	32	0.92702	57.572	0.04555	0.069004	<i>0</i>
	31	0.9185	57.153	0.050792	0.077028	<i>0</i>
	30	0.91029	56.553	0.055465	0.084668	<i>0</i>
	29	0.89871	55.622	0.063038	0.096362	<i>0</i>
	28	0.89035	54.957	0.068891	0.1044	<i>0</i>
	27	0.88063	54.473	0.074585	0.1132	<i>0</i>
	26	0.87293	53.976	0.079256	0.12022	<i>0</i>
	25	0.86147	53.068	0.086417	0.13089	<i>0</i>
	24	0.85125	52.436	0.092827	0.14009	<i>0</i>
	23	0.83595	51.552	0.10311	0.15477	<i>0</i>
	22	0.82648	50.696	0.10965	0.16364	<i>0</i>
	21	0.81353	49.401	0.11863	0.17585	<i>0</i>
	20	0.79828	48.71	0.12963	0.19073	<i>0</i>
	19	0.78238	47.826	0.13846	0.20505	<i>0</i>
	18	0.76675	47	0.14949	0.21932	<i>0</i>
	17	0.75003	45.858	0.16019	0.2345	<i>0</i>
	16	0.73761	44.483	0.1678	0.24623	<i>0</i>
	15	0.72018	43.468	0.18058	0.26232	<i>0</i>
	14	0.7034	41.989	0.19137	0.27826	<i>0</i>
	13	0.67917	40.288	0.20771	0.30075	<i>0</i>
	12	0.64037	38.85	0.23729	0.33629	<i>0</i>
	11	0.61215	37.101	0.2565	0.36286	<i>0</i>
	10	0.59166	34.742	0.27253	0.38234	<i>0</i>
	9	0.54822	33.02	0.30377	0.42364	<i>0</i>
	8	0.51257	29.642	0.33105	0.45863	<i>0</i>
	7	0.4631	28.402	0.36689	0.50588	<i>0</i>

	6	0.43613	25.614	0.39054	0.53439	<i>0</i>
	5	0.41263	22.822	0.41536	0.55633	<i>0</i>
	4	0.3582	18.463	0.45675	0.61106	<i>0</i>
	3	0.30582	16.712	0.50693	0.66796	<i>0</i>
	2	0.21759	5.7704	0.60668	0.76883	<i>0.01</i>
	1	0.12131	3.638	0.73944	0.86572	<i>0</i>
Espiral	2	1	7.3695	0.000015263	0.000026956	<i>0</i>
	1	0.5	4.2416	0.47874	0.61217	<i>0</i>
Vogal	10	1	24.282	6.6653E-006	0.000012401	<i>0.01</i>
	9	0.98759	23.994	0.012103	0.018735	<i>0.01</i>
	8	0.94621	23.239	0.04593	0.072768	<i>0</i>
	7	0.89146	21.969	0.084053	0.12801	<i>0</i>
	6	0.84032	20.882	0.1181	0.17937	<i>0</i>
	5	0.76175	19.292	0.17587	0.25388	<i>0.01</i>
	4	0.64718	16.335	0.25453	0.3687	<i>0</i>
	3	0.55614	13.347	0.32447	0.45567	<i>0</i>
	2	0.35047	10.192	0.49619	0.67002	<i>0</i>
	1	0.18834	6.5693	0.68586	0.82366	<i>0</i>

Cenário: Z e D Manhattan								
PCA								
Base de dados	k	Prop Var Ac	Strain	Stress-1 (M->E)	Stress-1 (M->M)	Sstress (M->E)	Sstress (M->M)	Tempo (s)
Triagem de Crédito	15	1	3E-015	0.60189	0.2717	0.83494	0.61315	0
	14	1	3E-015	0.60189	0.2717	0.83494	0.61315	0
	13	0.97399	0.086609	0.60662	0.23603	0.83798	0.5292	0
	12	0.93676	0.1512	0.61412	0.20426	0.84203	0.46562	0
	11	0.89765	0.19955	0.62249	0.18898	0.84598	0.41818	0
	10	0.85021	0.25451	0.63234	0.1882	0.85269	0.36344	0
	9	0.79967	0.3051	0.64349	0.21647	0.86044	0.35921	0
	8	0.7465	0.35274	0.65501	0.25686	0.86813	0.4102	0
	7	0.68711	0.40437	0.66997	0.32274	0.87485	0.48296	0
	6	0.62105	0.46033	0.68574	0.38839	0.89379	0.58643	0
	5	0.55322	0.51274	0.70538	0.46796	0.90187	0.6641	0
	4	0.47573	0.57398	0.72972	0.55069	0.91875	0.76975	0
	3	0.39198	0.63814	0.75891	0.64119	0.94203	0.87032	0
	2	0.28573	0.72963	0.79971	0.74439	0.95432	0.92242	0
	1	0.16898	0.82671	0.85732	0.85732	0.97239	0.97239	0
Imagens	18	1	2E-015	0.66658	0.3022	0.87796	0.52313	0
	17	1	2E-015	0.66658	0.3022	0.87796	0.52313	0
	16	1	2E-015	0.66658	0.3022	0.87796	0.52313	0
	15	1	2E-015	0.66658	0.3022	0.87796	0.52313	0
	14	1	2E-015	0.66658	0.3022	0.87796	0.52313	0
	13	0.99999	2.90E-005	0.66658	0.30269	0.87796	0.52361	0
	12	0.99779	0.004521	0.66693	0.30657	0.87817	0.53294	0
	11	0.99238	0.012037	0.66783	0.31834	0.87838	0.54412	0
	10	0.97959	0.029011	0.67002	0.3399	0.87901	0.56336	0
	9	0.96154	0.047183	0.67303	0.35538	0.87982	0.57757	0
	8	0.93462	0.07286	0.67745	0.38241	0.88223	0.61916	0
	7	0.90434	0.095962	0.68295	0.41995	0.88377	0.65179	0
	6	0.87005	0.11921	0.68916	0.46006	0.88538	0.68039	0
	5	0.82771	0.14778	0.69687	0.50771	0.88852	0.72937	0
	4	0.77082	0.18869	0.70874	0.56914	0.89236	0.77839	0
	3	0.70546	0.23189	0.72417	0.63248	0.89738	0.82767	0
	2	0.60157	0.31573	0.75404	0.71532	0.90239	0.86437	0
	1	0.42472	0.4824	0.80277	0.80277	0.96485	0.96485	0
Iris	4	1	1E-015	0.45931	0.26917	0.71472	0.47945	0
	3	0.99398	0.007878	0.46085	0.29101	0.71562	0.50836	0
	2	0.95691	0.049148	0.47165	0.35561	0.72136	0.58324	0
	1	0.72816	0.30334	0.55799	0.55799	0.773	0.773	0
iris	60	1	2E-015	0.83778	0.29353	0.97391	0.5295	0.01
	59	0.99989	0.000352	0.83779	0.2947	0.97391	0.53104	0.01
	58	0.9997	0.000694	0.8378	0.29623	0.97391	0.53302	0.01
	57	0.99951	0.000929	0.83782	0.29778	0.97392	0.53499	0.01
	56	0.99926	0.001218	0.83784	0.2995	0.97392	0.53718	0.01
	55	0.99899	0.001493	0.83786	0.30129	0.97393	0.53939	0.01
	54	0.99867	0.001813	0.83788	0.3032	0.97393	0.54177	0.01
	53	0.9983	0.002159	0.83791	0.30524	0.97394	0.54428	0.01
	52	0.99792	0.002483	0.83794	0.30744	0.97395	0.54704	0.01
	51	0.99745	0.002908	0.83798	0.30984	0.97396	0.55011	0.01
	50	0.99696	0.003299	0.83802	0.31233	0.97397	0.55317	0.01
	49	0.99644	0.003696	0.83807	0.31492	0.97398	0.55638	0.01
	48	0.99584	0.004154	0.83811	0.31765	0.97399	0.55968	0.01
	47	0.99514	0.00473	0.83817	0.32053	0.97401	0.56322	0.01
	46	0.9942	0.0056	0.83825	0.32407	0.97403	0.56763	0.01
	45	0.99325	0.006368	0.83832	0.32764	0.97405	0.57199	0.01
	44	0.99222	0.007165	0.83841	0.33126	0.97407	0.57626	0.01
	43	0.99096	0.008214	0.83851	0.33524	0.97409	0.58106	0.01
	42	0.9896	0.009304	0.83862	0.33959	0.97412	0.58623	0.01

Sonar	41	0.98801	0.010589	0.83875	0.3441	0.97415	0.59156	0.01
	40	0.98617	0.012121	0.8389	0.34901	0.97419	0.59727	0.01
	39	0.98407	0.013856	0.83908	0.35431	0.97424	0.60344	0.01
	38	0.98187	0.01553	0.83926	0.3597	0.97428	0.60976	0.01
	37	0.97943	0.017385	0.83946	0.36547	0.97433	0.61629	0.01
	36	0.97655	0.019666	0.8397	0.37172	0.97439	0.62342	0.01
	35	0.97352	0.021923	0.83994	0.37815	0.97446	0.63066	0.01
	34	0.97041	0.024063	0.8402	0.38483	0.97452	0.63828	0.01
	33	0.96705	0.026359	0.84048	0.39165	0.9746	0.64591	0.01
	32	0.96332	0.028921	0.84079	0.39895	0.97468	0.6539	0.01
	31	0.95896	0.032104	0.84116	0.40678	0.97477	0.66243	0.01
	30	0.9542	0.035528	0.84155	0.41494	0.97487	0.67108	0.01
	29	0.94924	0.038895	0.84197	0.42331	0.97497	0.67978	0.01
	28	0.944	0.042355	0.84241	0.432	0.97509	0.6887	0.01
	27	0.93868	0.045634	0.84286	0.44068	0.9752	0.69743	0.01
	26	0.93253	0.04969	0.84339	0.4503	0.97533	0.70697	0.01
	25	0.92556	0.054445	0.84398	0.46047	0.97547	0.71686	0.01
	24	0.91834	0.05913	0.8446	0.47112	0.97564	0.72742	0.01
	23	0.91064	0.064055	0.84527	0.48173	0.97581	0.73765	0.01
	22	0.90224	0.06945	0.846	0.49314	0.976	0.74828	0.01
	21	0.89336	0.075017	0.84678	0.50464	0.97619	0.75863	0.01
	20	0.88392	0.080856	0.84761	0.51678	0.9764	0.7696	0.01
	19	0.87413	0.086696	0.84848	0.52919	0.97662	0.7806	0.01
	18	0.86382	0.092742	0.84941	0.54157	0.97687	0.79149	0.01
	17	0.85214	0.099977	0.85047	0.55513	0.97713	0.80268	0.01
	16	0.83979	0.10748	0.85159	0.56903	0.9774	0.81377	0.01
	15	0.82618	0.11594	0.85284	0.58381	0.97772	0.82565	0.01
	14	0.8119	0.1246	0.85416	0.59914	0.97807	0.83777	0.01
	13	0.79652	0.13394	0.8556	0.61441	0.97844	0.84934	0.01
	12	0.77905	0.14511	0.85721	0.63094	0.97892	0.86237	0.01
	11	0.76003	0.15732	0.85906	0.64846	0.97933	0.87386	0.01
	10	0.73928	0.17073	0.86111	0.66618	0.97976	0.88468	0.01
	9	0.71436	0.18837	0.86358	0.68642	0.98037	0.89772	0.01
	8	0.6887	0.20544	0.86629	0.70667	0.98097	0.90957	0.01
	7	0.65824	0.22734	0.86938	0.72922	0.98182	0.9231	0.01
	6	0.62545	0.25032	0.87304	0.7527	0.98256	0.93432	0.01
	5	0.5848	0.28201	0.87758	0.77906	0.98356	0.94676	0.01
	4	0.53473	0.32424	0.88369	0.80784	0.9847	0.95792	0.01
	3	0.47794	0.37157	0.89097	0.83944	0.98626	0.97008	0.01
	2	0.39244	0.4612	0.90314	0.878	0.98837	0.9814	0.01
	1	0.20347	0.75981	0.9381	0.9381	0.99354	0.99354	0.01
Espiral	2	1	0	0.23407	0.19072	0.42319	0.36817	0
	1	0.53428	0.65709	0.54258	0.54258	0.70114	0.70114	0
Vogal	10	1	3E-015	0.62451	0.16268	0.8621	0.32432	0
	9	0.9915	0.021169	0.6261	0.17783	0.86297	0.34758	0
	8	0.96194	0.076647	0.63183	0.2144	0.86621	0.40083	0
	7	0.92078	0.12802	0.64013	0.26625	0.87091	0.4715	0
	6	0.86921	0.18137	0.65095	0.32692	0.87678	0.54702	0
	5	0.80424	0.24311	0.66528	0.39949	0.88456	0.63159	0
	4	0.73109	0.30383	0.68281	0.47783	0.89263	0.70811	0
	3	0.63073	0.39348	0.70974	0.57236	0.90418	0.78753	0
	2	0.46294	0.5741	0.76165	0.70072	0.92748	0.88285	0
	1	0.25204	0.77826	0.845	0.845	0.95942	0.95942	0

Cenário: Z e D Manhattan

SVD

Base de dados	k	Prop Var Ac	Strain	Stress-1 (M->E)	Stress-1 (M->M)	Sstress (M->E)	Sstress (M->M)	Tempo (s)	Prop Aprox Ac
	15	1	1E-015	0.60189	0.2717	0.83494	0.61315	0.08	1
	14	1	1E-015	0.60189	0.2717	0.83494	0.61315	0.08	1

Triagem de Crédito	13	0.97399	0.086609	0.60662	0.23603	0.83798	0.5292	0.08	0.9556
	12	0.93676	0.1512	0.61412	0.20426	0.84203	0.46562	0.08	0.9025
	11	0.89765	0.19955	0.62249	0.18898	0.84598	0.41818	0.08	0.84806
	10	0.85021	0.25451	0.63234	0.1882	0.85269	0.36344	0.08	0.7881
	9	0.79967	0.3051	0.64349	0.21647	0.86044	0.35921	0.08	0.72622
	8	0.7465	0.35274	0.65501	0.25686	0.86813	0.4102	0.08	0.66274
	7	0.68711	0.40437	0.66997	0.32274	0.87485	0.48296	0.08	0.59566
	6	0.62105	0.46033	0.68574	0.38839	0.89379	0.58643	0.08	0.52491
	5	0.55322	0.51274	0.70538	0.46796	0.90187	0.6641	0.08	0.45322
	4	0.47573	0.57398	0.72972	0.55069	0.91875	0.76975	0.08	0.3766
	3	0.39198	0.63814	0.75891	0.64119	0.94203	0.87032	0.08	0.29693
	2	0.28573	0.72963	0.79971	0.74439	0.95432	0.92242	0.08	0.20721
	1	0.16898	0.82671	0.85732	0.85732	0.97239	0.97239	0.08	0.11316
Imagens	18	1	1E-015	0.66658	0.3022	0.87796	0.52313	0.01	1
	17	1	1E-015	0.66658	0.3022	0.87796	0.52313	0.01	1
	16	1	1E-015	0.66658	0.3022	0.87796	0.52313	0.01	1
	15	1	1E-015	0.66658	0.3022	0.87796	0.52313	0.01	1
	14	1	1E-015	0.66658	0.3022	0.87796	0.52313	0.01	1
	13	0.99999	2.90E-005	0.66658	0.30269	0.87796	0.52361	0.01	0.99874
	12	0.99779	0.004521	0.66693	0.30657	0.87817	0.53294	0.01	0.98308
	11	0.99238	0.012037	0.66783	0.31834	0.87838	0.54412	0.01	0.95848
	10	0.97959	0.029011	0.67002	0.3399	0.87901	0.56336	0.01	0.92064
	9	0.96154	0.047183	0.67303	0.35538	0.87982	0.57757	0.01	0.8757
	8	0.93462	0.07286	0.67745	0.38241	0.88223	0.61916	0.01	0.82082
	7	0.90434	0.095962	0.68295	0.41995	0.88377	0.65179	0.01	0.76261
	6	0.87005	0.11921	0.68916	0.46006	0.88538	0.68039	0.01	0.70066
	5	0.82771	0.14778	0.69687	0.50771	0.88852	0.72937	0.01	0.63182
	4	0.77082	0.18869	0.70874	0.56914	0.89236	0.77839	0.01	0.55203
	3	0.70546	0.23189	0.72417	0.63248	0.89738	0.82767	0.01	0.46651
	2	0.60157	0.31573	0.75404	0.71532	0.90239	0.86437	0.01	0.35869
	1	0.42472	0.4824	0.80277	0.80277	0.96485	0.96485	0.01	0.21801
Iris	4	1	0	0.45931	0.26917	0.71472	0.47945	0	1
	3	0.99398	0.007878	0.46085	0.29101	0.71562	0.50836	0	0.95156
	2	0.95691	0.049148	0.47165	0.35561	0.72136	0.58324	0	0.83135
	1	0.72816	0.30334	0.55799	0.55799	0.773	0.773	0	0.53275
Santos	60	1	1E-015	0.83778	0.29353	0.97391	0.5295	0.01	1
	59	0.99989	0.000352	0.83779	0.2947	0.97391	0.53104	0.01	0.99813
	58	0.9997	0.000694	0.8378	0.29623	0.97391	0.53302	0.01	0.99568
	57	0.99951	0.000929	0.83782	0.29778	0.97392	0.53499	0.01	0.9932
	56	0.99926	0.001218	0.83784	0.2995	0.97392	0.53718	0.01	0.99039
	55	0.99899	0.001493	0.83786	0.30129	0.97393	0.53939	0.01	0.98745
	54	0.99867	0.001813	0.83788	0.3032	0.97393	0.54177	0.01	0.98425
	53	0.9983	0.002159	0.83791	0.30524	0.97394	0.54428	0.01	0.98083
	52	0.99792	0.002483	0.83794	0.30744	0.97395	0.54704	0.01	0.97733
	51	0.99745	0.002908	0.83798	0.30984	0.97396	0.55011	0.01	0.97344
	50	0.99696	0.003299	0.83802	0.31233	0.97397	0.55317	0.01	0.96949
	49	0.99644	0.003696	0.83807	0.31492	0.97398	0.55638	0.01	0.96541
	48	0.99584	0.004154	0.83811	0.31765	0.97399	0.55968	0.01	0.96106
	47	0.99514	0.00473	0.83817	0.32053	0.97401	0.56322	0.01	0.95631
	46	0.9942	0.0056	0.83825	0.32407	0.97403	0.56763	0.01	0.95083
	45	0.99325	0.006368	0.83832	0.32764	0.97405	0.57199	0.01	0.94533
	44	0.99222	0.007165	0.83841	0.33126	0.97407	0.57626	0.01	0.9396
	43	0.99096	0.008214	0.83851	0.33524	0.97409	0.58106	0.01	0.93327
	42	0.9896	0.009304	0.83862	0.33959	0.97412	0.58623	0.01	0.92666
	41	0.98801	0.010589	0.83875	0.3441	0.97415	0.59156	0.01	0.91955
	40	0.98617	0.012121	0.8389	0.34901	0.97419	0.59727	0.01	0.91188
	39	0.98407	0.013856	0.83908	0.35431	0.97424	0.60344	0.01	0.90369
	38	0.98187	0.01553	0.83926	0.3597	0.97428	0.60976	0.01	0.89532
	37	0.97943	0.017385	0.83946	0.36547	0.97433	0.61629	0.01	0.88648
	36	0.97655	0.019666	0.8397	0.37172	0.97439	0.62342	0.01	0.8769

Cenário: Z e D Manhattan									
CMDS									
Base de dados	k	Prop Var Ac	Strain	Stress-1 (M->E)	Stress-1 (M->M)	Sstress (M->E)	Sstress (M->M)	Tempo (s)	
Sonar	35	0.97352	0.021923	0.83994	0.37815	0.97446	0.63066	0.01	0.86706
	34	0.97041	0.024063	0.8402	0.38483	0.97452	0.63828	0.01	0.8571
	33	0.96705	0.026359	0.84048	0.39165	0.9746	0.64591	0.01	0.84673
	32	0.96332	0.028921	0.84079	0.39895	0.97468	0.6539	0.01	0.83583
	31	0.95896	0.032104	0.84116	0.40678	0.97477	0.66243	0.01	0.82403
	30	0.9542	0.035528	0.84155	0.41494	0.97487	0.67108	0.01	0.8117
	29	0.94924	0.038895	0.84197	0.42331	0.97497	0.67978	0.01	0.79912
	28	0.944	0.042355	0.84241	0.432	0.97509	0.6887	0.01	0.78618
	27	0.93868	0.045634	0.84286	0.44068	0.9752	0.69743	0.01	0.77315
	26	0.93253	0.04969	0.84339	0.4503	0.97533	0.70697	0.01	0.75914
	25	0.92556	0.054445	0.84398	0.46047	0.97547	0.71686	0.01	0.74423
	24	0.91834	0.05913	0.8446	0.47112	0.97564	0.72742	0.01	0.72905
	23	0.91064	0.064055	0.84527	0.48173	0.97581	0.73765	0.01	0.71336
	22	0.90224	0.06945	0.846	0.49314	0.976	0.74828	0.01	0.69699
	21	0.89336	0.075017	0.84678	0.50464	0.97619	0.75863	0.01	0.68015
	20	0.88392	0.080856	0.84761	0.51678	0.9764	0.7696	0.01	0.66279
	19	0.87413	0.086696	0.84848	0.52919	0.97662	0.7806	0.01	0.64511
	18	0.86382	0.092742	0.84941	0.54157	0.97687	0.79149	0.01	0.62697
	17	0.85214	0.099977	0.85047	0.55513	0.97713	0.80268	0.01	0.60766
	16	0.83979	0.10748	0.85159	0.56903	0.9774	0.81377	0.01	0.58781
	15	0.82618	0.11594	0.85284	0.58381	0.97772	0.82565	0.01	0.56696
	14	0.8119	0.1246	0.85416	0.59914	0.97807	0.83777	0.01	0.54561
	13	0.79652	0.13394	0.8556	0.61441	0.97844	0.84934	0.01	0.52345
	12	0.77905	0.14511	0.85721	0.63094	0.97892	0.86237	0.01	0.49983
	11	0.76003	0.15732	0.85906	0.64846	0.97933	0.87386	0.01	0.47519
	10	0.73928	0.17073	0.86111	0.66618	0.97976	0.88468	0.01	0.44945
	9	0.71436	0.18837	0.86358	0.68642	0.98037	0.89772	0.01	0.42125
	8	0.6887	0.20544	0.86629	0.70667	0.98097	0.90957	0.01	0.39263
	7	0.65824	0.22734	0.86938	0.72922	0.98182	0.9231	0.01	0.36145
	6	0.62545	0.25032	0.87304	0.7527	0.98256	0.93432	0.01	0.32909
	5	0.5848	0.28201	0.87758	0.77906	0.98356	0.94676	0.01	0.29307
	4	0.53473	0.32424	0.88369	0.80784	0.9847	0.95792	0.01	0.25309
	3	0.47794	0.37157	0.89097	0.83944	0.98626	0.97008	0.01	0.21051
	2	0.39244	0.4612	0.90314	0.878	0.98837	0.9814	0.01	0.15826
	1	0.20347	0.75981	0.9381	0.9381	0.99354	0.99354	0.01	0.08059
Espiral	2	1	0	0.23407	0.19072	0.42319	0.36817	0	1
	1	0.53428	0.65709	0.54258	0.54258	0.70114	0.70114	0	0.51716
Vogal	10	1	1E-015	0.62451	0.16268	0.8621	0.32432	0.13	1
	9	0.9915	0.021169	0.6261	0.17783	0.86297	0.34758	0.13	0.96829
	8	0.96194	0.076647	0.63183	0.2144	0.86621	0.40083	0.13	0.90914
	7	0.92078	0.12802	0.64013	0.26625	0.87091	0.4715	0.13	0.83936
	6	0.86921	0.18137	0.65095	0.32692	0.87678	0.54702	0.13	0.76125
	5	0.80424	0.24311	0.66528	0.39949	0.88456	0.63159	0.13	0.67357
	4	0.73109	0.30383	0.68281	0.47783	0.89263	0.70811	0.13	0.58053
	3	0.63073	0.39348	0.70974	0.57236	0.90418	0.78753	0.13	0.47156
	2	0.46294	0.5741	0.76165	0.70072	0.92748	0.88285	0.13	0.33066
	1	0.25204	0.77826	0.845	0.845	0.95942	0.95942	0.13	0.17269

	6	6.4946	8.9343	0.14618	1.1026	0.21222	3.0906	1.82
	5	5.9369	8.7759	0.16135	0.87817	0.23235	2.2453	1.82
	4	5.1838	8.4742	0.20446	0.62488	0.28163	1.5103	1.82
	3	4.3267	8.0742	0.27999	0.40322	0.38477	0.77766	1.82
	2	3.2755	7.4274	0.39549	0.34147	0.49227	0.4181	1.82
	1	2.1172	6.5575	0.55357	0.55357	0.6274	0.6274	1.82
Imagens	18	10.398	10.531	0.12147	1.849	0.14554	5.4311	0.07
	17	10.353	10.531	0.11869	1.794	0.1431	5.2476	0.07
	16	10.306	10.53	0.11611	1.7375	0.14058	5.0542	0.07
	15	10.254	10.53	0.11305	1.6778	0.13779	4.8533	0.07
	14	10.202	10.529	0.11007	1.6177	0.13415	4.6233	0.07
	13	10.134	10.529	0.10621	1.5461	0.12948	4.3552	0.07
	12	10.057	10.528	0.10205	1.4734	0.1253	4.113	0.07
	11	9.9692	10.526	0.097944	1.4003	0.12111	3.8805	0.07
	10	9.8667	10.525	0.093496	1.3158	0.11623	3.6263	0.07
	9	9.7486	10.522	0.087886	1.2255	0.10966	3.3462	0.07
	8	9.6022	10.519	0.083901	1.1313	0.10488	3.1069	0.07
	7	9.4309	10.514	0.078645	1.0207	0.095067	2.7237	0.07
	6	9.2126	10.507	0.076702	0.89832	0.087039	2.3202	0.07
	5	8.9145	10.493	0.08097	0.75481	0.086983	1.921	0.07
	4	8.4945	10.466	0.090822	0.59647	0.09934	1.5415	0.07
	3	7.8585	10.399	0.12965	0.4088	0.12421	1.1624	0.07
	2	6.8555	10.225	0.23803	0.28018	0.19696	0.70371	0.07
	1	5.104	9.6856	0.36627	0.36627	0.56332	0.56332	0.07
Iris	4	3.6164	2.6578	0.082671	0.56823	0.10752	1.3287	0.03
	3	3.5252	2.6564	0.079769	0.44344	0.099731	1.0182	0.03
	2	3.3478	2.6464	0.079901	0.26633	0.095284	0.59423	0.03
	1	2.6969	2.5994	0.23803	0.23803	0.25501	0.25501	0.03
Sonar	60	45.038	41.352	0.098679	4.4318	0.17283	26.231	0.07
	59	44.958	41.351	0.097646	4.3954	0.17113	25.887	0.07
	58	44.876	41.35	0.096571	4.3593	0.1694	25.556	0.07
	57	44.792	41.349	0.095461	4.323	0.16762	25.222	0.07
	56	44.705	41.348	0.09434	4.2855	0.16573	24.874	0.07
	55	44.615	41.347	0.093166	4.2473	0.16383	24.528	0.07
	54	44.52	41.346	0.091969	4.208	0.16184	24.172	0.07
	53	44.421	41.345	0.090685	4.1678	0.1597	23.808	0.07
	52	44.319	41.344	0.089409	4.127	0.15753	23.441	0.07
	51	44.212	41.343	0.088048	4.0847	0.15524	23.061	0.07
	50	44.103	41.341	0.086634	4.0425	0.15297	22.695	0.07
	49	43.988	41.34	0.085186	3.9994	0.15062	22.321	0.07
	48	43.865	41.338	0.083628	3.9544	0.14808	21.931	0.07
	47	43.736	41.336	0.081995	3.9079	0.14523	21.521	0.07
	46	43.604	41.334	0.080386	3.8615	0.14236	21.112	0.07
	45	43.463	41.331	0.078679	3.8135	0.1395	20.707	0.07
	44	43.315	41.329	0.076924	3.7636	0.13635	20.273	0.07
	43	43.165	41.326	0.07501	3.7141	0.13333	19.868	0.07
	42	42.997	41.322	0.072965	3.6613	0.12974	19.426	0.07
	41	42.827	41.319	0.070953	3.6092	0.12642	19.003	0.07
	40	42.648	41.315	0.068835	3.5549	0.12281	18.562	0.07
	39	42.469	41.311	0.066672	3.5004	0.11913	18.126	0.07
	38	42.275	41.307	0.064435	3.4442	0.11528	17.68	0.07
	37	42.075	41.302	0.062132	3.387	0.11129	17.227	0.07
	36	41.867	41.297	0.059805	3.3284	0.10728	16.772	0.07
	35	41.648	41.291	0.057369	3.2683	0.10326	16.325	0.07
	34	41.42	41.285	0.054943	3.2076	0.099212	15.871	0.07
	33	41.187	41.278	0.05266	3.1463	0.095053	15.408	0.07
	32	40.945	41.271	0.050318	3.0833	0.090942	14.944	0.07
	31	40.692	41.263	0.047761	3.0191	0.085964	14.469	0.07
	30	40.434	41.255	0.045673	2.9551	0.082077	14.012	0.07
	29	40.157	41.246	0.043637	2.8876	0.078263	13.543	0.07

28	39.854	41.235	0.041661	2.8175	0.074295	13.051	0.07	
27	39.533	41.223	0.0401	2.7443	0.070636	12.546	0.07	
26	39.195	41.209	0.038533	2.6697	0.066894	12.025	0.07	
25	38.846	41.194	0.037583	2.5933	0.063997	11.51	0.07	
24	38.486	41.179	0.037588	2.515	0.062474	10.993	0.07	
23	38.109	41.162	0.038731	2.4368	0.062865	10.498	0.07	
22	37.715	41.143	0.040266	2.3567	0.063788	10.01	0.07	
21	37.295	41.122	0.043857	2.2742	0.068106	9.5191	0.07	
20	36.833	41.096	0.048642	2.1879	0.07413	9.0091	0.07	
19	36.362	41.07	0.053116	2.1002	0.080279	8.4791	0.07	
18	35.858	41.039	0.057939	2.0091	0.087688	7.9461	0.07	
17	35.326	41.005	0.064128	1.9159	0.096124	7.437	0.07	
16	34.759	40.967	0.070556	1.8176	0.10652	6.8868	0.07	
15	34.172	40.926	0.077994	1.7177	0.11805	6.3568	0.07	
14	33.542	40.878	0.087782	1.6183	0.13295	5.8607	0.07	
13	32.88	40.825	0.096509	1.5136	0.14672	5.3369	0.07	
12	32.199	40.769	0.10664	1.4097	0.16255	4.8428	0.07	
11	31.406	40.694	0.11785	1.2943	0.1806	4.3039	0.07	
10	30.575	40.61	0.13143	1.1785	0.19974	3.816	0.07	
9	29.691	40.516	0.14707	1.0615	0.22119	3.354	0.07	
8	28.674	40.393	0.16397	0.93275	0.24532	2.853	0.07	
7	27.537	40.236	0.18571	0.8042	0.27331	2.3993	0.07	
6	26.175	40.01	0.21097	0.66111	0.3086	1.9082	0.07	
5	24.693	39.744	0.24035	0.52397	0.34449	1.4766	0.07	
4	22.542	39.172	0.28593	0.38213	0.39702	1.0319	0.07	
3	20.262	38.52	0.33428	0.2816	0.45924	0.61252	0.07	
2	16.594	36.778	0.41772	0.32823	0.54975	0.44057	0.07	
1	9.0509	28.31	0.62174	0.62174	0.76439	0.76439	0.07	
Espiral	2	1.6521	0.71319	0.11802	0.33491	0.18137	0.68927	0.01
	1	0.90358	0.86013	0.48344	0.48344	0.59072	0.59072	0.01
Vogal	10	7.6873	6.7021	0.075407	1.5392	0.12958	5.1987	6.12
	9	7.4835	6.6837	0.069404	1.4049	0.11821	4.5972	6.12
	8	7.2168	6.6552	0.067334	1.252	0.1125	3.957	6.12
	7	6.8982	6.6178	0.075539	1.0839	0.1203	3.2887	6.12
	6	6.5046	6.5607	0.09637	0.90103	0.14811	2.6141	6.12
	5	6.0636	6.4925	0.12995	0.7118	0.19334	1.9744	6.12
	4	5.5674	6.406	0.17753	0.52877	0.24956	1.4299	6.12
	3	4.848	6.2187	0.25524	0.35252	0.34009	0.90429	6.12
	2	3.6052	5.6226	0.39884	0.33987	0.50954	0.50461	6.12
	1	2.0201	4.4832	0.61125	0.61125	0.72453	0.72453	6.12
Cenário: Z e D Manhattan								
MDS Stress-1								
Base de dados	k	Prop Var Ac	Strain	Stress-1 (M->E)	Stress-1 (M->M)	Sstress (M->E)	Sstress (M->M)	Tempo (s)
Triagem de Crédito	15	6.0153	7.2692	0.078823	1.5904	0.12347	5.2652	116.84
	14	6.0155	7.2709	0.078823	1.5808	0.12345	5.2205	118.88
	13	6.0155	7.2706	0.078823	1.5748	0.12346	5.1797	166.72
	12	6.0151	7.2821	0.078843	1.5398	0.12345	5.0103	87.95
	11	6.0154	7.3278	0.079091	1.4758	0.12381	4.7099	153.51
	10	6.0145	7.3933	0.079527	1.4132	0.12451	4.4253	115.53
	9	6.0138	7.4826	0.080452	1.3368	0.1256	4.1157	97.28
	8	6.011	7.6559	0.082967	1.2389	0.12962	3.6858	110.81
	7	6.0066	7.9348	0.087581	1.1337	0.13666	3.2967	106.72
	6	5.9991	8.3369	0.094525	1.0078	0.14856	2.7604	113.08
	5	5.9808	8.853	0.10889	0.86882	0.16717	2.3059	53
	4	5.941	9.699	0.13626	0.71292	0.20616	1.8414	50.96
	3	5.8621	10.982	0.17779	0.53739	0.26273	1.2842	118.36
	2	5.6353	13.016	0.2627	0.38882	0.37106	0.82493	184.92
	1	4.4835	14.925	0.5092	0.5092	0.68897	0.68897	43.21
	18	8.7632	10.116	0.035211	1.1466	0.051862	2.9689	10.84

Imagens	17	8.7633	10.117	0.035211	1.1466	0.051862	2.9685	10.42
	16	8.7628	10.115	0.035211	1.1425	0.051867	2.9609	9.86
	15	8.7632	10.116	0.035211	1.1425	0.051866	2.9567	11.29
	14	8.763	10.116	0.035211	1.1396	0.051868	2.9489	12.97
	13	8.7628	10.116	0.035211	1.1305	0.051867	2.9237	12.3
	12	8.7627	10.118	0.035225	1.107	0.051883	2.8459	25.55
	11	8.7628	10.122	0.035259	1.0825	0.051874	2.7839	20.83
	10	8.7626	10.13	0.035339	1.046	0.051912	2.6824	14.38
	9	8.7628	10.143	0.035593	1.0035	0.052024	2.5562	12.11
	8	8.7627	10.16	0.035937	0.95633	0.052176	2.4192	5.21
	7	8.7614	10.195	0.037239	0.89835	0.053642	2.2787	5.83
	6	8.7598	10.29	0.04004	0.81846	0.056153	2.0741	7.33
	5	8.7567	10.399	0.044582	0.73228	0.060699	1.8933	7.57
	4	8.748	10.664	0.054254	0.61249	0.074096	1.5565	7.37
	3	8.7205	11.213	0.077746	0.49851	0.10446	1.377	3.42
	2	8.6085	12.678	0.13752	0.30941	0.22068	0.85449	4.22
	1	7.2712	14.807	0.41384	0.41384	0.61122	0.61122	1.71
Iris	4	3.3685	2.571	0.04114	0.4422	0.069519	0.96129	1.38
	3	3.3674	2.591	0.044689	0.37057	0.073083	0.79919	1.05
	2	3.362	2.6579	0.060265	0.26206	0.079996	0.57071	0.64
	1	2.3278	3.1784	0.55694	0.55694	0.83464	0.83464	0.51
Sonar	60	37.857	39.661	0.019074	3.3204	0.035975	16.574	31.96
	59	37.856	39.66	0.019074	3.3139	0.035976	16.523	42.57
	58	37.855	39.66	0.019074	3.3135	0.035977	16.521	48.79
	57	37.857	39.661	0.019074	3.3122	0.035975	16.51	54.75
	56	37.856	39.66	0.019074	3.308	0.035976	16.478	28.48
	55	37.855	39.66	0.019073	3.3085	0.035977	16.483	51.65
	54	37.857	39.662	0.019074	3.3022	0.035972	16.436	43.69
	53	37.856	39.66	0.019073	3.3029	0.035975	16.439	32.55
	52	37.857	39.662	0.019074	3.2948	0.035973	16.377	59.94
	51	37.855	39.662	0.019074	3.2889	0.035977	16.331	27.11
	50	37.855	39.663	0.019075	3.2782	0.035975	16.252	49.88
	49	37.855	39.664	0.019076	3.2681	0.035975	16.175	60.98
	48	37.856	39.668	0.019079	3.2546	0.035972	16.075	67.41
	47	37.856	39.671	0.019081	3.2399	0.035974	15.961	71.61
	46	37.857	39.675	0.019086	3.2229	0.035969	15.838	68.45
	45	37.855	39.679	0.019094	3.203	0.035981	15.687	59.78
	44	37.857	39.689	0.01911	3.182	0.035991	15.532	59.65
	43	37.857	39.701	0.019139	3.1571	0.036015	15.347	75.92
	42	37.856	39.707	0.019153	3.1372	0.036025	15.2	52.94
	41	37.855	39.719	0.019178	3.1103	0.036062	15	52.34
	40	37.857	39.736	0.019211	3.0853	0.036091	14.814	67.17
	39	37.855	39.75	0.019249	3.0584	0.036132	14.619	55.42
	38	37.853	39.767	0.019303	3.0288	0.036216	14.396	48.05
	37	37.855	39.788	0.019356	3	0.036283	14.191	63.82
	36	37.855	39.813	0.019429	2.9668	0.036374	13.951	53.91
	35	37.854	39.842	0.019525	2.9331	0.036525	13.708	44
	34	37.856	39.875	0.019641	2.8983	0.036705	13.459	65.35
	33	37.855	39.909	0.019763	2.8634	0.03685	13.209	38.37
	32	37.854	39.947	0.019912	2.8258	0.037049	12.95	61.2
	31	37.854	39.987	0.020099	2.7904	0.037252	12.711	37.33
	30	37.855	40.035	0.020321	2.7515	0.037574	12.447	59.9
	29	37.855	40.092	0.020565	2.708	0.03789	12.155	36.29
	28	37.85	40.15	0.020866	2.6643	0.0384	11.851	59.49
	27	37.855	40.232	0.021217	2.6218	0.03889	11.563	32.1
	26	37.851	40.314	0.021622	2.5763	0.039569	11.265	54.32
	25	37.854	40.402	0.022105	2.5288	0.040244	10.954	31.86
	24	37.849	40.501	0.022681	2.4794	0.041163	10.642	51.52
	23	37.846	40.616	0.023444	2.4304	0.042227	10.333	30.37
	22	37.846	40.758	0.024211	2.3746	0.043467	9.9838	40.27

21	37.846	40.923	0.025147	2.3181	0.044879	9.6332	32.84	
20	37.838	41.104	0.026282	2.2599	0.046752	9.2823	31.71	
19	37.842	41.317	0.027478	2.1976	0.048597	8.907	47.56	
18	37.839	41.569	0.028957	2.1342	0.051089	8.538	22.13	
17	37.834	41.861	0.030645	2.0673	0.053848	8.1552	40.23	
16	37.83	42.179	0.032436	1.9995	0.056733	7.7739	32.64	
15	37.827	42.56	0.034556	1.9232	0.060376	7.3503	20.75	
14	37.816	42.981	0.036982	1.8482	0.064218	6.9511	32.83	
13	37.808	43.482	0.039928	1.7693	0.068875	6.5425	37.19	
12	37.8	44.122	0.043283	1.6844	0.074492	6.1029	12.91	
11	37.784	44.875	0.047401	1.5899	0.081285	5.6328	28.84	
10	37.767	45.777	0.05256	1.4912	0.088947	5.1656	20.61	
9	37.737	46.903	0.058827	1.3877	0.098875	4.6968	31.98	
8	37.7	48.297	0.066389	1.276	0.11023	4.2083	11.3	
7	37.644	50.107	0.076791	1.1568	0.12589	3.7212	12.03	
6	37.539	52.577	0.090458	1.0242	0.14676	3.1935	19.29	
5	37.421	55.936	0.10934	0.87757	0.17345	2.6525	12.12	
4	37.151	60.772	0.13797	0.71893	0.21804	2.1146	15.53	
3	36.606	67.784	0.18301	0.5496	0.28226	1.5685	6.54	
2	35.177	78.83	0.26639	0.39557	0.40996	1.0107	10.58	
1	28.891	92.248	0.48696	0.48696	0.77318	0.77318	8.89	
Espiral	2	1.617	0.62769	0.092762	0.33584	0.16377	0.73137	0.24
	1	1.3537	1.4532	0.41235	0.41235	0.54807	0.54807	0.32
Vogal	10	7.0261	6.231	0.047665	1.3984	0.08465	4.5128	299.21
	9	7.0249	6.3147	0.049333	1.3157	0.087037	4.1658	266.08
	8	7.0223	6.4575	0.053125	1.2187	0.092014	3.7756	194.97
	7	7.0169	6.6566	0.060127	1.1095	0.1013	3.3538	191.09
	6	7.0066	6.965	0.071253	0.98491	0.11731	2.8928	176.57
	5	6.9866	7.4013	0.088291	0.84628	0.14149	2.4102	228.1
	4	6.9461	8.0184	0.11658	0.69849	0.17643	1.9448	139.95
	3	6.8452	9.0513	0.16755	0.53514	0.24847	1.4634	150.76
	2	6.5474	10.825	0.26532	0.39138	0.40886	0.97351	267.9
	1	5.1711	12.842	0.51543	0.51543	0.78316	0.78316	131.74
Cenário: Z e D Manhattan								
MDS Sstress								
Base de dados	k	Prop Var Ac	Strain	Stress-1 (M->E)	Stress-1 (M->M)	Sstress (M->E)	Sstress (M->M)	Tempo (s)
Triagem de Crédito	15	6.2677	7.46	0.083304	1.6393	0.11704	5.4325	430.77
	14	6.2681	7.4566	0.083302	1.649	0.11703	5.4892	459.32
	13	6.2673	7.4605	0.083289	1.6341	0.11704	5.415	474.85
	12	6.2672	7.4647	0.083303	1.6132	0.11706	5.2814	621.7
	11	6.261	7.5162	0.083549	1.5486	0.11771	4.9982	675.9
	10	6.2535	7.5861	0.083853	1.4761	0.11844	4.6866	646.89
	9	6.2406	7.6789	0.084551	1.3928	0.11971	4.3593	473.15
	8	6.2143	7.8495	0.086666	1.2844	0.12361	3.9099	675.81
	7	6.1748	8.0938	0.091151	1.1679	0.13063	3.5012	446.88
	6	6.12	8.4559	0.099075	1.0337	0.14315	2.9203	686.67
	5	6.0308	8.8868	0.11324	0.8827	0.16101	2.3617	321.58
	4	5.8577	9.5323	0.14342	0.70621	0.196	1.8054	595.56
	3	5.5711	10.419	0.19092	0.50835	0.24814	1.1497	473.35
	2	5.0449	11.528	0.28077	0.35905	0.33127	0.646	692.08
	1	3.7909	12.265	0.45708	0.45708	0.5212	0.5212	341.31
Censo Demográfico	18	8.884	10.143	0.045401	1.1401	0.043973	2.9709	43.9
	17	8.8834	10.143	0.045399	1.1401	0.043973	2.9719	52
	16	8.8834	10.143	0.045421	1.1343	0.043972	2.9552	65.83
	15	8.8835	10.144	0.045417	1.1383	0.043973	2.9657	41.86
	14	8.8837	10.143	0.045425	1.1363	0.043973	2.9605	68.2
	13	8.8832	10.143	0.045422	1.1283	0.043972	2.941	47.15
	12	8.8839	10.143	0.045438	1.1244	0.043976	2.9304	73.92
	11	8.8816	10.148	0.045411	1.0979	0.043995	2.8544	56.02

Imagenes	10	8.8792	10.155	0.045619	1.067	0.044062	2.776	65.11
	9	8.8765	10.166	0.045754	1.0288	0.044173	2.6666	53.79
	8	8.8662	10.189	0.045901	0.97294	0.044573	2.4933	66.58
	7	8.8487	10.223	0.04846	0.91145	0.045371	2.3164	67.24
	6	8.8277	10.289	0.051438	0.83452	0.047084	2.0986	43.92
	5	8.785	10.41	0.056486	0.73795	0.051283	1.8781	39.04
	4	8.7083	10.619	0.065383	0.62498	0.061079	1.6049	71.75
	3	8.5065	11.036	0.098854	0.47625	0.084797	1.2955	19.77
	2	8.0129	11.88	0.16535	0.30445	0.15664	0.81008	14.43
	1	6.9937	13.656	0.29638	0.29638	0.30043	0.30043	7.16
Iris	4	3.4159	2.5712	0.050349	0.46835	0.061595	1.0231	9.12
	3	3.3995	2.5918	0.057269	0.39281	0.063562	0.85717	12.36
	2	3.3545	2.6601	0.066347	0.26072	0.07346	0.56325	3.57
	1	3.0009	3.0058	0.21446	0.21446	0.21367	0.21367	2.75
Sonar	60	38.001	39.959	0.020102	3.2545	0.034429	16.082	124.91
	59	38.001	39.959	0.020102	3.2516	0.034429	16.06	136.89
	58	38.001	39.96	0.020102	3.2509	0.034429	16.056	127.98
	57	38.002	39.96	0.020104	3.2468	0.03443	16.024	135.08
	56	38	39.958	0.020101	3.2489	0.03443	16.039	127.69
	55	38	39.96	0.020103	3.2417	0.034431	15.986	119.68
	54	38.001	39.96	0.020104	3.2413	0.03443	15.983	129.1
	53	38.001	39.959	0.020104	3.2409	0.03443	15.981	105.12
	52	38.001	39.961	0.020106	3.2326	0.03443	15.92	122.88
	51	38.001	39.961	0.020106	3.2267	0.03443	15.874	103.31
	50	38	39.962	0.020111	3.2251	0.034434	15.864	119.08
	49	38	39.958	0.020108	3.2244	0.034432	15.863	124.57
	48	38.001	39.966	0.020114	3.2014	0.034439	15.684	111.33
	47	38	39.969	0.020116	3.2019	0.034438	15.688	116.04
	46	37.999	39.965	0.020121	3.1924	0.03445	15.614	128.93
	45	38	39.972	0.020125	3.1798	0.034453	15.518	128.19
	44	37.998	39.981	0.02014	3.1595	0.034463	15.373	108.59
	43	37.998	39.991	0.020169	3.1393	0.034489	15.222	127.76
	42	37.996	39.996	0.020176	3.1196	0.034495	15.083	109.06
	41	37.995	40.01	0.020213	3.0975	0.034524	14.925	90.82
	40	37.992	40.021	0.020251	3.0706	0.034568	14.725	68.27
	39	37.992	40.041	0.020293	3.0459	0.034609	14.551	87.1
	38	37.99	40.049	0.02035	3.0213	0.034668	14.371	72.83
	37	37.987	40.072	0.020413	2.9898	0.034744	14.138	78.3
	36	37.984	40.097	0.020473	2.9602	0.034807	13.93	79.51
	35	37.98	40.126	0.020598	2.9254	0.03498	13.674	67.29
	34	37.977	40.159	0.02072	2.8938	0.035071	13.466	66.11
	33	37.973	40.192	0.020848	2.8585	0.035245	13.211	68.01
	32	37.967	40.231	0.02105	2.8249	0.035486	12.983	64.71
	31	37.962	40.271	0.021194	2.7867	0.035623	12.72	64.6
	30	37.956	40.321	0.021422	2.7474	0.035902	12.454	64.06
	29	37.949	40.374	0.021671	2.7081	0.0362	12.192	63.79
	28	37.939	40.448	0.022039	2.6652	0.036685	11.897	63.77
	27	37.931	40.518	0.022503	2.6189	0.037246	11.585	62.86
	26	37.922	40.609	0.022827	2.5721	0.037806	11.278	62.43
	25	37.909	40.7	0.023325	2.5227	0.038456	10.947	62.1
	24	37.895	40.81	0.023987	2.4718	0.039305	10.624	61.79
	23	37.88	40.928	0.024704	2.4194	0.040273	10.295	62.75
	22	37.861	41.056	0.025573	2.3653	0.041369	9.9607	57.64
	21	37.842	41.212	0.026609	2.3118	0.042837	9.627	55.23
	20	37.817	41.399	0.027813	2.2497	0.044629	9.2547	55.29
	19	37.788	41.612	0.02902	2.1845	0.046415	8.855	54.88
	18	37.757	41.835	0.030647	2.1218	0.048519	8.4963	54.91
	17	37.719	42.102	0.032486	2.0537	0.051002	8.1059	54.6
	16	37.677	42.417	0.03442	1.9816	0.053798	7.7001	54.66
	15	37.626	42.766	0.036831	1.9032	0.057096	7.2635	53.65

14	37.568	43.194	0.039364	1.8237	0.060952	6.8296	48.34	
13	37.497	43.661	0.042599	1.7425	0.065407	6.4092	53.05	
12	37.416	44.228	0.046243	1.6579	0.070449	5.9737	39.58	
11	37.313	44.887	0.050877	1.5629	0.076671	5.5101	52.05	
10	37.184	45.658	0.05628	1.4614	0.083864	5.033	50.44	
9	37.018	46.584	0.063344	1.3575	0.092703	4.5754	42.58	
8	36.81	47.771	0.071373	1.2408	0.1034	4.067	49.39	
7	36.528	49.224	0.082675	1.1183	0.1177	3.566	28.97	
6	36.131	51.138	0.097601	0.98231	0.13678	3.0342	21.49	
5	35.564	53.597	0.11714	0.82832	0.16085	2.4709	27.66	
4	34.664	57.057	0.14743	0.66119	0.19967	1.893	32.46	
3	33.163	61.466	0.198	0.49245	0.25155	1.3358	25.36	
2	30.276	67.641	0.28529	0.35439	0.34041	0.77131	16.18	
1	21.579	68.623	0.49484	0.49484	0.61953	0.61953	34.45	
Espiral	2	1.6355	0.67697	0.10021	0.33263	0.15368	0.69863	2.58
	1	1.2306	1.2893	0.40054	0.40054	0.49816	0.49816	2.76
Vogal	10	7.0975	6.315	0.049136	1.4076	0.08207	4.5553	1138.4
	9	7.0825	6.3913	0.050824	1.3223	0.084478	4.1948	667.09
	8	7.0563	6.5178	0.05455	1.2208	0.089676	3.7851	660.13
	7	7.0161	6.6933	0.061698	1.1048	0.098717	3.3339	959.56
	6	6.9544	6.9391	0.073235	0.97368	0.11394	2.8494	818.81
	5	6.8602	7.2823	0.091398	0.82528	0.1358	2.3314	778.42
	4	6.7076	7.7196	0.12131	0.66827	0.16744	1.8347	669.7
	3	6.3895	8.3862	0.175	0.48754	0.2312	1.2959	519.86
	2	5.7019	9.3173	0.2799	0.34533	0.3642	0.74953	1330.4
	1	4.1434	9.9247	0.50328	0.50328	0.61106	0.61106	1031.3

Cenário: Z e D Manhattan

FastMap

Base de dados	k	Prop Var Ac	Strain	Stress-1 (M->E)	Stress-1 (M->M)	Sstress (M->E)	Sstress (M->M)	Tempo (s)
Triagem de Crédito	15	0.99999	74.717	0.60189	0.2644	0.83494	0.52748	0.03
	14	0.99999	74.717	0.60189	0.2644	0.83494	0.52748	0
	13	0.94905	73.545	0.6129	0.2166	0.84012	0.44745	0
	12	0.8869	71.186	0.62715	0.18819	0.84644	0.39079	0.01
	11	0.82026	69.44	0.64347	0.19521	0.85315	0.3796	0
	10	0.76748	68.527	0.65669	0.22532	0.85864	0.39319	0
	9	0.7151	66.384	0.66983	0.26847	0.86507	0.4257	0
	8	0.64591	62.024	0.68903	0.33598	0.87289	0.49194	0
	7	0.57534	57.926	0.70996	0.41032	0.88068	0.56283	0
	6	0.50294	53.149	0.73229	0.48814	0.89073	0.64283	0
	5	0.402	47.898	0.76789	0.58745	0.90353	0.7265	0
	4	0.30582	40.358	0.80741	0.68603	0.91648	0.7931	0
	3	0.2363	32.612	0.83693	0.76229	0.92732	0.84892	0.01
	2	0.16442	27.539	0.87091	0.83538	0.94212	0.90573	0.01
	1	0.0924	12.745	0.91373	0.91373	0.95689	0.95689	0
Imagens	18	1	10.935	0.66658	0.21042	0.87796	0.49784	0.03
	17	1	10.935	0.66658	0.21042	0.87796	0.49784	0
	16	1	10.935	0.66658	0.21042	0.87796	0.49784	0
	15	1	10.935	0.66658	0.21042	0.87796	0.49784	0
	14	1	10.935	0.66658	0.21042	0.87796	0.49784	0.01
	13	0.99997	10.935	0.66659	0.2109	0.87796	0.49847	0
	12	0.99331	10.869	0.6677	0.2194	0.87821	0.50768	0.01
	11	0.98735	10.822	0.66868	0.22847	0.8785	0.51668	0
	10	0.96603	10.661	0.67226	0.25175	0.87949	0.54149	0
	9	0.92813	9.8844	0.67884	0.28947	0.8813	0.5738	0
	8	0.88573	9.5059	0.68678	0.33716	0.88388	0.6124	0
	7	0.82517	8.5829	0.69922	0.39686	0.88717	0.65735	0

	3	0.36324	3.8062	0.82199	0.75901	0.92141	0.89195	0
	2	0.31842	2.5785	0.8371	0.79947	0.92433	0.90805	0.01
	1	0.18953	1.1179	0.90204	0.90204	0.93539	0.93539	0
Iris	4	1	6.7013	0.45931	0.2533	0.71472	0.45272	0.01
	3	0.98755	6.5942	0.46258	0.28336	0.71667	0.4928	0.01
	2	0.94509	6.3036	0.47486	0.35524	0.72403	0.57272	0
	1	0.67739	3.7666	0.58432	0.58432	0.79158	0.79158	0
Sonar	60	1	61.566	0.83778	0.22854	0.97391	0.44623	0.05
	59	0.99956	61.553	0.83781	0.23076	0.97392	0.44943	0.05
	58	0.99899	61.544	0.83786	0.23309	0.97393	0.45268	0.03
	57	0.99841	61.524	0.8379	0.23554	0.97394	0.45621	0.01
	56	0.9976	61.499	0.83797	0.23844	0.97396	0.46033	0
	55	0.9968	61.456	0.83804	0.2413	0.97397	0.46433	0.01
	54	0.99578	61.407	0.83812	0.24453	0.97399	0.46883	0
	53	0.99472	61.37	0.8382	0.24779	0.97401	0.47325	0
	52	0.99359	61.314	0.8383	0.25128	0.97404	0.47799	0
	51	0.99245	61.255	0.83839	0.25488	0.97406	0.48292	0
	50	0.99067	61.189	0.83853	0.25932	0.9741	0.48887	0
	49	0.9893	61.069	0.83865	0.26324	0.97413	0.49415	0
	48	0.98789	60.985	0.83876	0.26718	0.97416	0.49939	0
	47	0.98513	60.892	0.83899	0.27257	0.97421	0.50649	0
	46	0.98266	60.804	0.83919	0.27792	0.97426	0.51337	0
	45	0.98041	60.713	0.83937	0.2832	0.97431	0.52022	0
	44	0.97815	60.596	0.83956	0.28829	0.97435	0.52662	0
	43	0.97518	60.501	0.83981	0.29457	0.97442	0.53475	0.01
	42	0.97179	60.269	0.84009	0.30085	0.97449	0.54247	0
	41	0.96829	60.139	0.84038	0.30754	0.97456	0.55098	0
	40	0.96454	59.898	0.84069	0.31424	0.97464	0.55919	0
	39	0.96173	59.723	0.84092	0.32017	0.9747	0.56636	0
	38	0.95538	59.503	0.84146	0.3292	0.97483	0.57707	0
	37	0.95128	59.268	0.8418	0.33634	0.97491	0.58535	0
	36	0.94511	59.002	0.84233	0.34536	0.97503	0.59569	0
	35	0.94023	58.755	0.84274	0.35334	0.97513	0.60465	0
	34	0.93567	58.318	0.84313	0.36066	0.97522	0.61269	0
	33	0.93196	57.98	0.84345	0.3674	0.97529	0.62014	0
	32	0.92702	57.572	0.84387	0.37527	0.97539	0.62879	0
	31	0.9185	57.153	0.84461	0.38602	0.97556	0.64039	0
	30	0.91029	56.553	0.84531	0.39685	0.97575	0.65268	0
	29	0.89871	55.622	0.84633	0.40972	0.97601	0.66633	0
	28	0.89035	54.957	0.84707	0.4198	0.97617	0.6765	0
	27	0.88063	54.473	0.84793	0.43182	0.97638	0.68897	0
	26	0.87293	53.976	0.84861	0.44234	0.97655	0.69956	0
	25	0.86147	53.068	0.84964	0.45502	0.97679	0.71189	0
	24	0.85125	52.436	0.85056	0.46724	0.97701	0.72367	0
	23	0.83595	51.552	0.85197	0.48228	0.97733	0.73773	0
	22	0.82648	50.696	0.85286	0.49332	0.97752	0.74738	0
	21	0.81353	49.401	0.85409	0.50658	0.97778	0.75899	0
	20	0.79828	48.71	0.85556	0.52132	0.97809	0.7715	0
	19	0.78238	47.826	0.85704	0.53723	0.97849	0.78618	0
	18	0.76675	47	0.85858	0.55179	0.97881	0.79759	0
	17	0.75003	45.858	0.86022	0.56771	0.97919	0.81041	0
	16	0.73761	44.483	0.86144	0.5812	0.9795	0.82144	0
	15	0.72018	43.468	0.86323	0.59679	0.97987	0.83275	0
	14	0.7034	41.989	0.86494	0.61304	0.98029	0.8453	0
	13	0.67917	40.288	0.86745	0.63203	0.98086	0.85854	0
	12	0.64037	38.85	0.87171	0.65713	0.98168	0.87443	0
	11	0.61215	37.101	0.87482	0.67864	0.98238	0.88847	0
	10	0.59166	34.742	0.87716	0.69583	0.98285	0.89846	0
	9	0.54822	33.02	0.88219	0.72314	0.98393	0.91419	0
	8	0.51257	29.642	0.88655	0.74715	0.98487	0.92715	0

	7	0.4631	28.402	0.89261	0.77614	0.98619	0.94217	0
	6	0.43613	25.614	0.89632	0.79626	0.98692	0.95105	0
	5	0.41263	22.822	0.89985	0.81413	0.98743	0.95722	0
	4	0.3582	18.463	0.90724	0.84458	0.98905	0.96961	0
	3	0.30582	16.712	0.91556	0.87383	0.99064	0.97923	0
	2	0.21759	5.7704	0.93153	0.91245	0.99324	0.9889	0.01
	1	0.12131	3.638	0.9538	0.9538	0.99603	0.99603	0
Espiral	2	1	7.3695	0.23407	1.55E-005	0.42319	2.68E-005	0
	1	0.5	4.2416	0.55367	0.55367	0.73726	0.73726	0
Vogal	10	1	24.282	0.62451	0.14291	0.8621	0.28629	0.01
	9	0.98759	23.994	0.62685	0.16069	0.86345	0.31491	0.01
	8	0.94621	23.239	0.63502	0.20594	0.8683	0.38234	0
	7	0.89146	21.969	0.64628	0.26624	0.8744	0.46265	0
	6	0.84032	20.882	0.65727	0.32833	0.88046	0.54149	0
	5	0.76175	19.292	0.67555	0.40842	0.88909	0.6276	0.01
	4	0.64718	16.335	0.70403	0.51132	0.90442	0.73614	0
	3	0.55614	13.347	0.72931	0.60139	0.91641	0.81608	0
	2	0.35047	10.192	0.79611	0.74418	0.94884	0.91854	0
	1	0.18834	6.5693	0.86975	0.86975	0.97181	0.97181	0

B Gráficos dos resultados

Abaixo seguem os gráficos dos resultados do tempo de execução e das métricas de qualidade da representação dos dados de cada método para cada dimensão possível das seis bases avaliadas nos quatro cenários definidos. Também, ao final, estão os gráficos da acurácia do classificador quando executados sobre os dados não transformados (originais ou padronizados) e o dados obtidos pelos métodos analisados, nos quatro cenários e para cada dimensão possível.

Figura B.1: Legendas utilizada nos gráficos de tempo de execução e das métricas de qualidade das representações obtidas

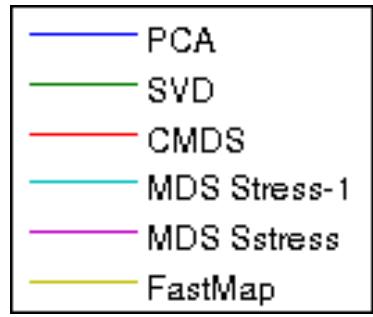


Figura B.2: Tempo de execução para base Triagem de Crédito (Legenda Fig. B.1)

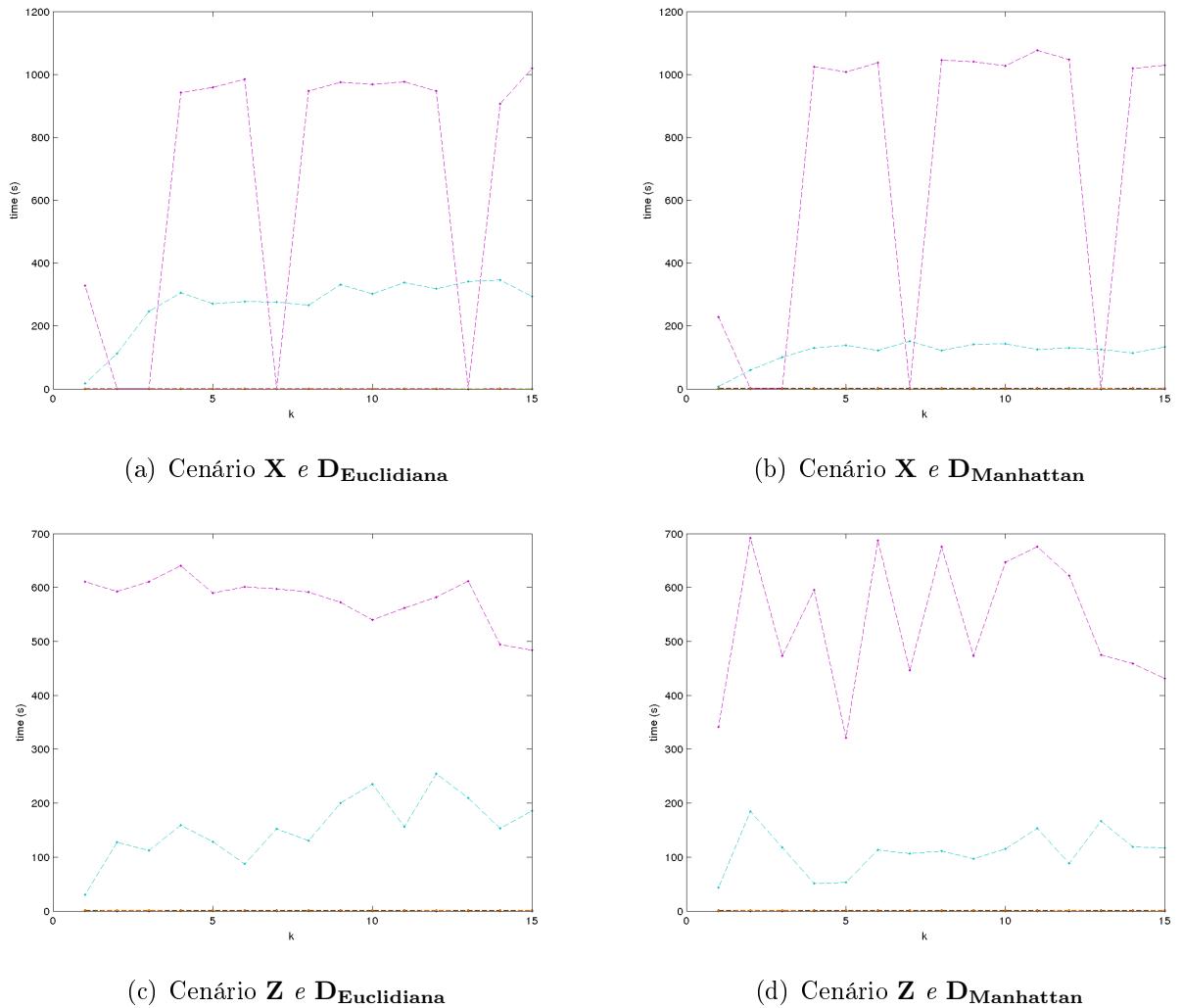


Figura B.3: Tempo de execução para base Imagens (Legenda Fig. B.1)

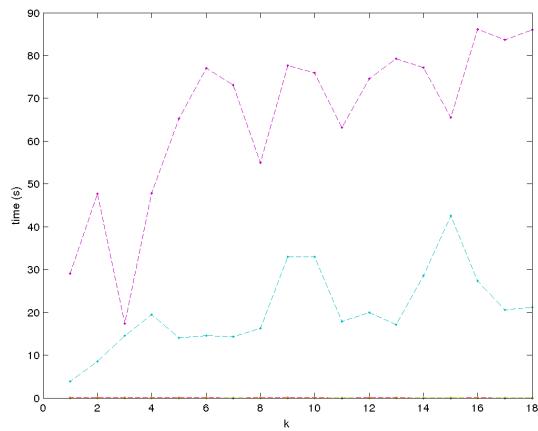
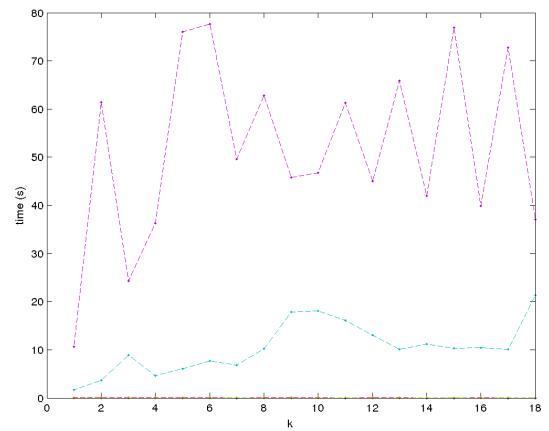
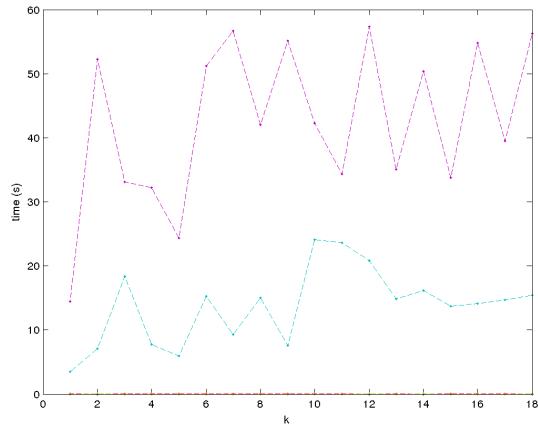
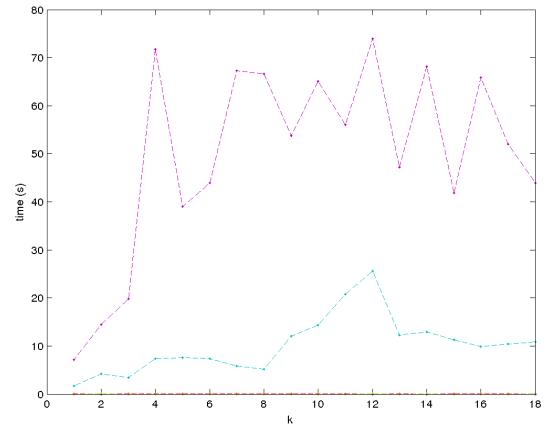
(a) Cenário X e D_{Euclidiana}(b) Cenário X e D_{Manhattan}(c) Cenário Z e D_{Euclidiana}(d) Cenário Z e D_{Manhattan}

Figura B.4: Tempo de execução para base Iris (Legenda Fig. B.1)

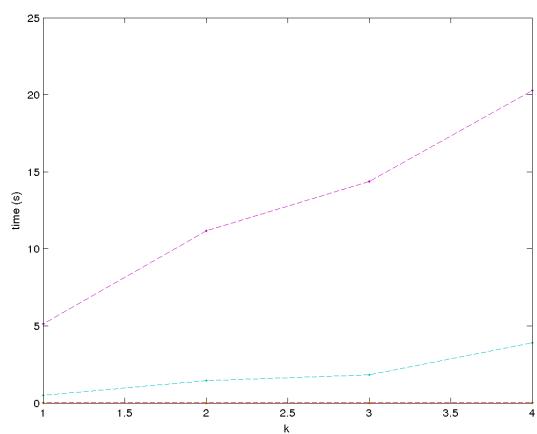
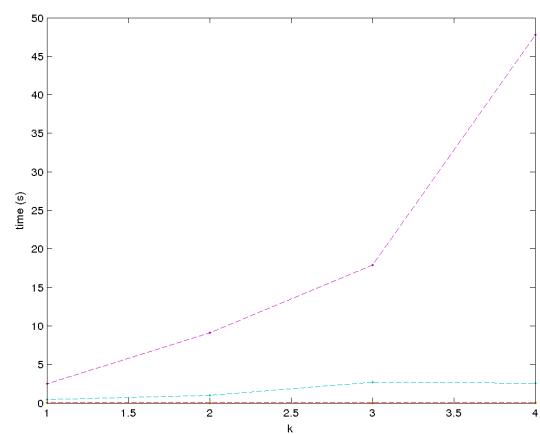
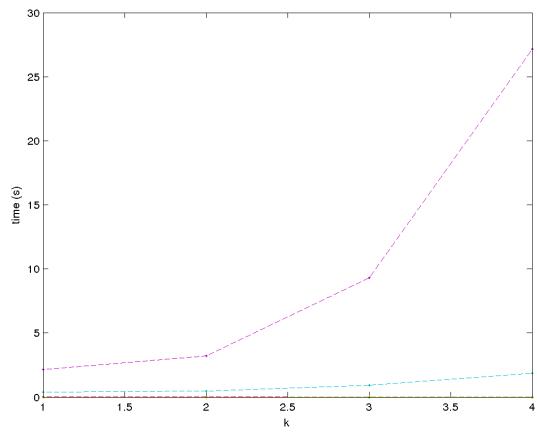
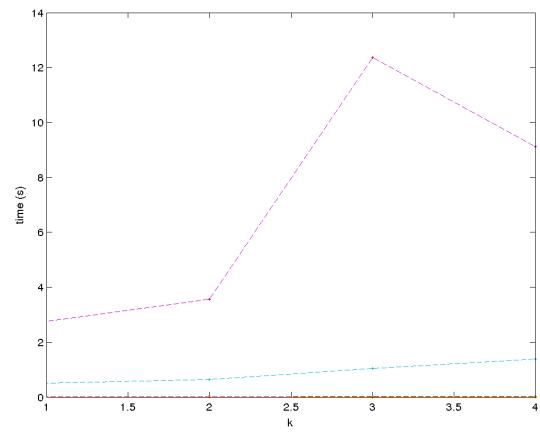
(a) Cenário **X** e $\mathbf{D}_{\text{Euclidiana}}$ (b) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$ (c) Cenário **Z** e $\mathbf{D}_{\text{Euclidiana}}$ (d) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$

Figura B.5: Tempo de execução para base Sonar (Legenda Fig. B.1)

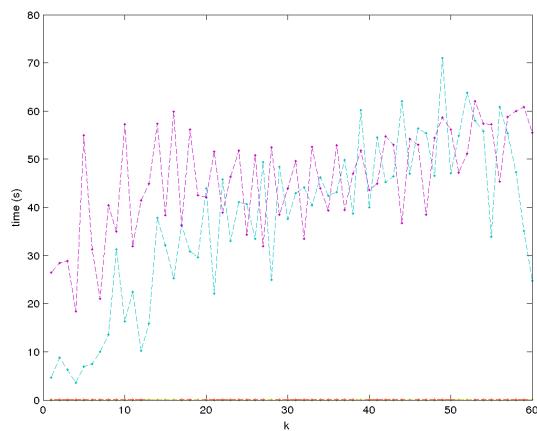
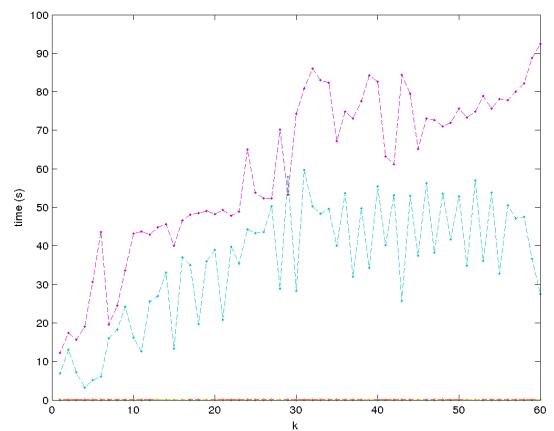
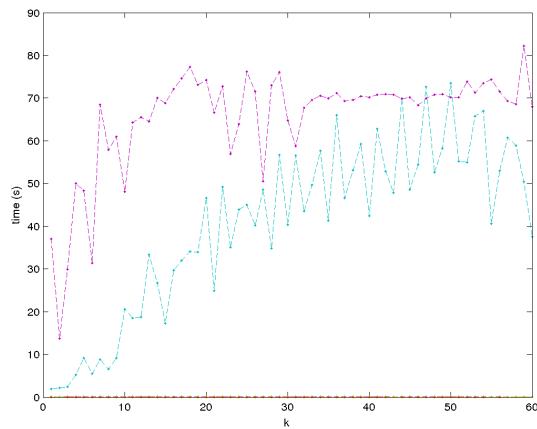
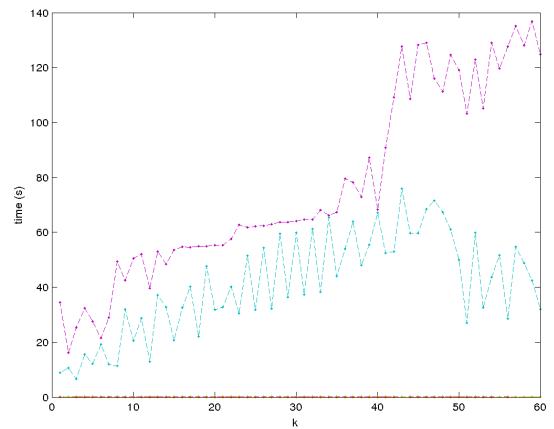
(a) Cenário **X** e $\mathbf{D}_{\text{Euclidian}}$ (b) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$ (c) Cenário **Z** e $\mathbf{D}_{\text{Euclidian}}$ (d) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$

Figura B.6: Tempo de execução para base Espiral (Legenda Fig. B.1)

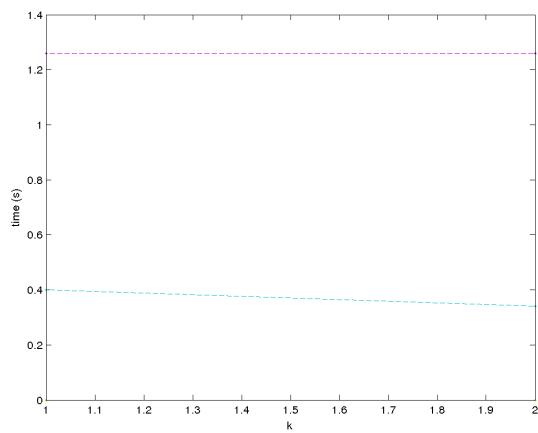
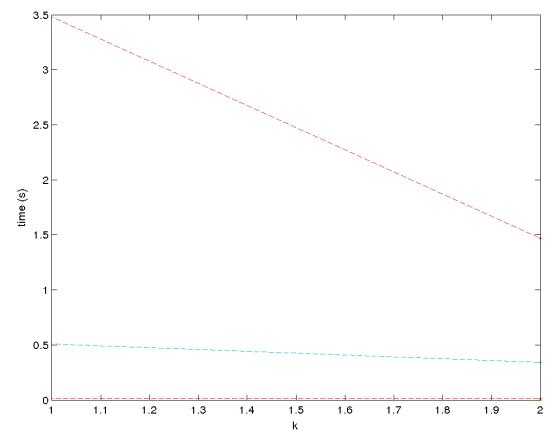
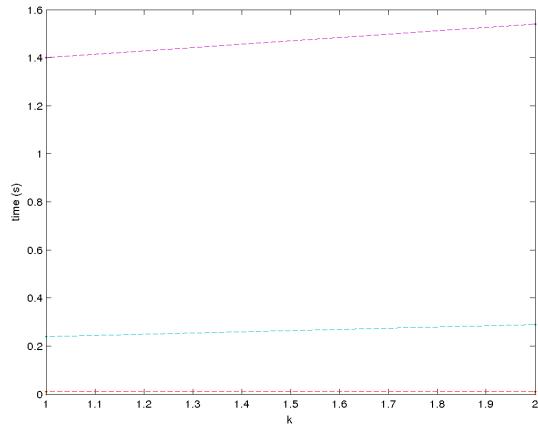
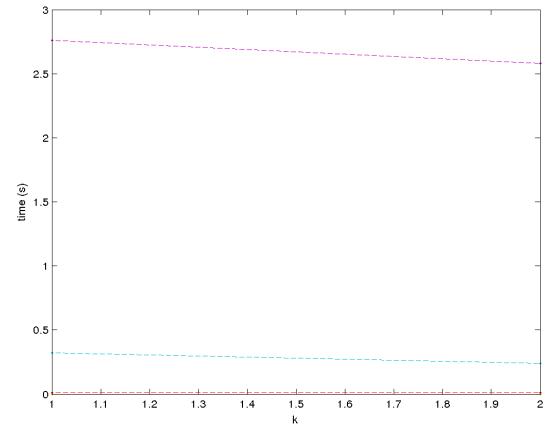
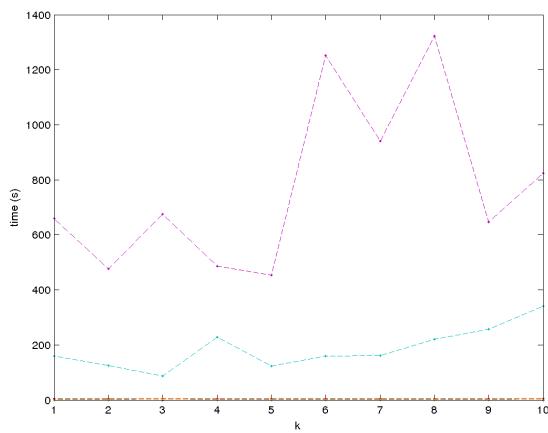
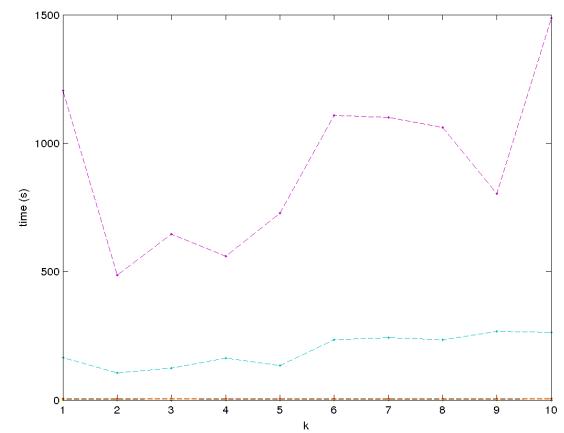
(a) Cenário **X** e **D_{Euclidian}**(b) Cenário **X** e **D_{Manhattan}**(c) Cenário **Z** e **D_{Euclidian}**(d) Cenário **Z** e **D_{Manhattan}**

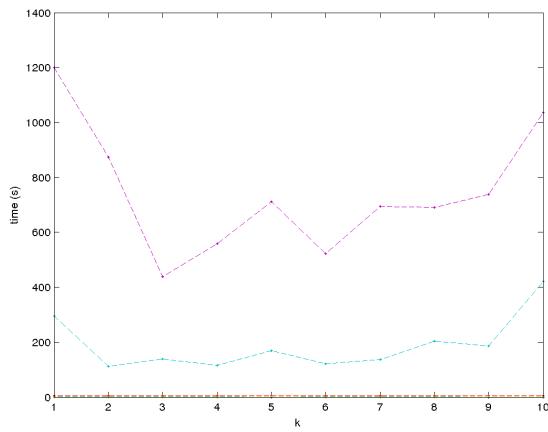
Figura B.7: Tempo de execução para base Vogal (Legenda Fig. B.1)



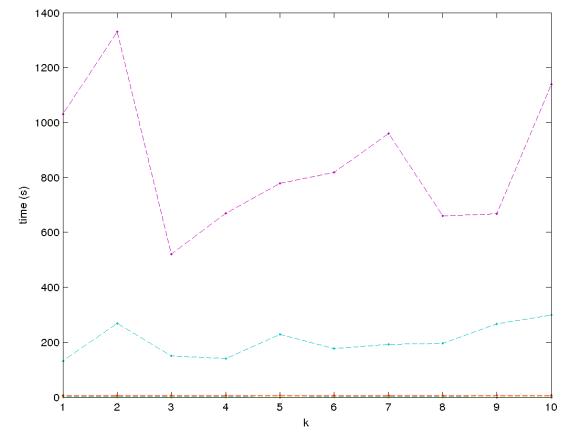
(a) Cenário X e D_{Euclidiana}



(b) Cenário X e D_{Manhattan}

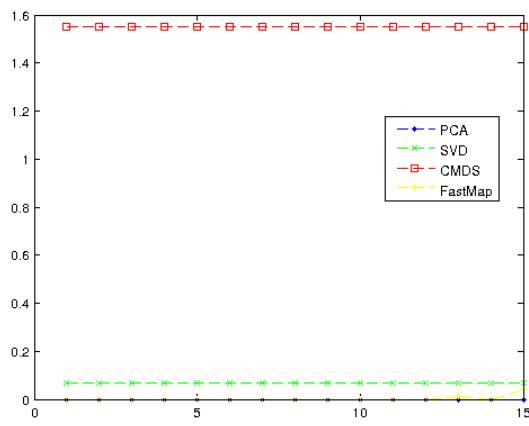


(c) Cenário Z e D_{Euclidiana}

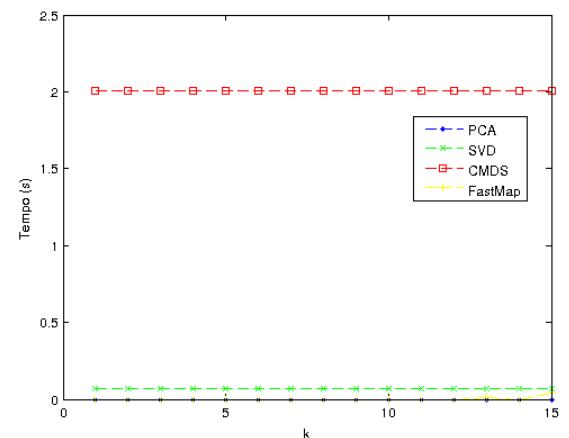


(d) Cenário Z e D_{Manhattan}

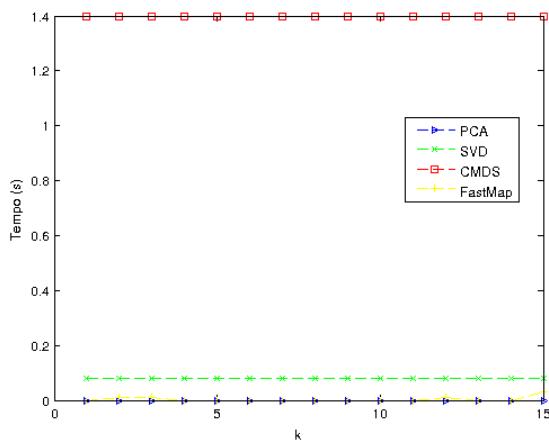
Figura B.8: Tempo de execução para base Triagem de Crédito - Excluindo métodos baseados em otimização



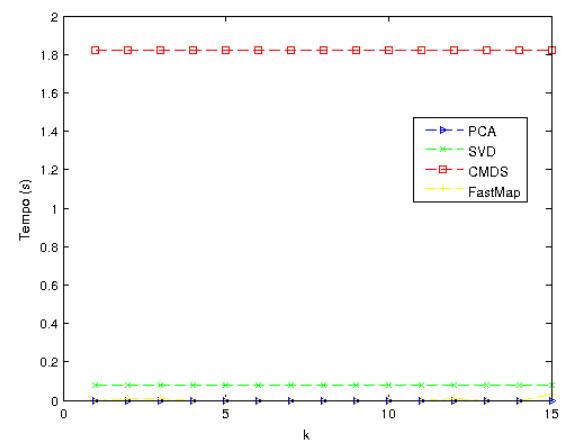
(a) Cenário \mathbf{X} e $\mathbf{D}_{\text{Euclidiana}}$



(b) Cenário \mathbf{X} e $\mathbf{D}_{\text{Manhattan}}$

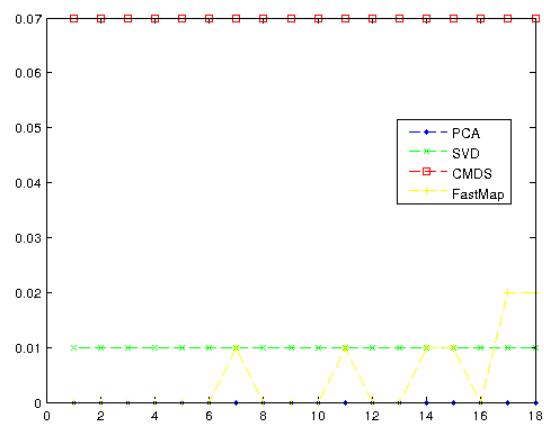


(c) Cenário \mathbf{Z} e $\mathbf{D}_{\text{Euclidiana}}$

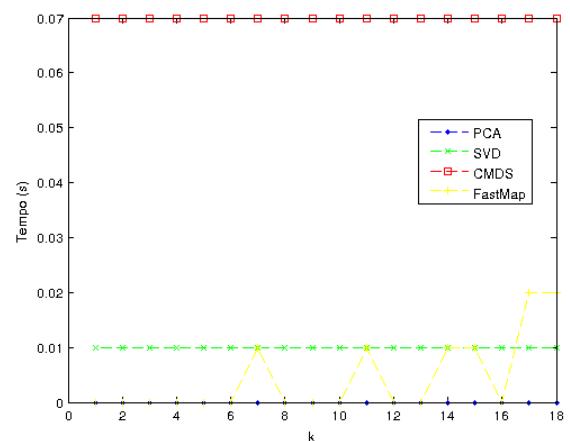


(d) Cenário \mathbf{Z} e $\mathbf{D}_{\text{Manhattan}}$

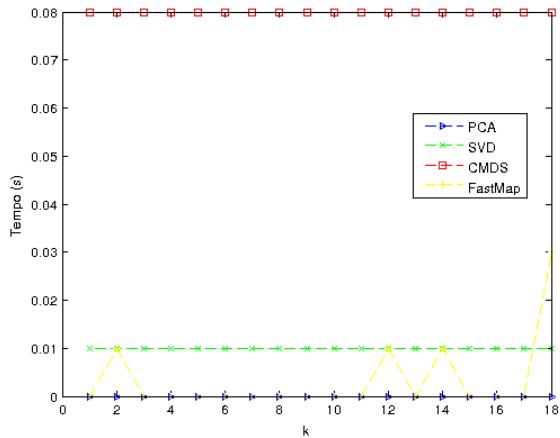
Figura B.9: Tempo de execução para base Imagens - Excluindo métodos baseados em otimização



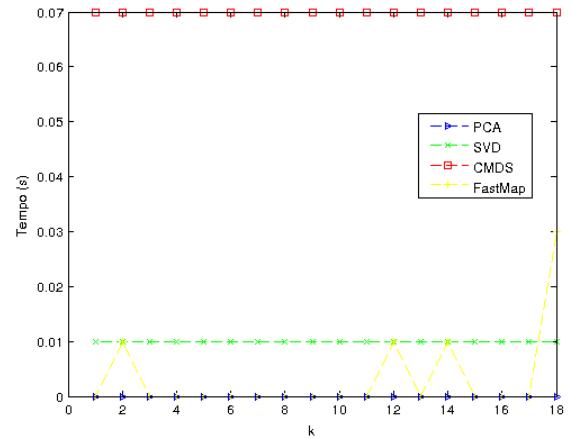
(a) Cenário \mathbf{X} e $\mathbf{D}_{\text{Euclidiana}}$



(b) Cenário \mathbf{X} e $\mathbf{D}_{\text{Manhattan}}$

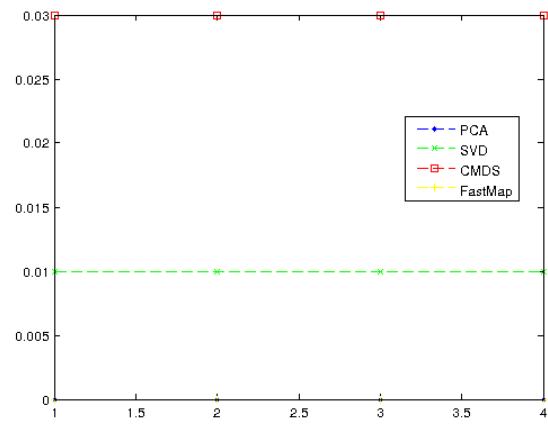


(c) Cenário \mathbf{Z} e $\mathbf{D}_{\text{Euclidiana}}$

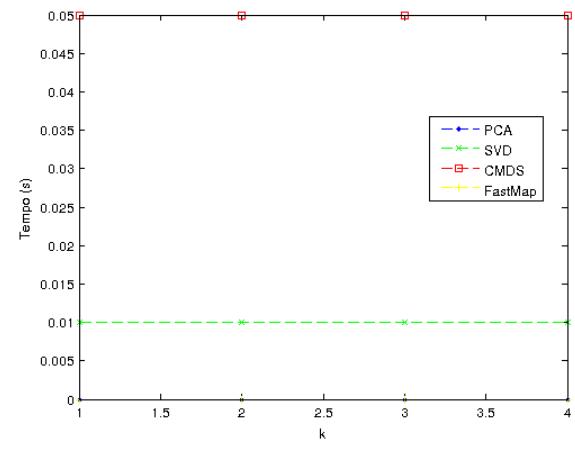


(d) Cenário \mathbf{Z} e $\mathbf{D}_{\text{Manhattan}}$

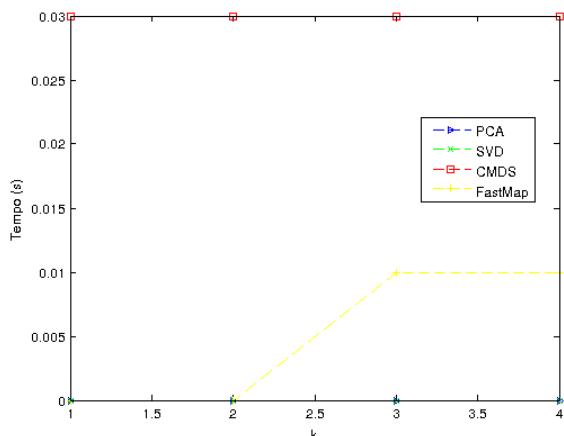
Figura B.10: Tempo de execução para base Iris - Excluindo métodos baseados em otimização



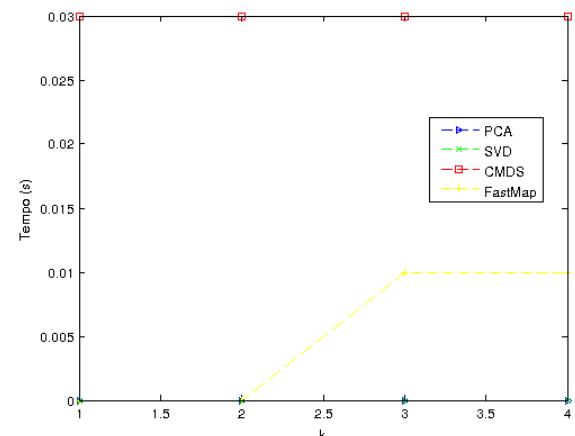
(a) Cenário \mathbf{X} e $\mathbf{D}_{\text{Euclidiana}}$



(b) Cenário \mathbf{X} e $\mathbf{D}_{\text{Manhattan}}$

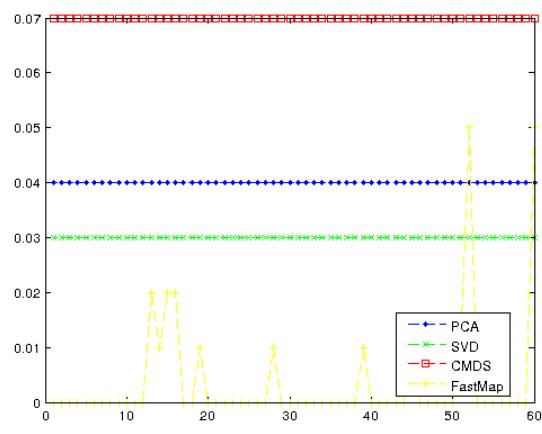


(c) Cenário \mathbf{Z} e $\mathbf{D}_{\text{Euclidiana}}$

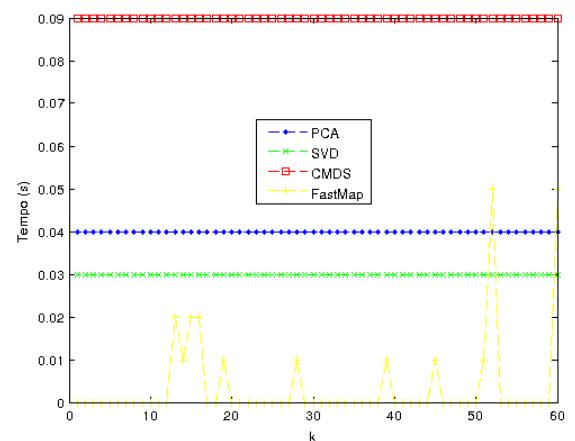


(d) Cenário \mathbf{Z} e $\mathbf{D}_{\text{Manhattan}}$

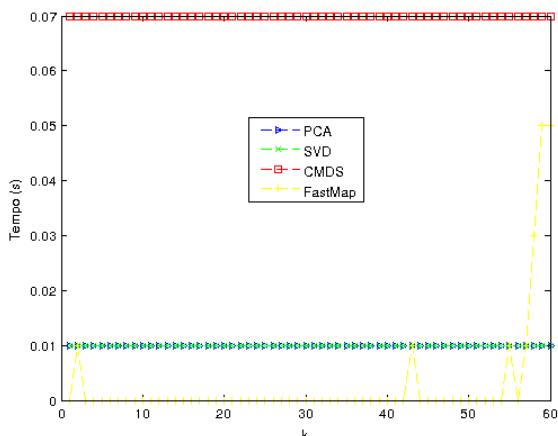
Figura B.11: Tempo de execução para base Sonar - Excluindo métodos baseados em otimização



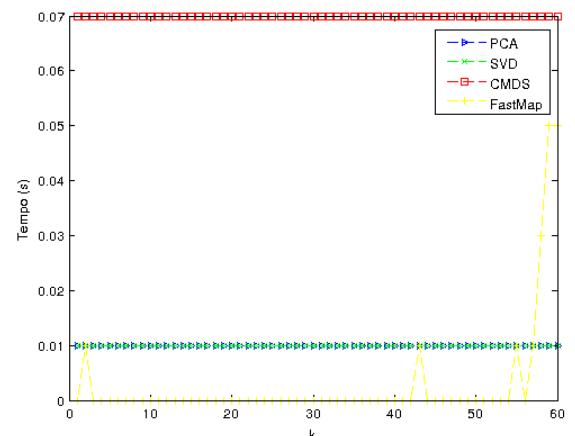
(a) Cenário \mathbf{X} e $\mathbf{D}_{\text{Euclidiana}}$



(b) Cenário \mathbf{X} e $\mathbf{D}_{\text{Manhattan}}$

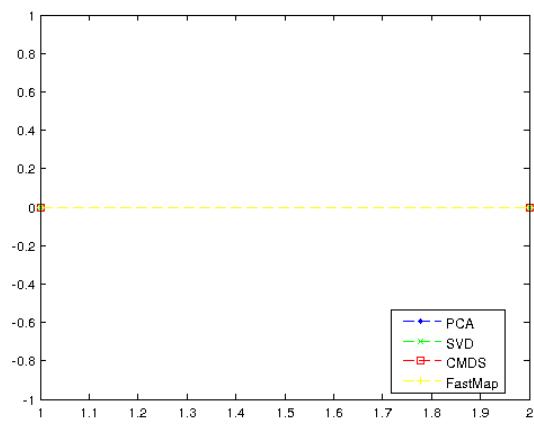


(c) Cenário \mathbf{Z} e $\mathbf{D}_{\text{Euclidiana}}$

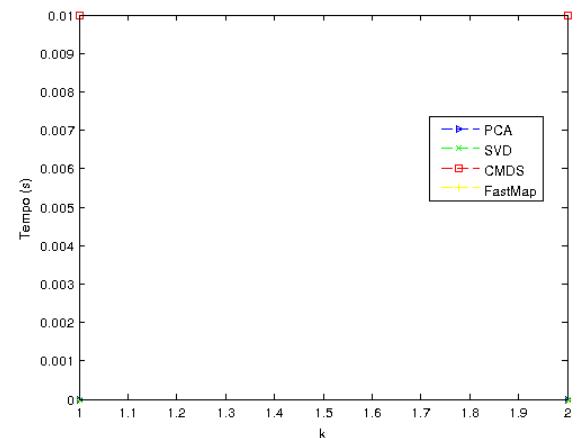


(d) Cenário \mathbf{Z} e $\mathbf{D}_{\text{Manhattan}}$

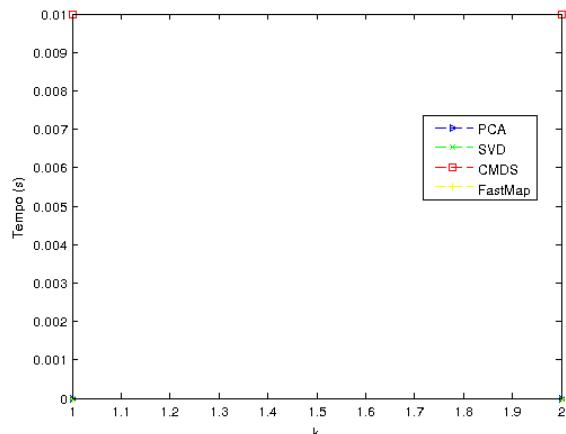
Figura B.12: Tempo de execução para base Espiral - Excluindo métodos baseados em otimização



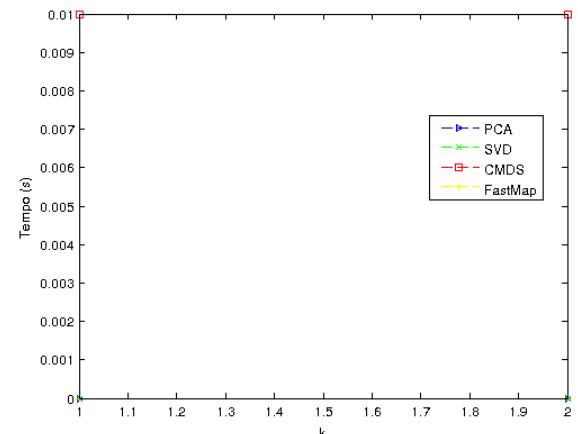
(a) Cenário \mathbf{X} e $\mathbf{D}_{\text{Euclidian}}$



(b) Cenário \mathbf{X} e $\mathbf{D}_{\text{Manhattan}}$

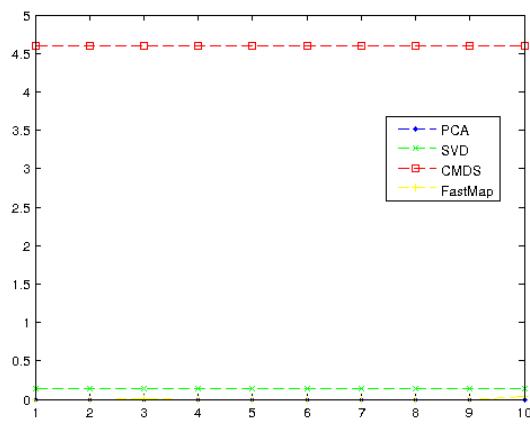


(c) Cenário \mathbf{Z} e $\mathbf{D}_{\text{Euclidian}}$

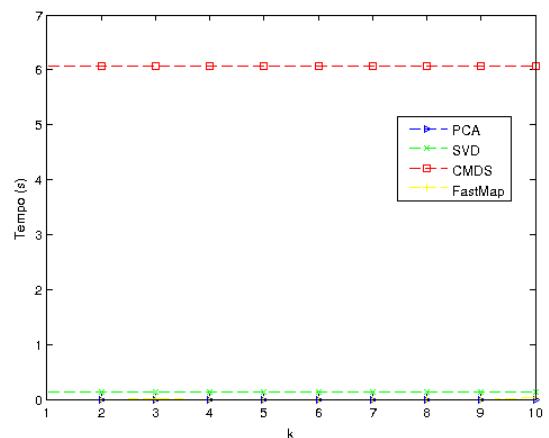


(d) Cenário \mathbf{Z} e $\mathbf{D}_{\text{Manhattan}}$

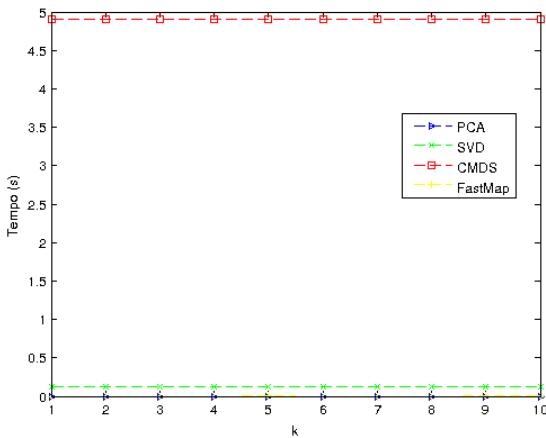
Figura B.13: Tempo de execução para base Vogal - Excluindo métodos baseados em otimização



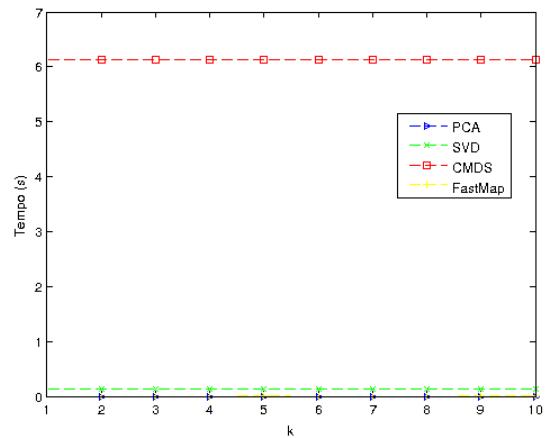
(a) Cenário \mathbf{X} e $D_{\text{Euclidiana}}$



(b) Cenário \mathbf{X} e $D_{\text{Manhattan}}$

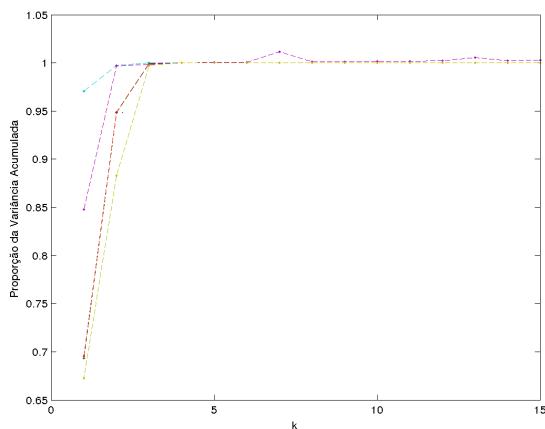


(c) Cenário \mathbf{Z} e $D_{\text{Euclidiana}}$

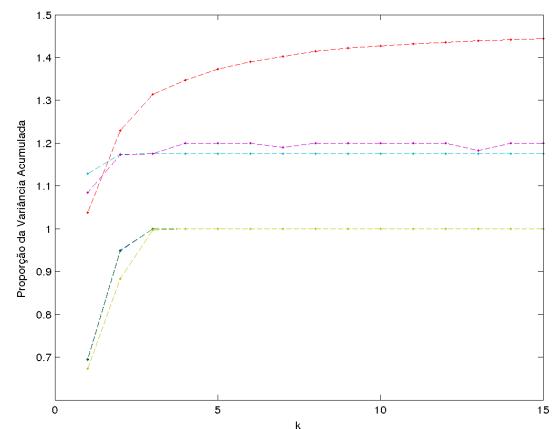


(d) Cenário \mathbf{Z} e $D_{\text{Manhattan}}$

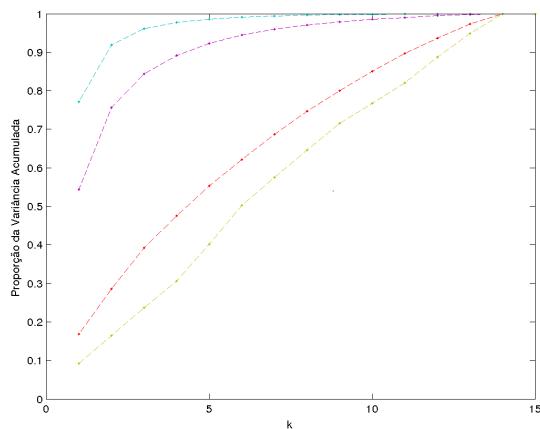
Figura B.14: Proporção da variância acumulada para base Triagem de Crédito (Legenda Fig. B.1)



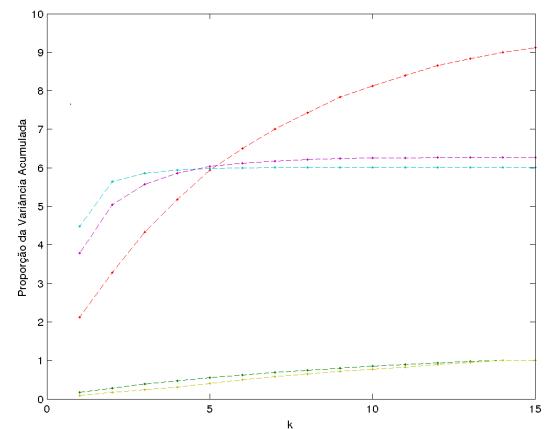
(a) Cenário **X** e $\mathbf{D}_{\text{Euclidiana}}$



(b) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$

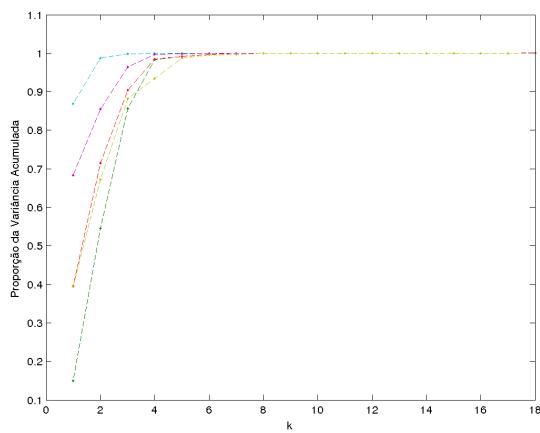


(c) Cenário **Z** e $\mathbf{D}_{\text{Euclidiana}}$

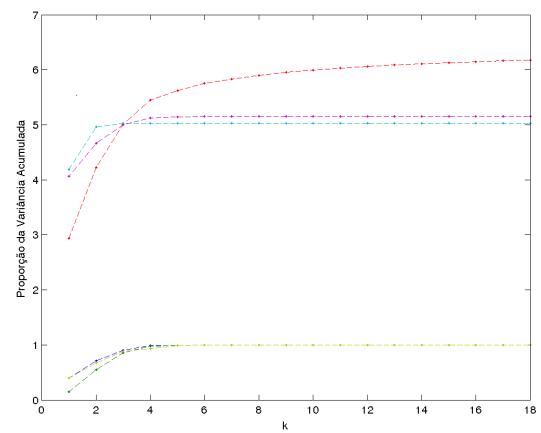


(d) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$

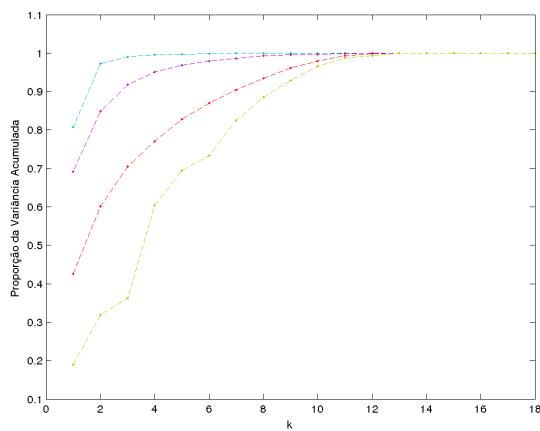
Figura B.15: Proporção da variância acumulada para base Imagens (Legenda Fig. B.1)



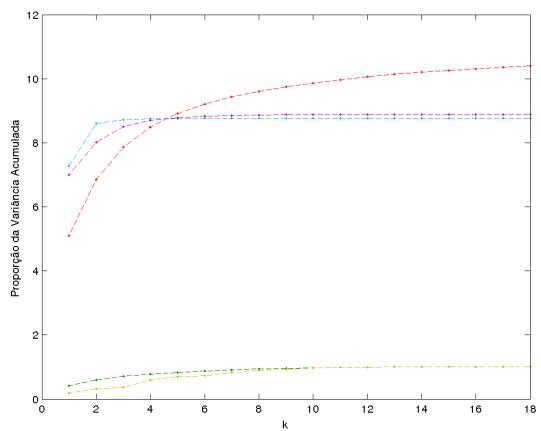
(a) Cenário \mathbf{X} e $\mathbf{D}_{\text{Euclidiana}}$



(b) Cenário \mathbf{X} e $\mathbf{D}_{\text{Manhattan}}$



(c) Cenário \mathbf{Z} e $\mathbf{D}_{\text{Euclidiana}}$



(d) Cenário \mathbf{Z} e $\mathbf{D}_{\text{Manhattan}}$

Figura B.16: Proporção da variância acumulada para base Iris (Legenda Fig. B.1)

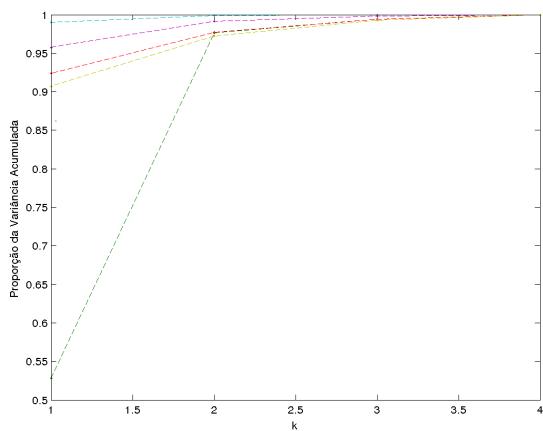
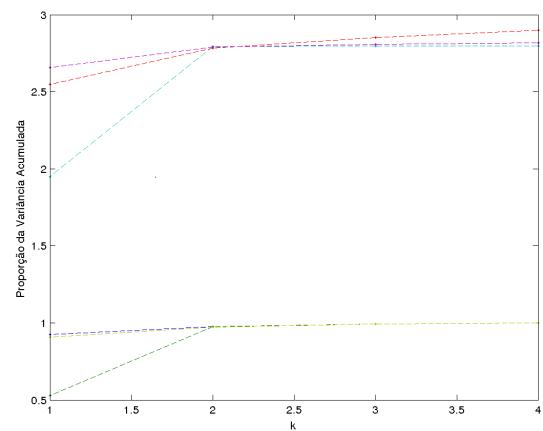
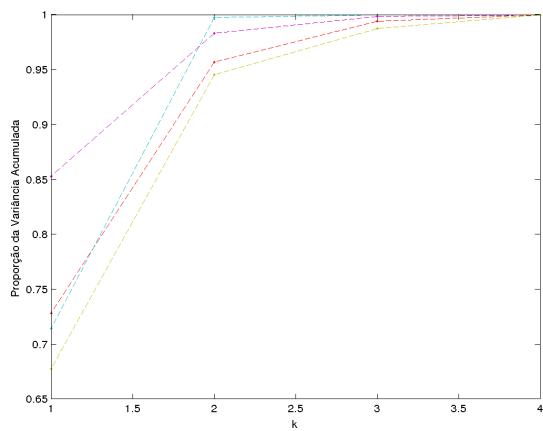
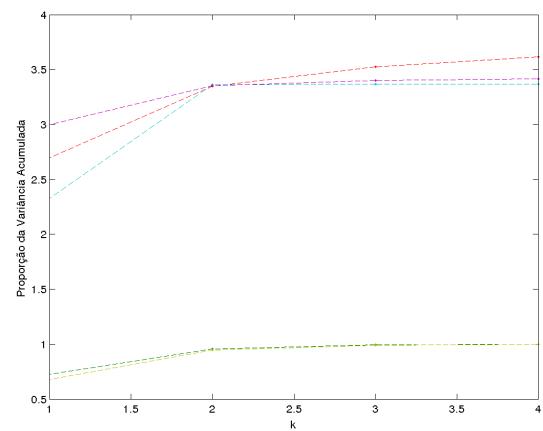
(a) Cenário **X** e $\mathbf{D}_{\text{Euclidiana}}$ (b) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$ (c) Cenário **Z** e $\mathbf{D}_{\text{Euclidiana}}$ (d) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$

Figura B.17: Proporção da variância acumulada para base Sonar (Legenda Fig. B.1)

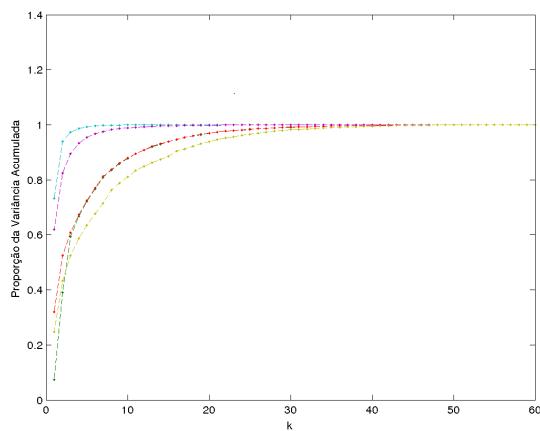
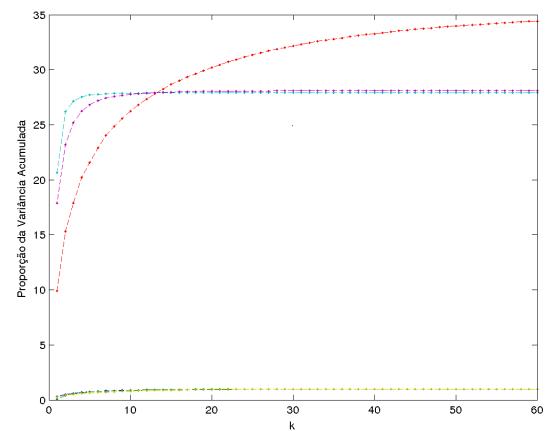
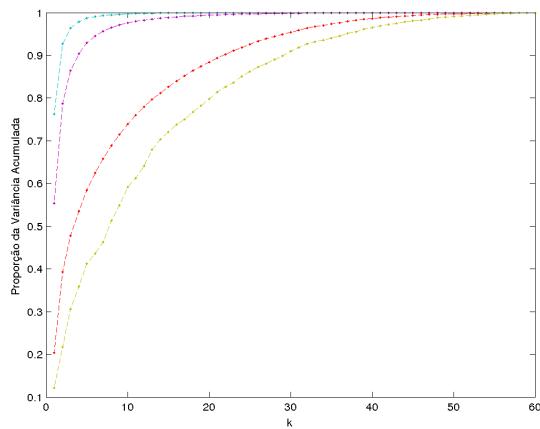
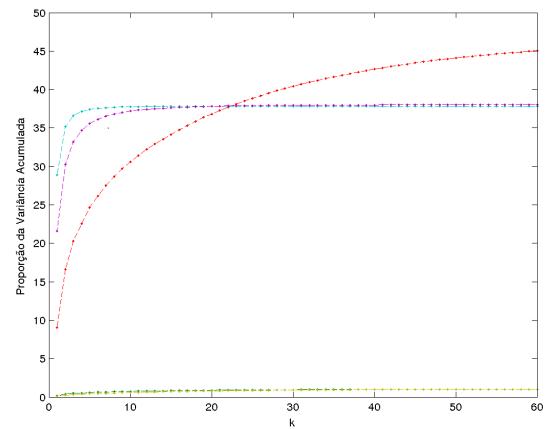
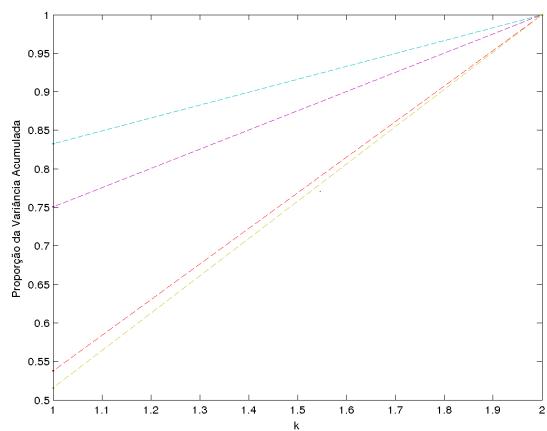
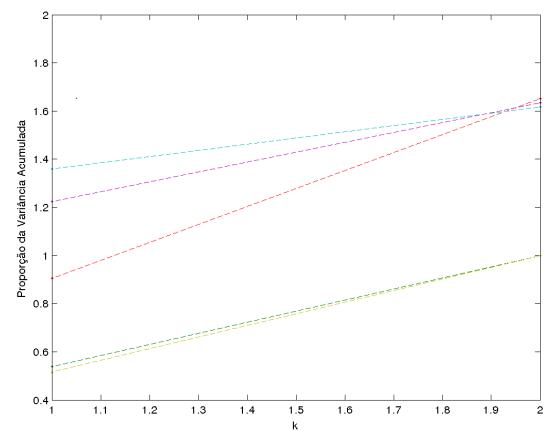
(a) Cenário **X** e **D_{Euclidian}**(b) Cenário **X** e **D_{Manhattan}**(c) Cenário **Z** e **D_{Euclidian}**(d) Cenário **Z** e **D_{Manhattan}**

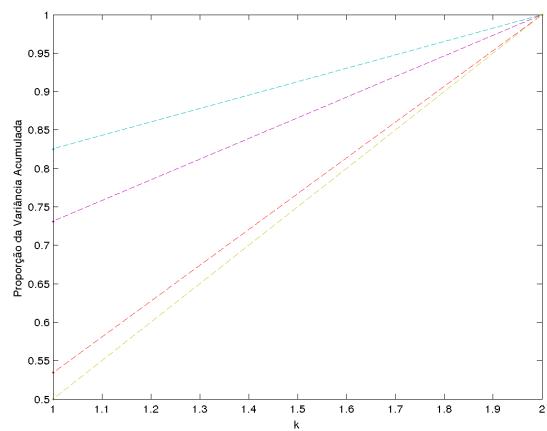
Figura B.18: Proporção da variância acumulada para base Espiral (Legenda Fig. B.1)



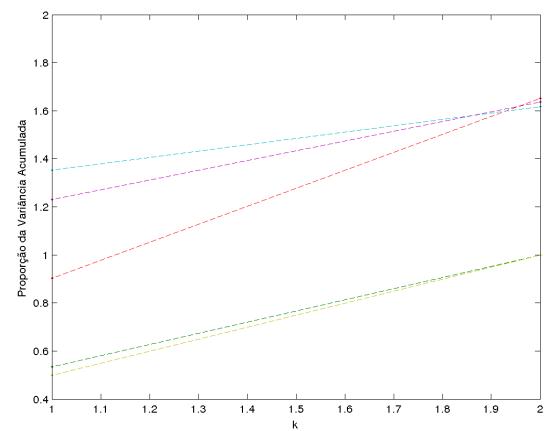
(a) Cenário X e D_{Euclidiana}



(b) Cenário X e D_{Manhattan}

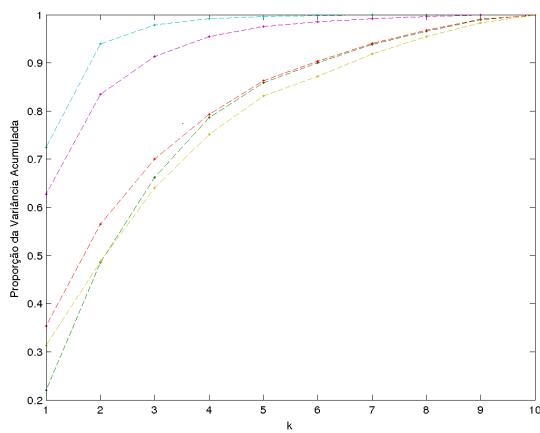


(c) Cenário Z e D_{Euclidiana}

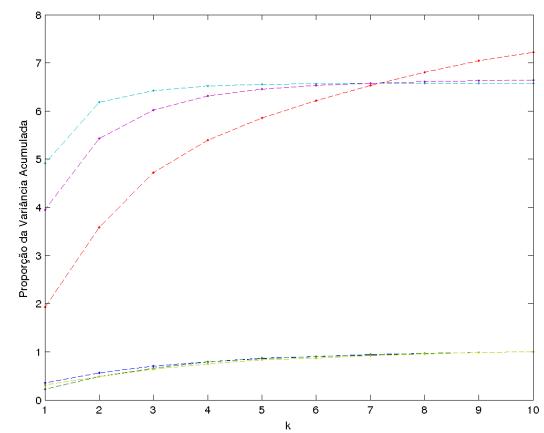


(d) Cenário Z e D_{Manhattan}

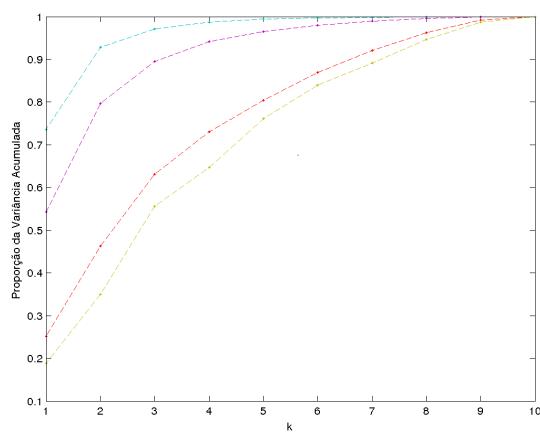
Figura B.19: Proporção da variância acumulada para base Vogal (Legenda Fig. B.1)



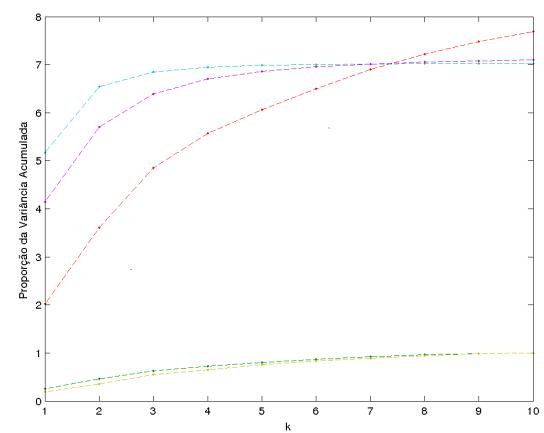
(a) Cenário **X** e $\mathbf{D}_{\text{Euclidiana}}$



(b) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$



(c) Cenário **Z** e $\mathbf{D}_{\text{Euclidiana}}$



(d) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$

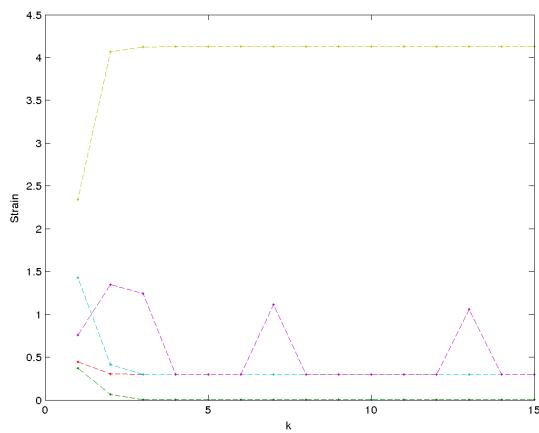
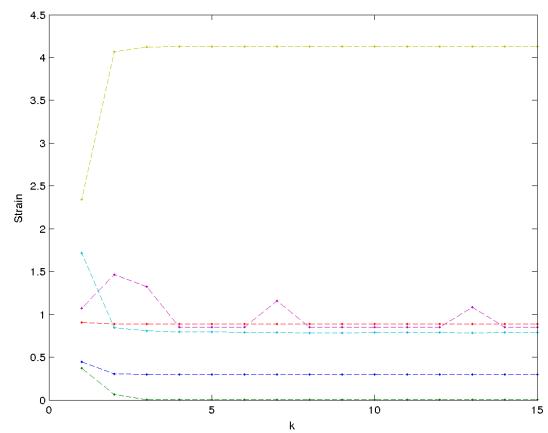
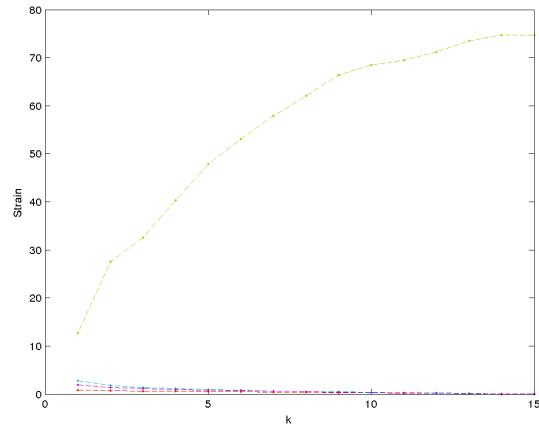
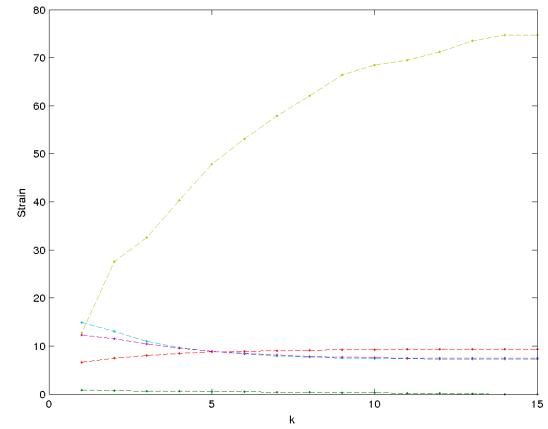
Figura B.20: *Strain* para base Triagem de Crédito (Legenda Fig. B.1)(a) Cenário **X** e $\mathbf{D}_{\text{Euclidiana}}$ (b) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$ (c) Cenário **Z** e $\mathbf{D}_{\text{Euclidiana}}$ (d) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$

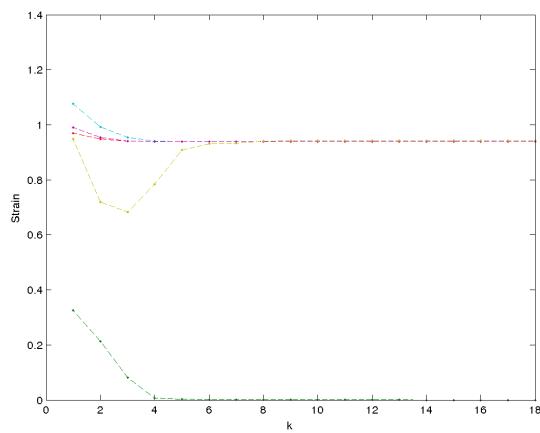
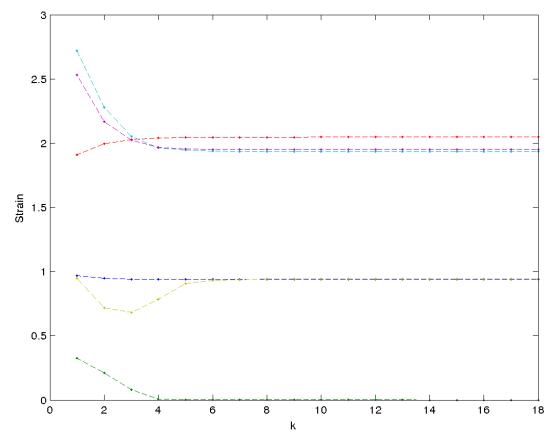
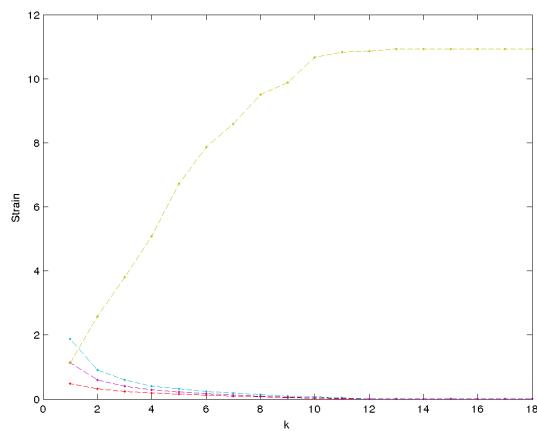
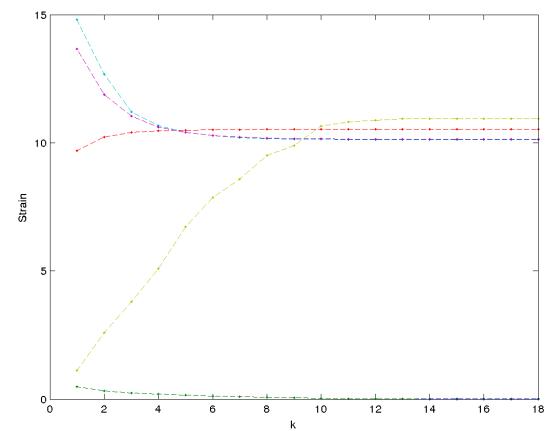
Figura B.21: *Strain* para base Imagens (Legenda Fig. B.1)(a) Cenário **X** e $\mathbf{D}_{\text{Euclidian}}$ (b) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$ (c) Cenário **Z** e $\mathbf{D}_{\text{Euclidian}}$ (d) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$

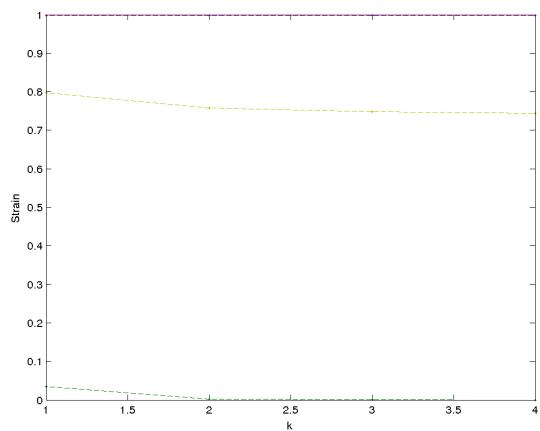
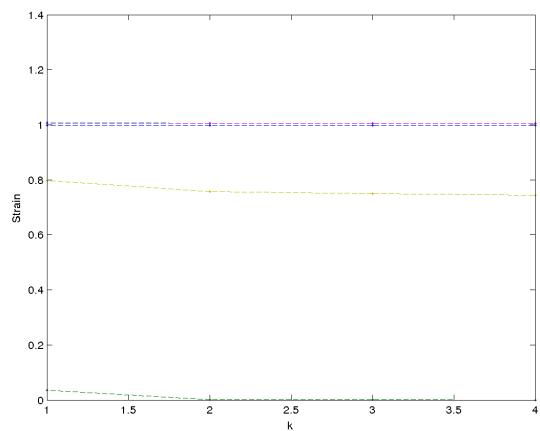
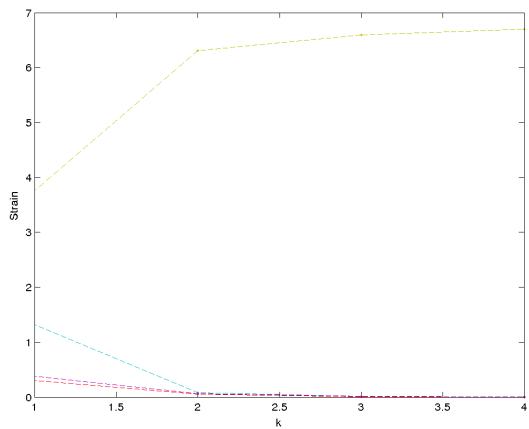
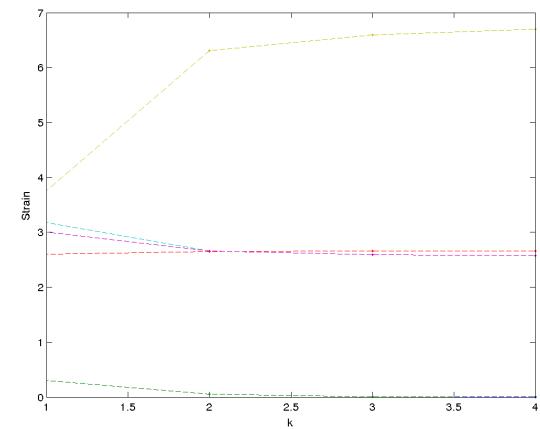
Figura B.22: *Strain* para base Iris (Legenda Fig. B.1)(a) Cenário **X** e $\mathbf{D}_{\text{Euclidiana}}$ (b) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$ (c) Cenário **Z** e $\mathbf{D}_{\text{Euclidiana}}$ (d) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$

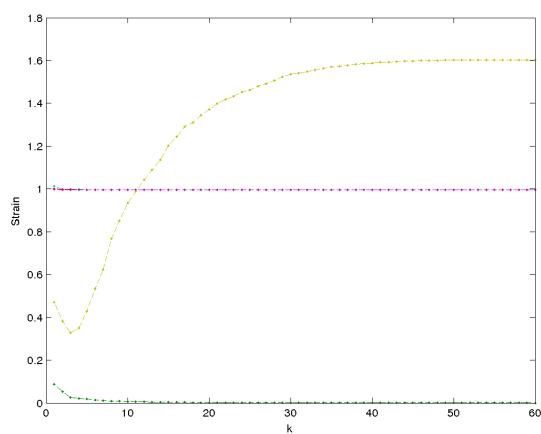
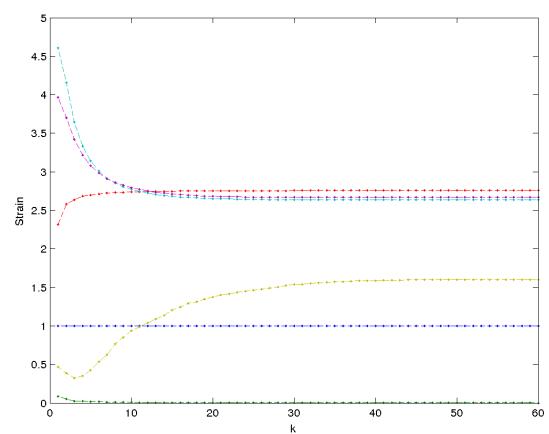
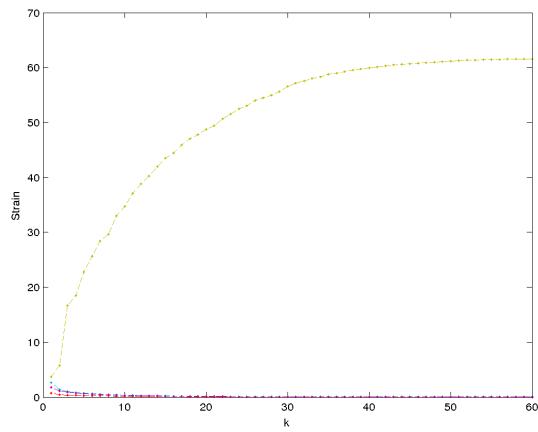
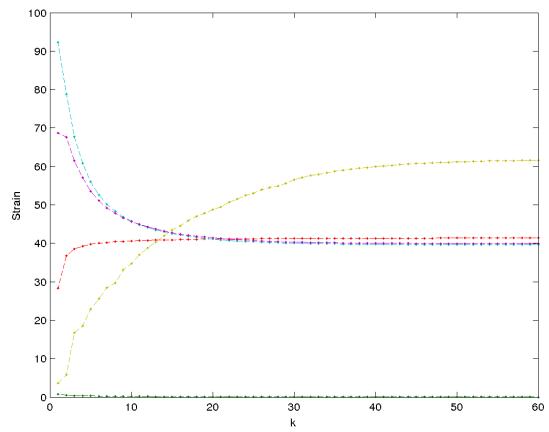
Figura B.23: *Strain* para base Sonar (Legenda Fig. B.1)(a) Cenário **X** e $\mathbf{D}_{\text{Euclidiana}}$ (b) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$ (c) Cenário **Z** e $\mathbf{D}_{\text{Euclidiana}}$ (d) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$

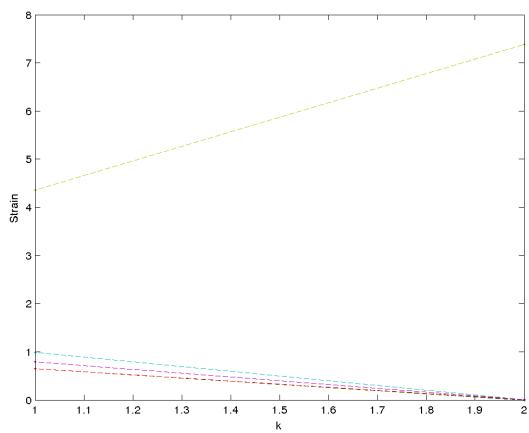
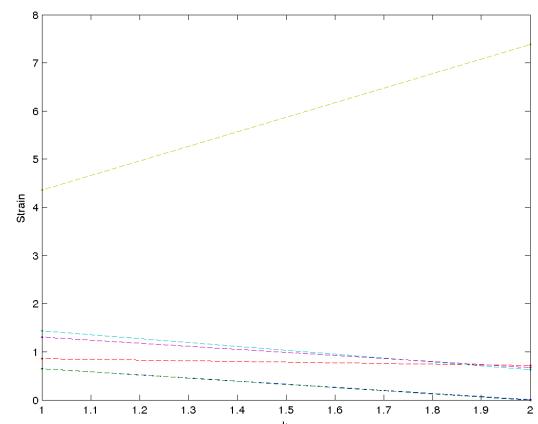
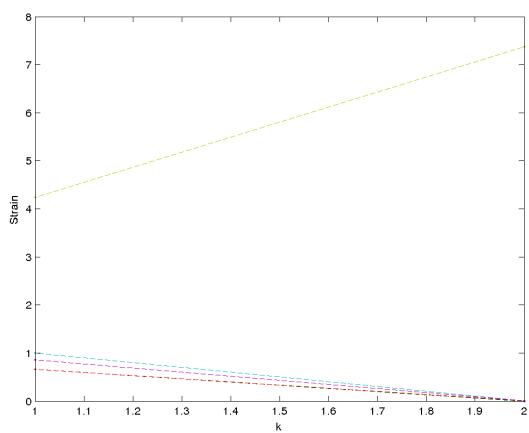
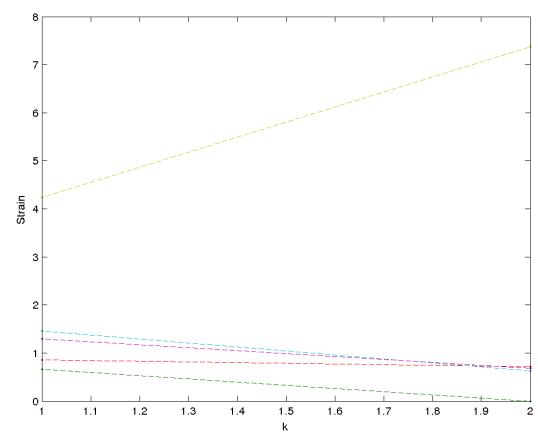
Figura B.24: *Strain* para base Espiral (Legenda Fig. B.1)(a) Cenário **X** e $\mathbf{D}_{\text{Euclidiana}}$ (b) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$ (c) Cenário **Z** e $\mathbf{D}_{\text{Euclidiana}}$ (d) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$

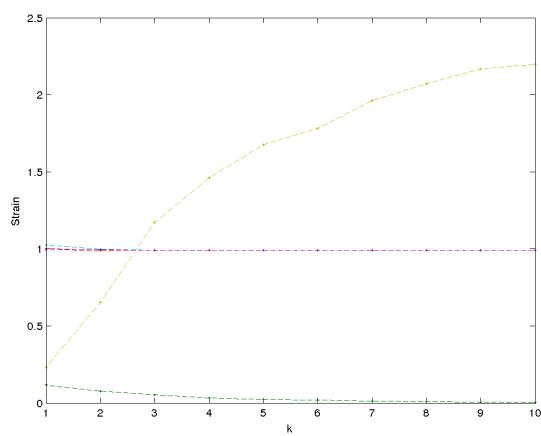
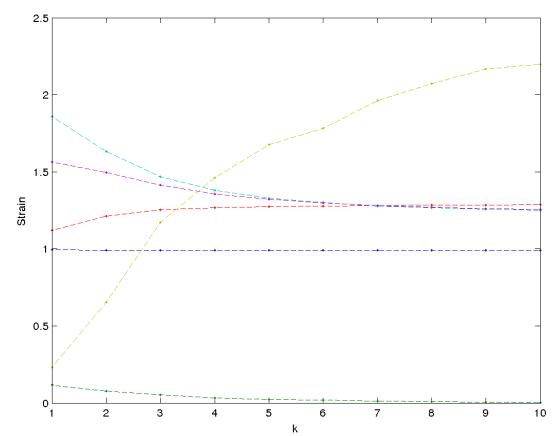
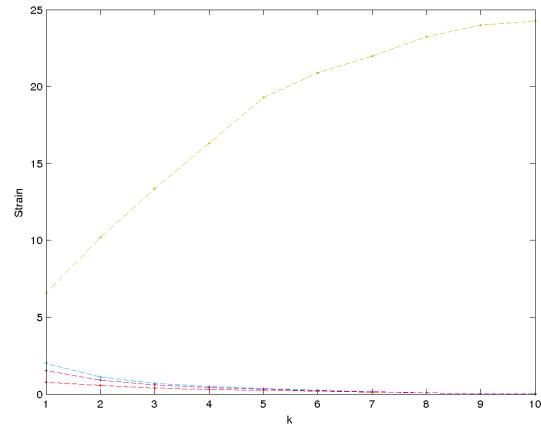
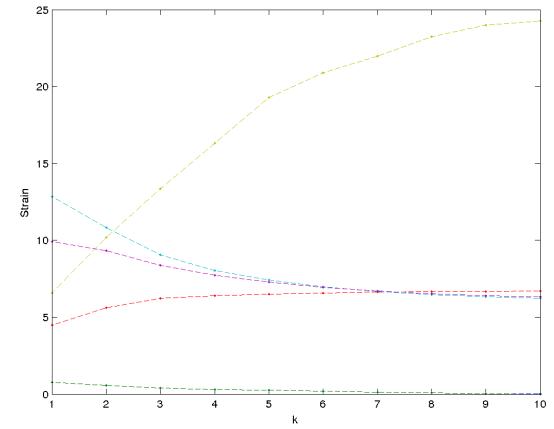
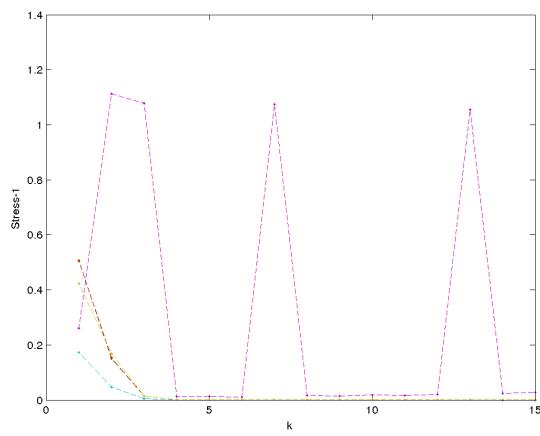
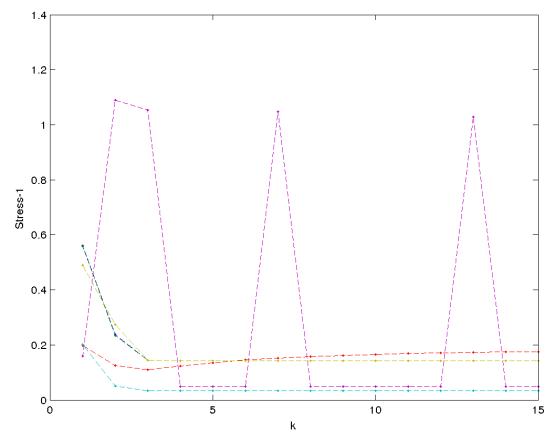
Figura B.25: *Strain* para base Vogal (Legenda Fig. B.1)(a) Cenário **X** e $\mathbf{D}_{\text{Euclidiana}}$ (b) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$ (c) Cenário **Z** e $\mathbf{D}_{\text{Euclidiana}}$ (d) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$

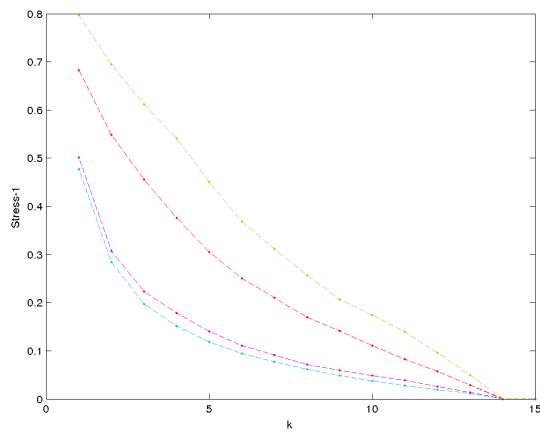
Figura B.26: Stress – 1 (formulação tradicional) para base Triagem de Crédito (Legenda Fig. B.1)



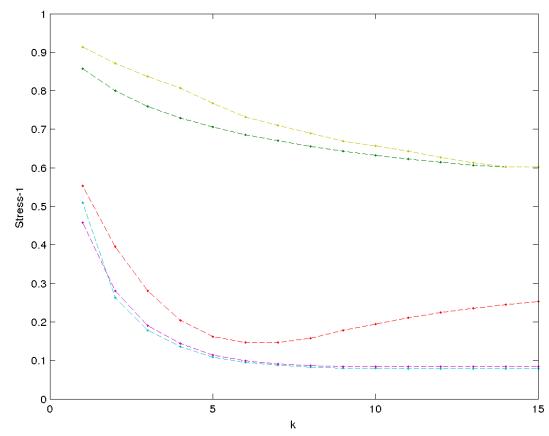
(a) Cenário **X** e **D_{Euclidiana}**



(b) Cenário **X** e **D_{Manhattan}**

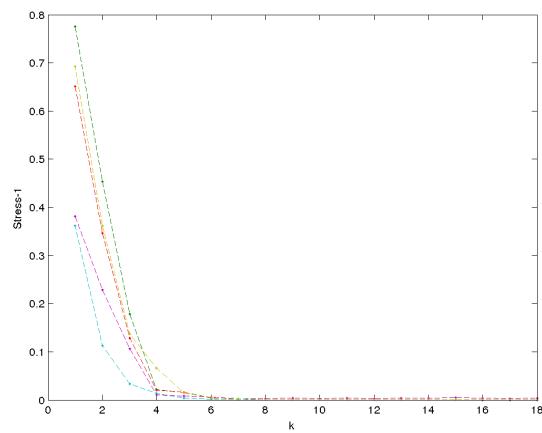


(c) Cenário **Z** e **D_{Euclidiana}**

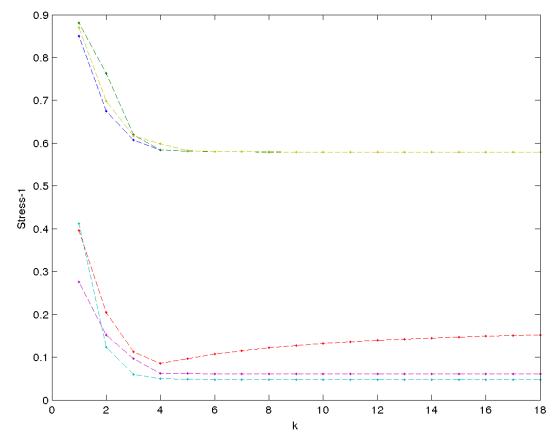


(d) Cenário **Z** e **D_{Manhattan}**

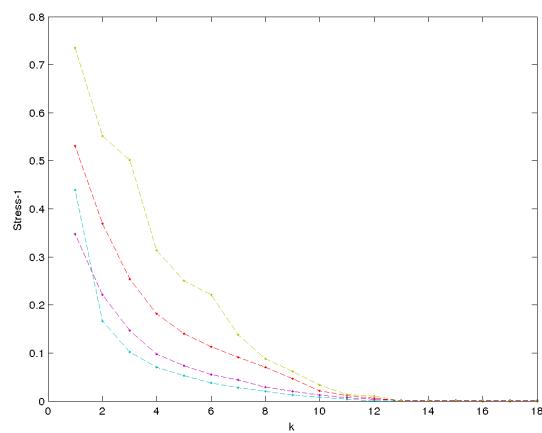
Figura B.27: *Stress - 1* (formulação tradicional) para base Imagens (Legenda Fig. B.1)



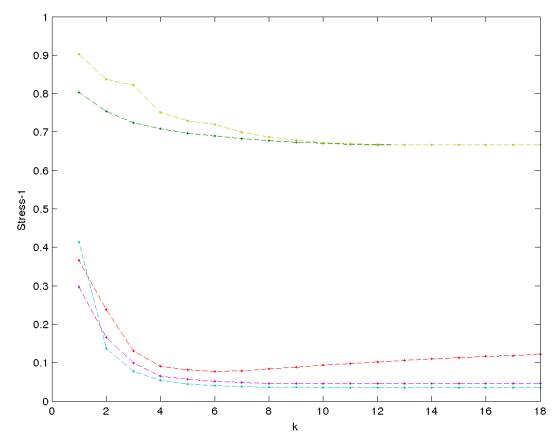
(a) Cenário **X** e $\mathbf{D}_{\text{Euclidian}}$



(b) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$

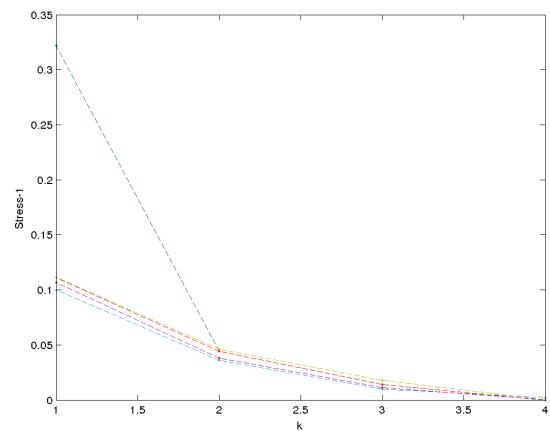


(c) Cenário **Z** e $\mathbf{D}_{\text{Euclidian}}$

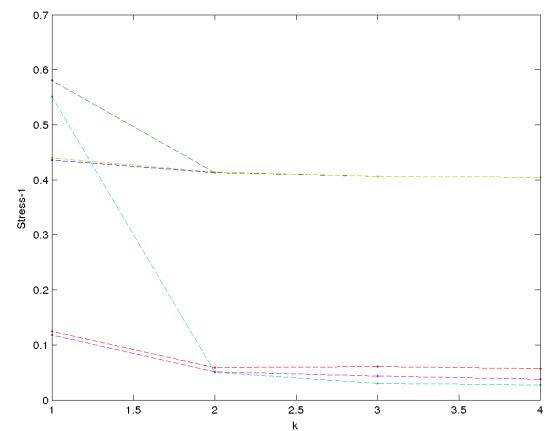


(d) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$

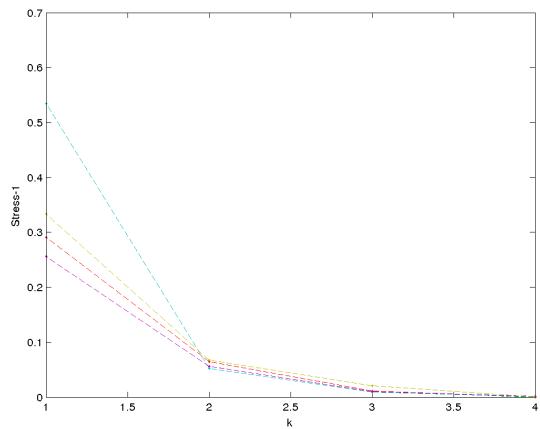
Figura B.28: $Stress - 1$ (formulação tradicional) para base Iris (Legenda Fig. B.1)



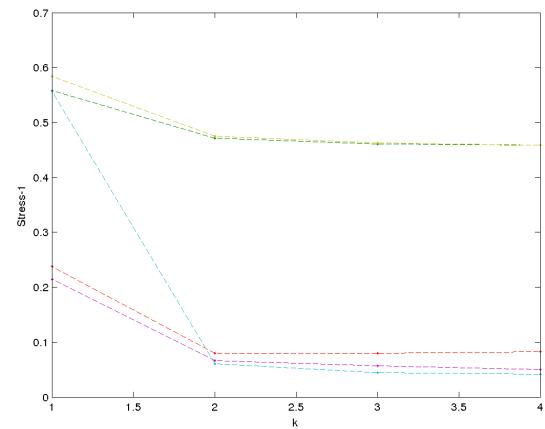
(a) Cenário **X** e $\mathbf{D}_{\text{Euclidian}}$



(b) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$

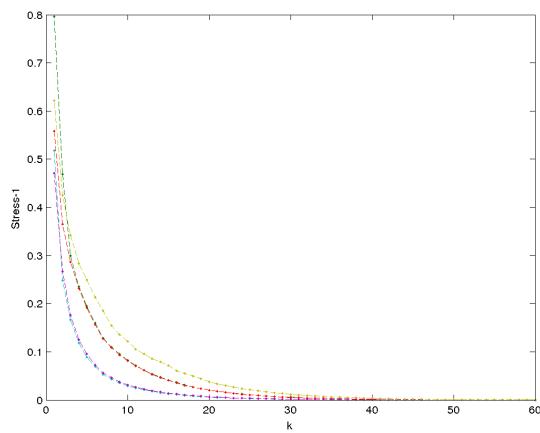


(c) Cenário **Z** e $\mathbf{D}_{\text{Euclidian}}$

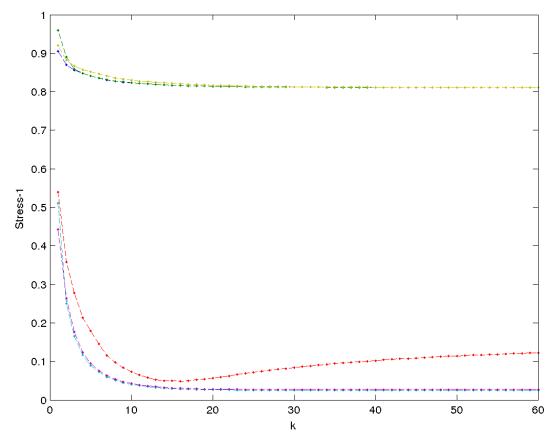


(d) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$

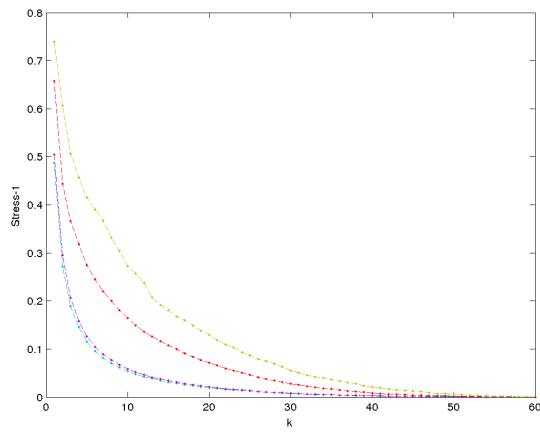
Figura B.29: *Stress – 1* (formulação tradicional) para base Sonar (Legenda Fig. B.1)



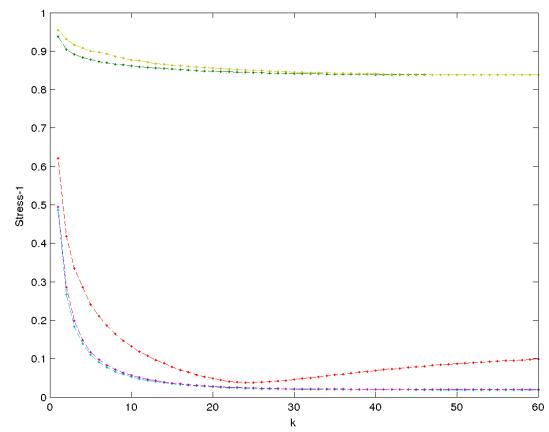
(a) Cenário **X** e $\mathbf{D}_{\text{Euclidian}}$



(b) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$

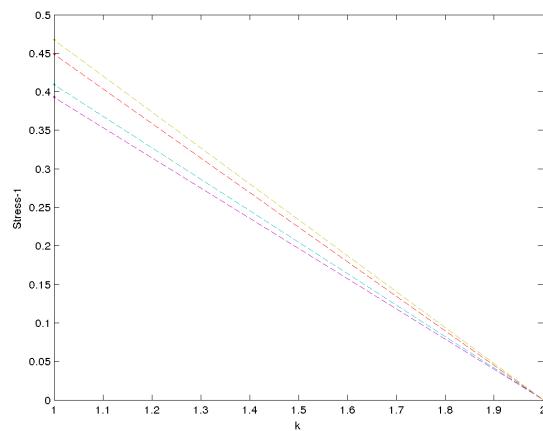


(c) Cenário **Z** e $\mathbf{D}_{\text{Euclidian}}$

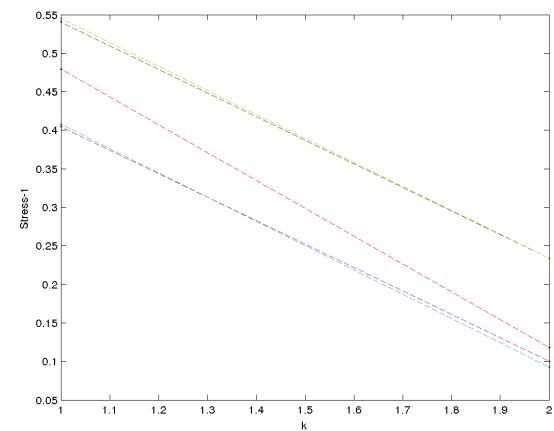


(d) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$

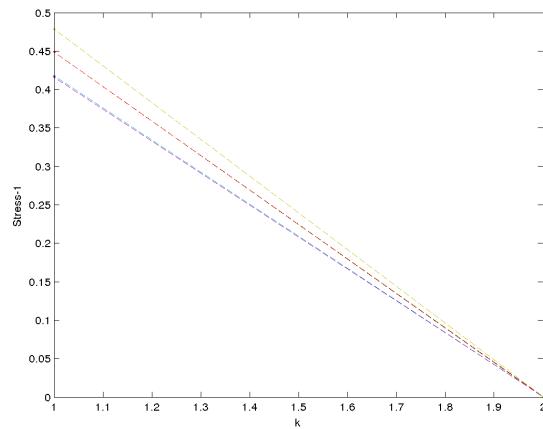
Figura B.30: $Stress - 1$ (formulação tradicional) para base Espiral (Legenda Fig. B.1)



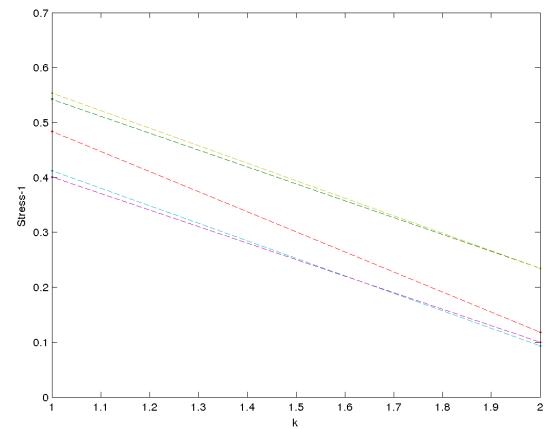
(a) Cenário **X** e $\mathbf{D}_{\text{Euclidian}}$



(b) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$



(c) Cenário **Z** e $\mathbf{D}_{\text{Euclidian}}$



(d) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$

Figura B.31: Stress – 1 (formulação tradicional) para base Vogal (Legenda Fig. B.1)

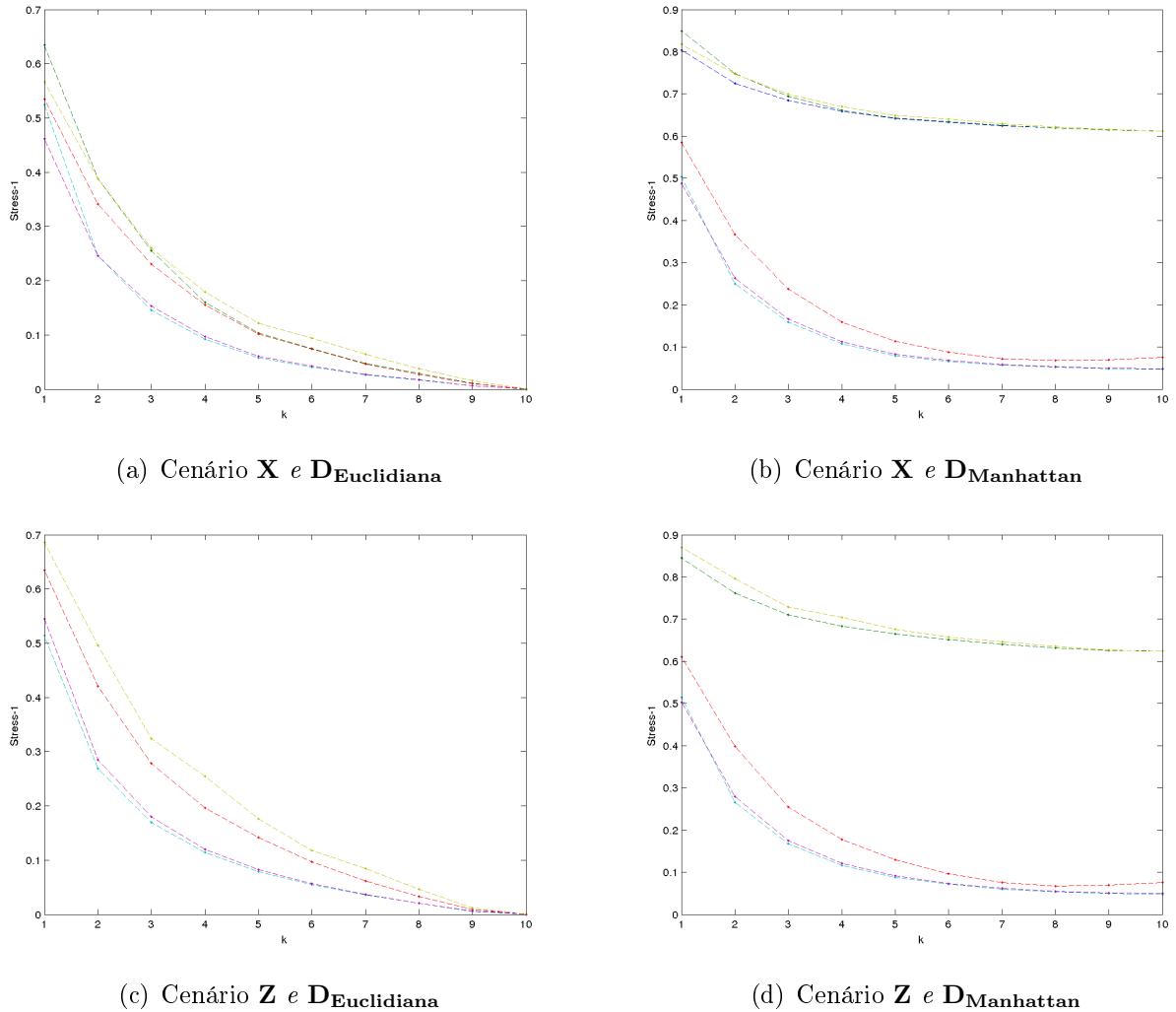


Figura B.32: Stress – 1 (formulação alternativa) para base Triagem de Crédito (Legenda Fig. B.1)

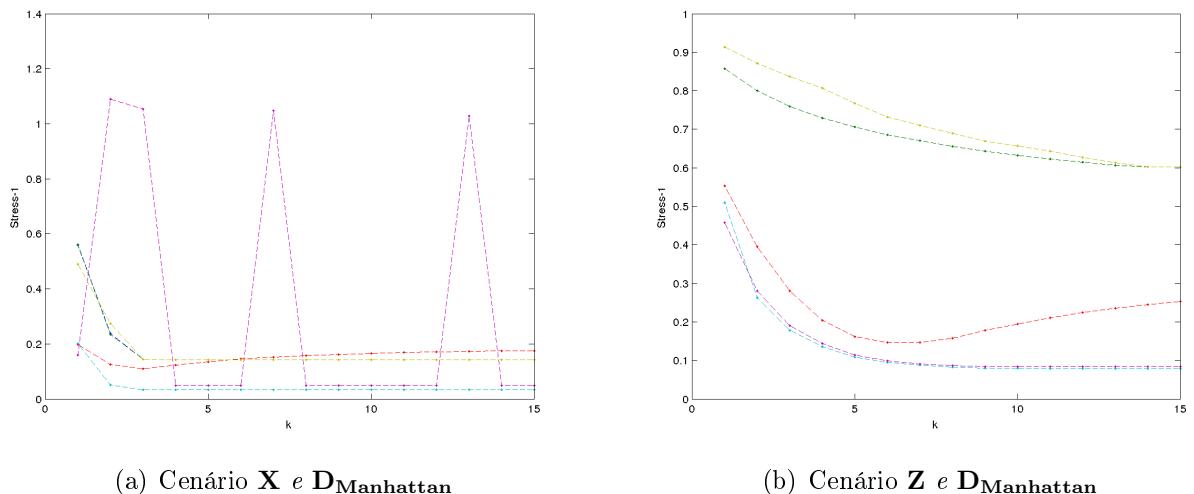


Figura B.33: Stress – 1 (formulação alternativa) para base Imagens (Legenda Fig. B.1)

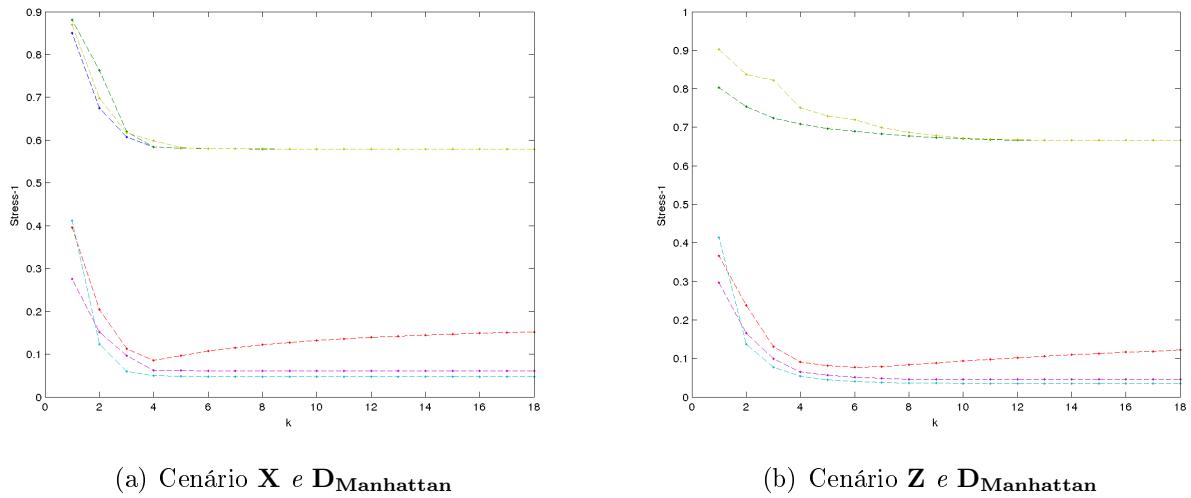


Figura B.34: Stress – 1 (formulação alternativa) para base Iris (Legenda Fig. B.1)

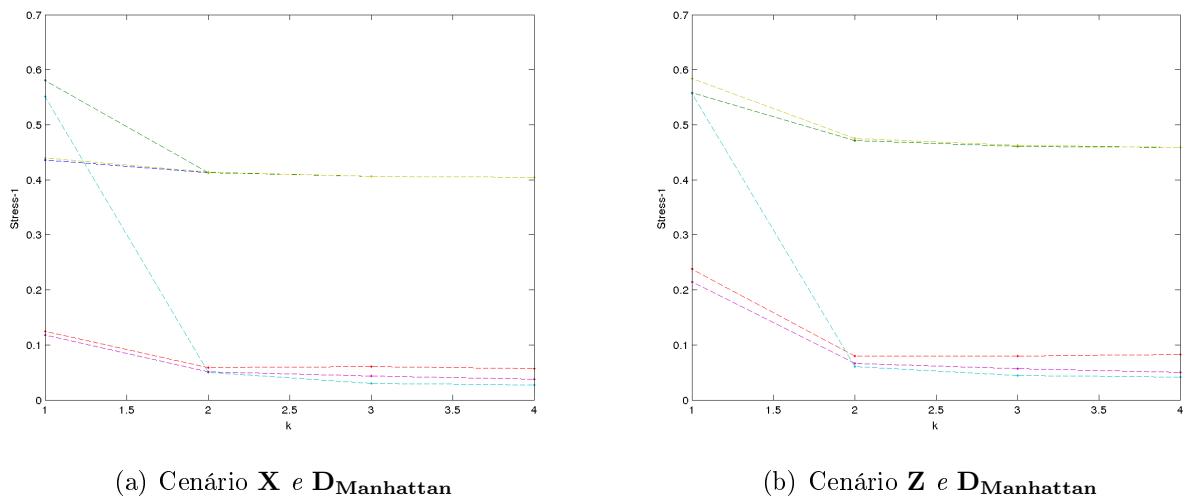


Figura B.35: Stress – 1 (formulação alternativa) para base Sonar (Legenda Fig. B.1)

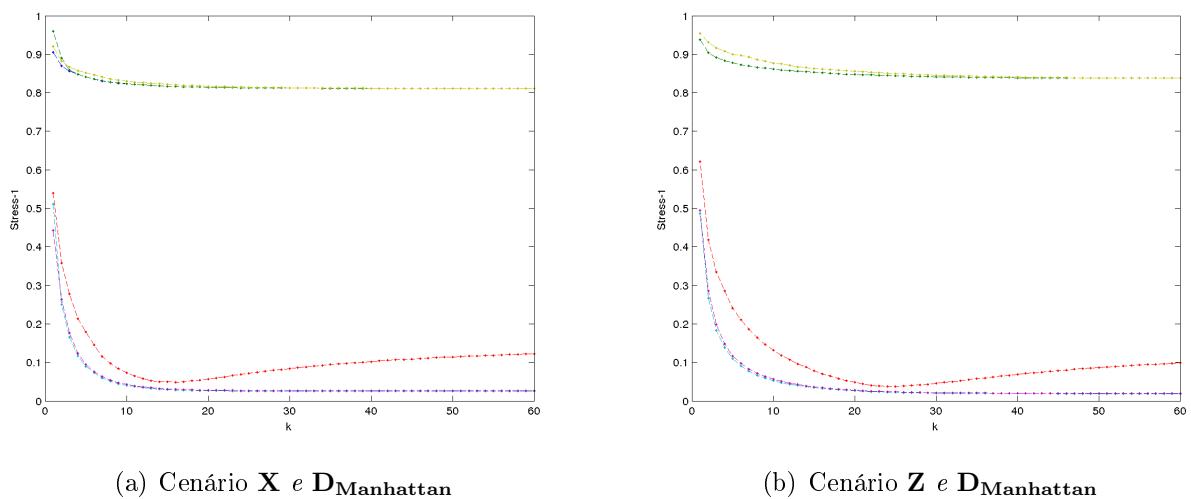
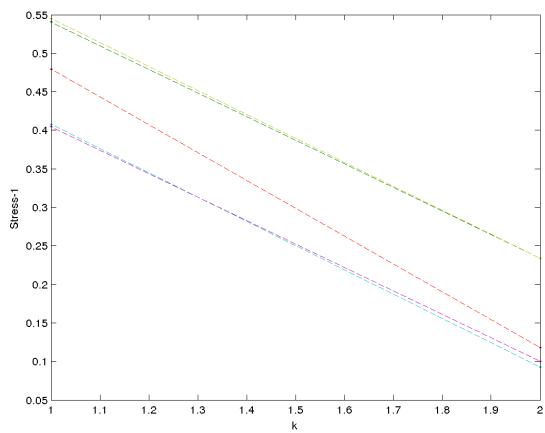
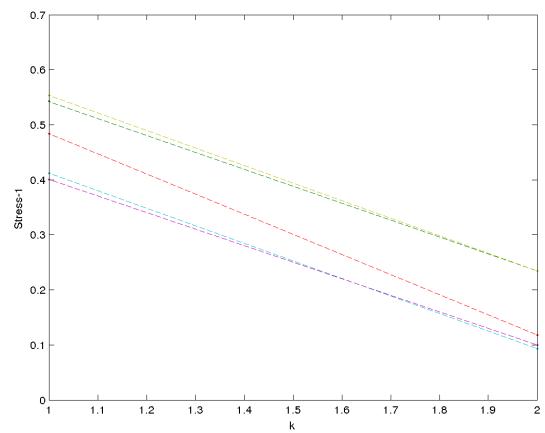


Figura B.36: *Stress – 1* (formulação alternativa) para base Espiral (Legenda Fig. B.1)



(a) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$



(b) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$

Figura B.37: Stress – 1 (formulação alternativa) para base Vogal (Legenda Fig. B.1)

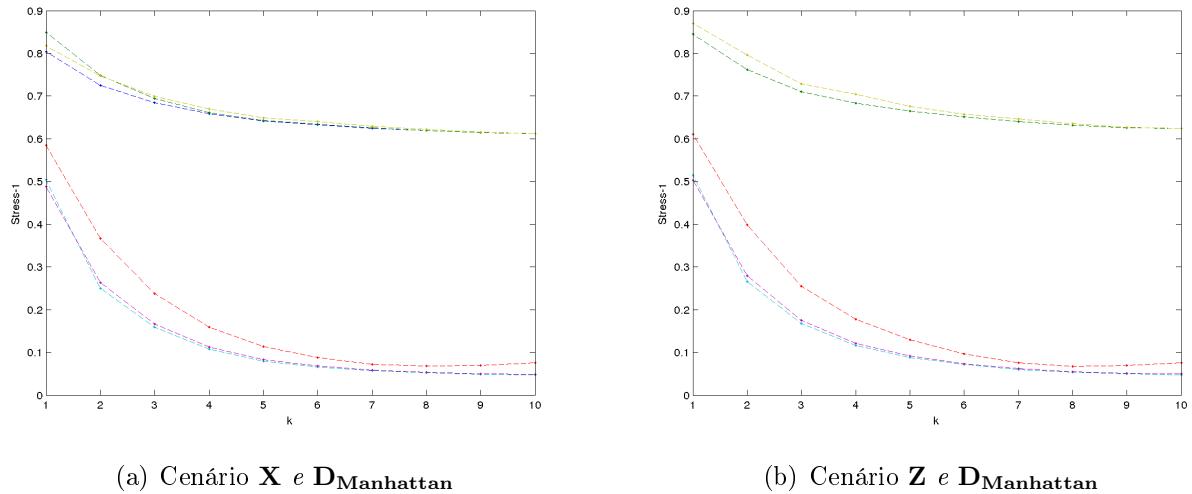


Figura B.38: Sstress (formulação tradicional) para base Triagem de Crédito (Legenda Fig. B.1)

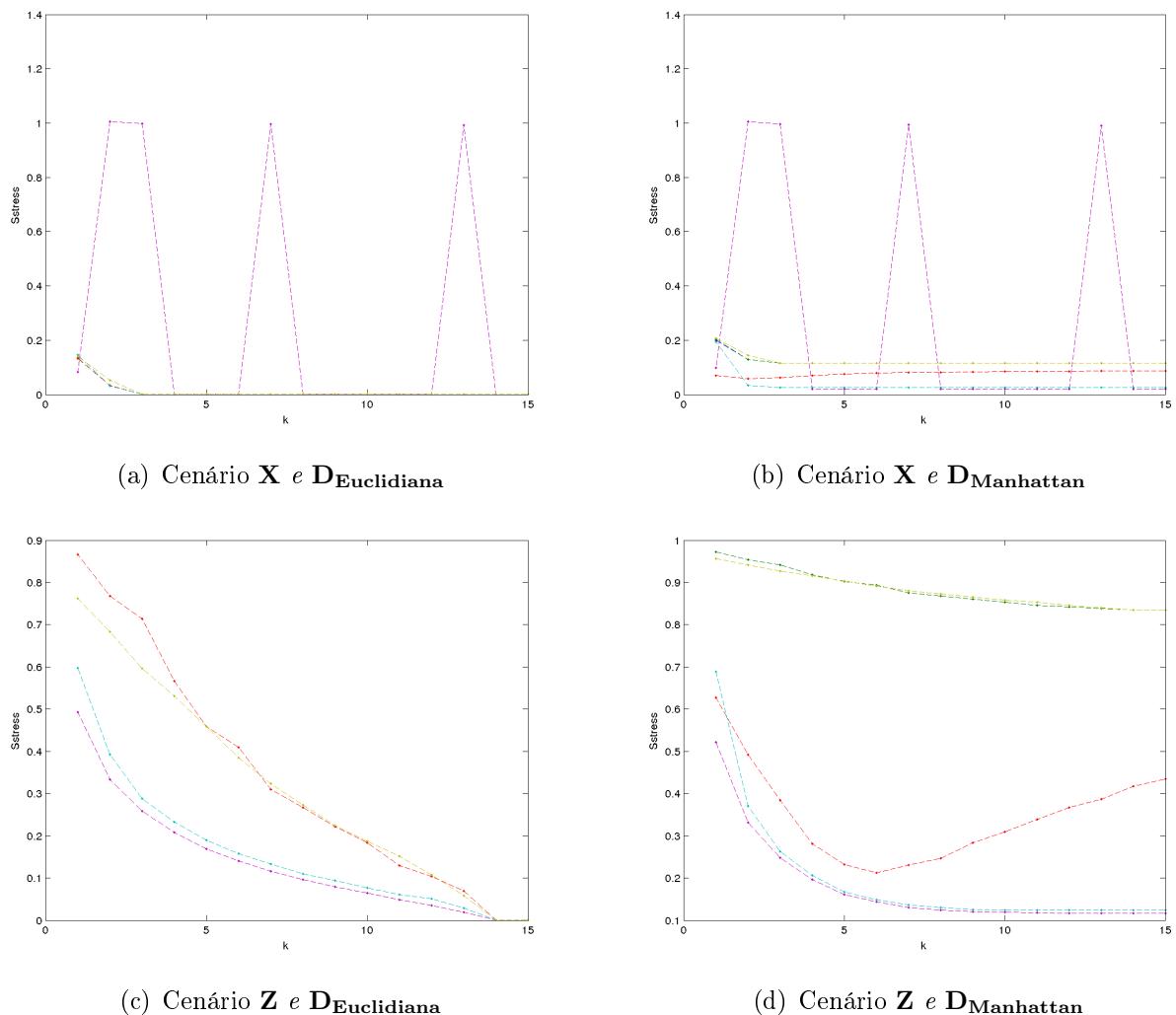
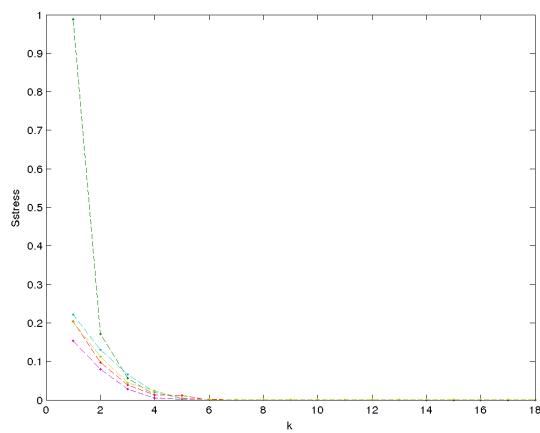
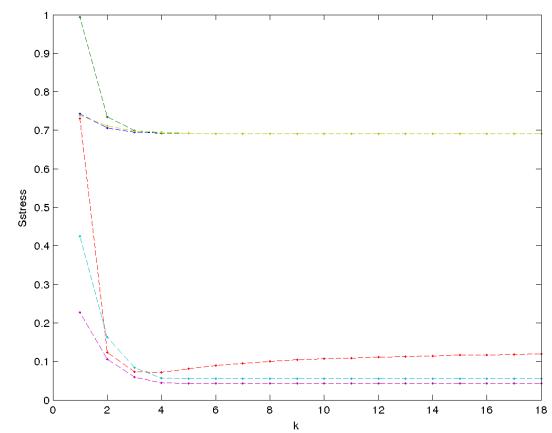


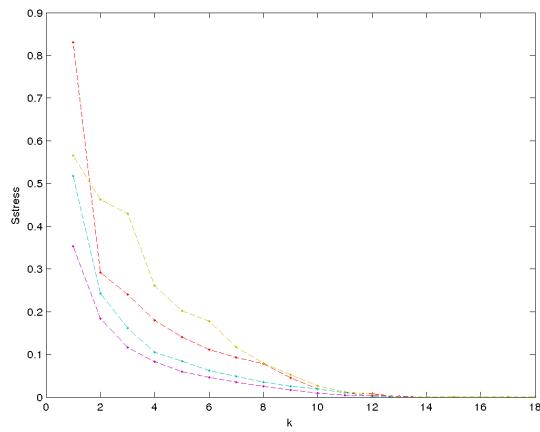
Figura B.39: *Sstress* (formulação tradicional) para base Imagens (Legenda Fig. B.1)



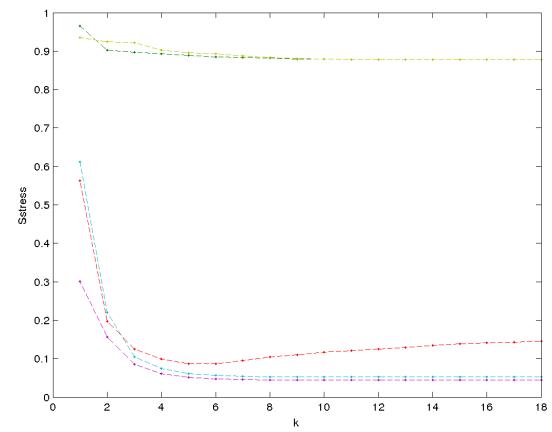
(a) Cenário **X** e $\mathbf{D}_{\text{Euclidiana}}$



(b) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$

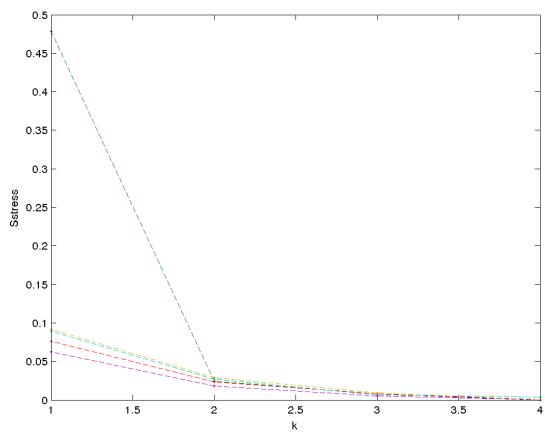


(c) Cenário **Z** e $\mathbf{D}_{\text{Euclidiana}}$

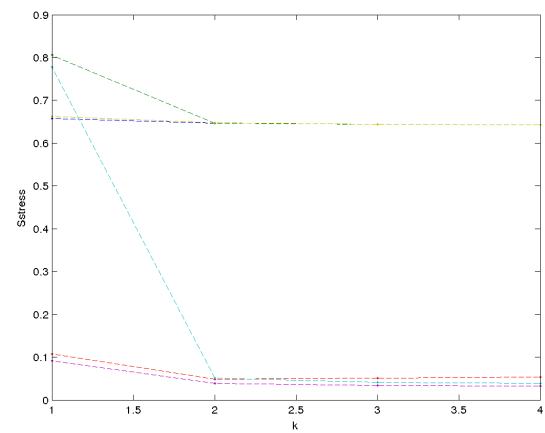


(d) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$

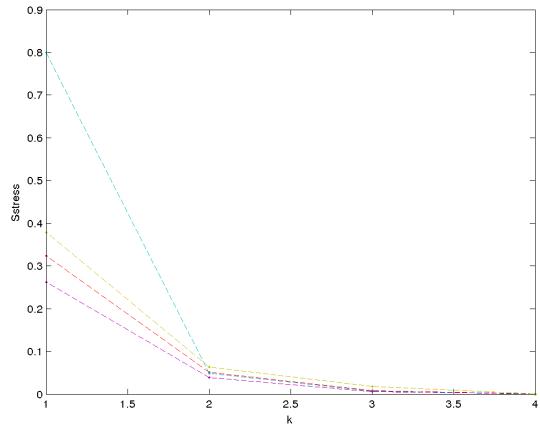
Figura B.40: *Sstress* (formulação tradicional) para base Iris (Legenda Fig. B.1)



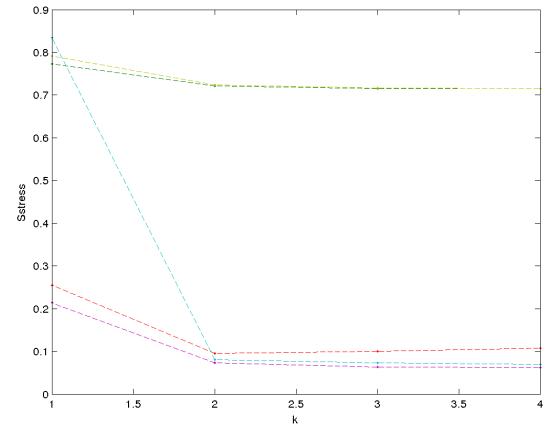
(a) Cenário **X** e $\mathbf{D}_{\text{Euclidian}}$



(b) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$

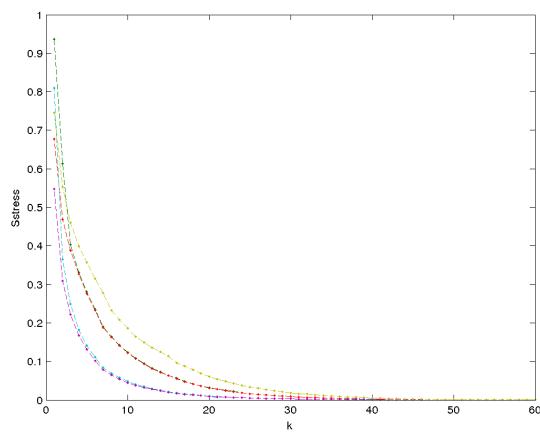


(c) Cenário **Z** e $\mathbf{D}_{\text{Euclidian}}$

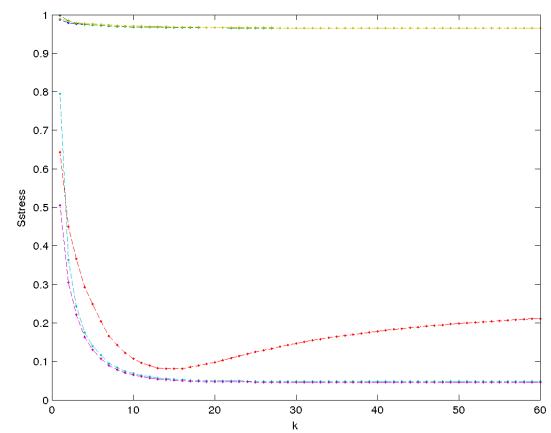


(d) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$

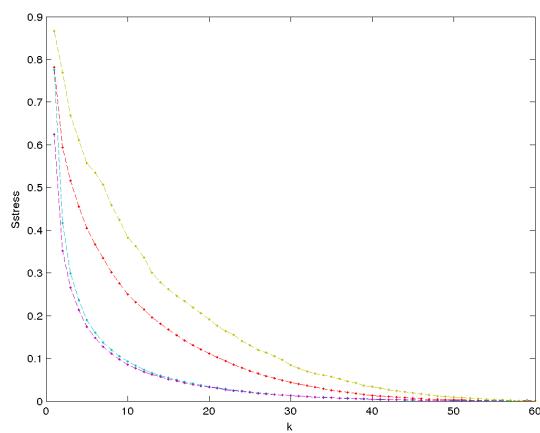
Figura B.41: *Sstress* (formulação tradicional) para base Sonar (Legenda Fig. B.1)



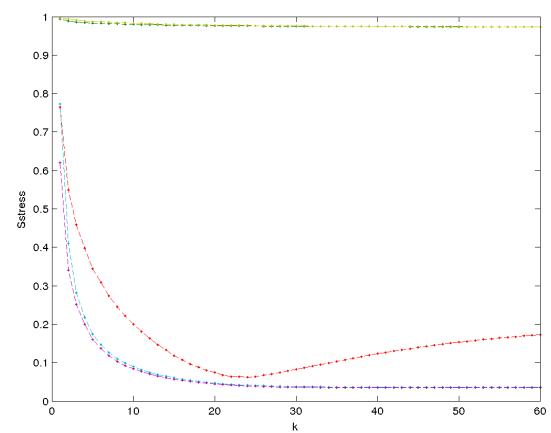
(a) Cenário **X** e $\mathbf{D}_{\text{Euclidiana}}$



(b) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$

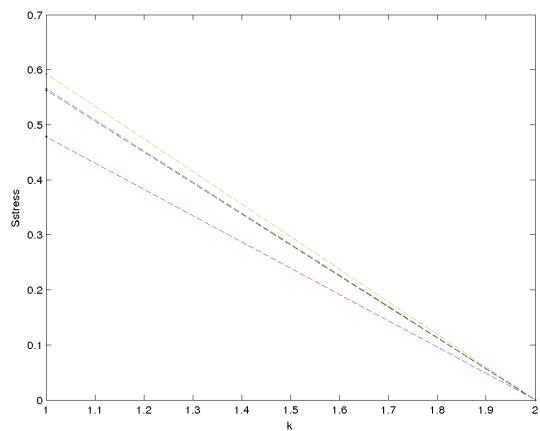


(c) Cenário **Z** e $\mathbf{D}_{\text{Euclidiana}}$

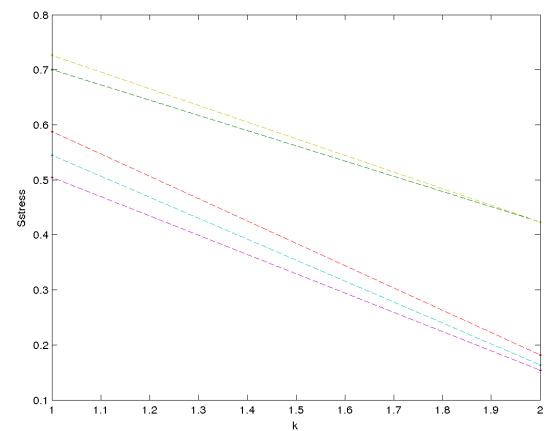


(d) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$

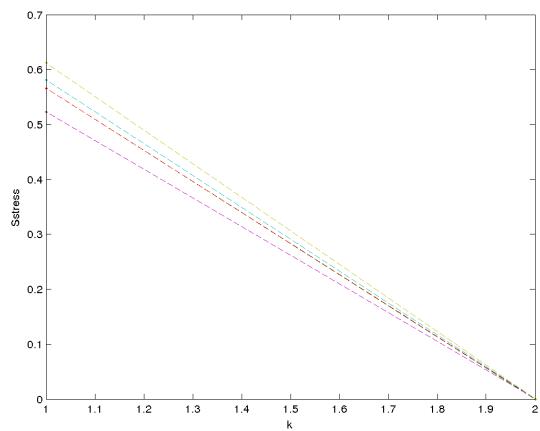
Figura B.42: *Sstress* (formulação tradicional) para base Espiral (Legenda Fig. B.1)



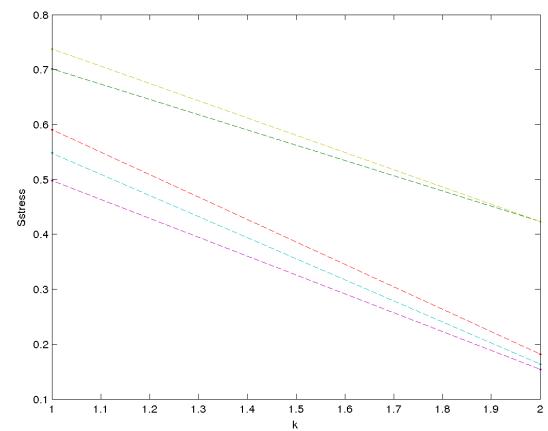
(a) Cenário **X** e **D_{Euclidiana}**



(b) Cenário **X** e **D_{Manhattan}**



(c) Cenário **Z** e **D_{Euclidiana}**



(d) Cenário **Z** e **D_{Manhattan}**

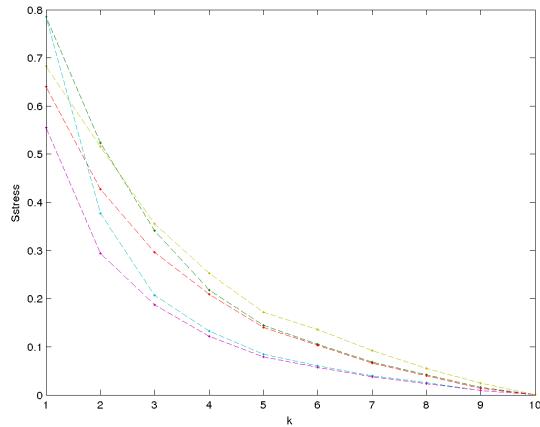
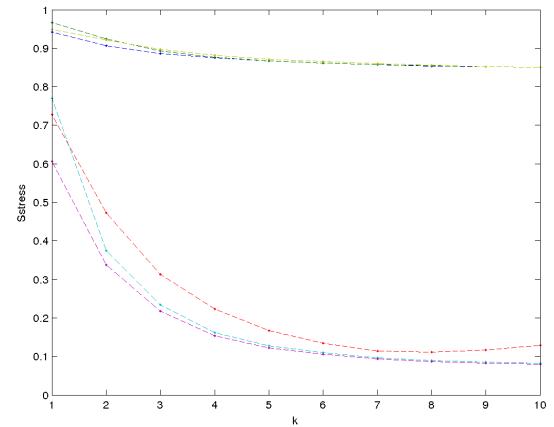
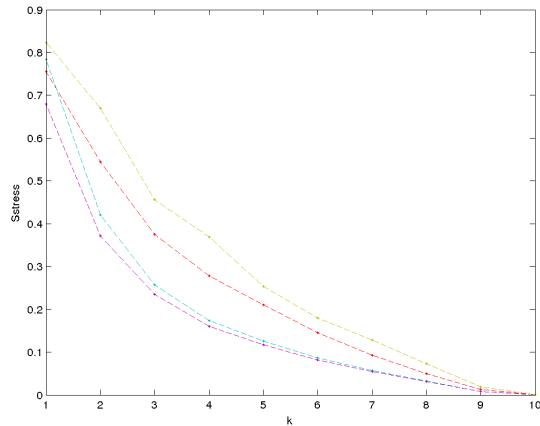
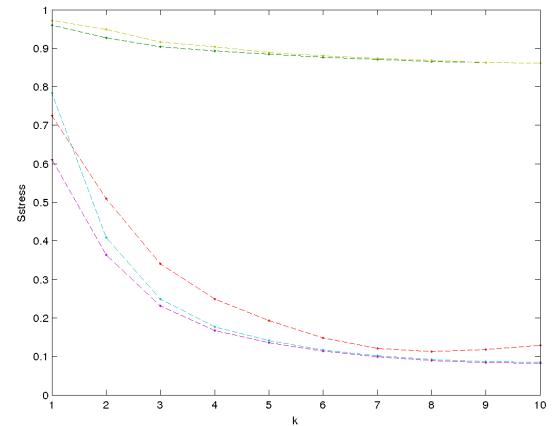
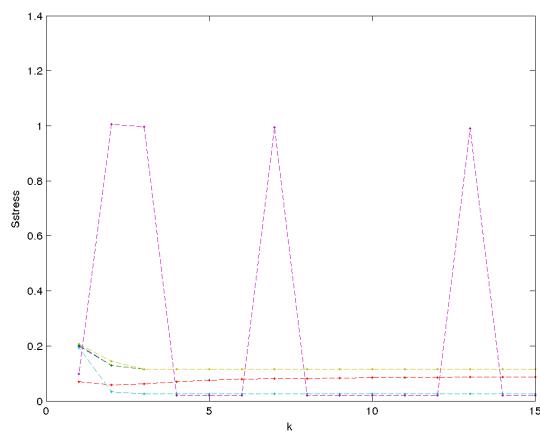
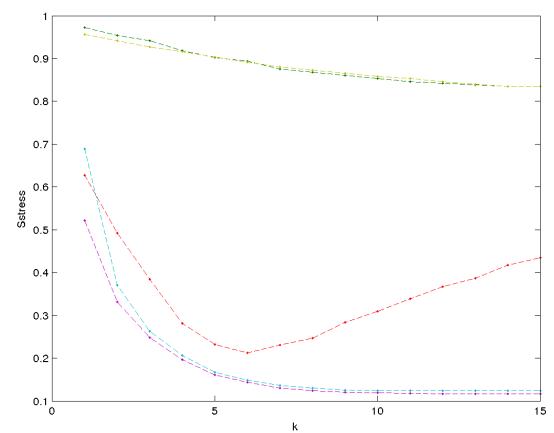
Figura B.43: *Sstress* (formulação tradicional) para base Vogal (Legenda Fig. B.1)(a) Cenário **X** e $\mathbf{D}_{\text{Euclidiana}}$ (b) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$ (c) Cenário **Z** e $\mathbf{D}_{\text{Euclidiana}}$ (d) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$ Figura B.44: *Sstress* (formulação alternativa) para base Triagem de Crédito (Legenda Fig. B.1)(a) Cenário **X** e $\mathbf{D}_{\text{Manhattan}}$ (b) Cenário **Z** e $\mathbf{D}_{\text{Manhattan}}$

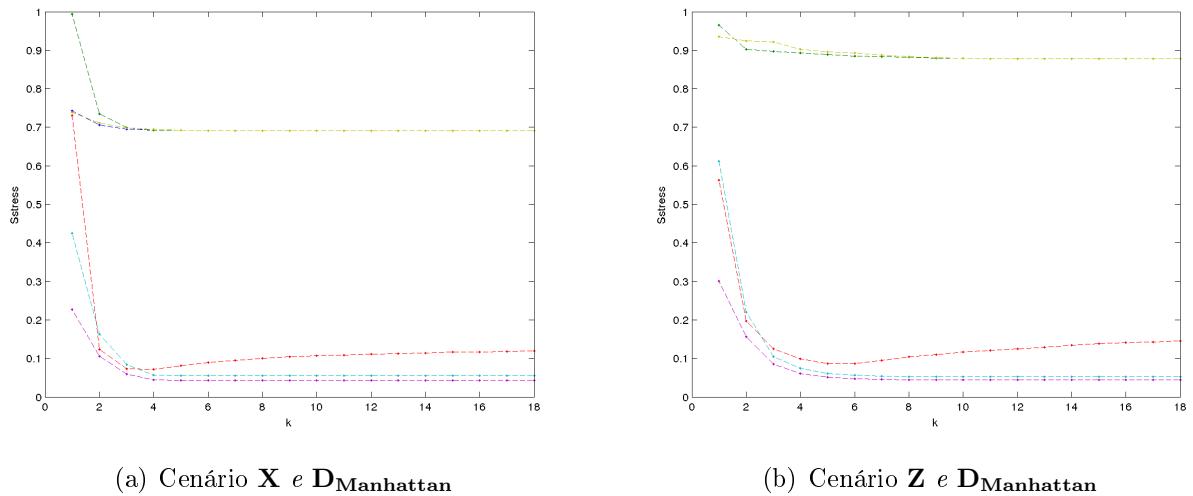
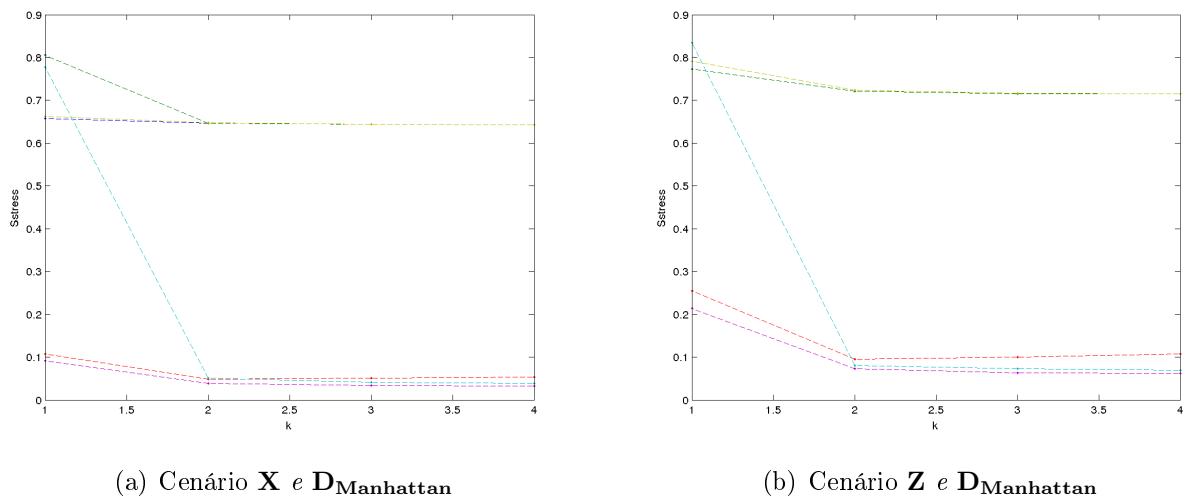
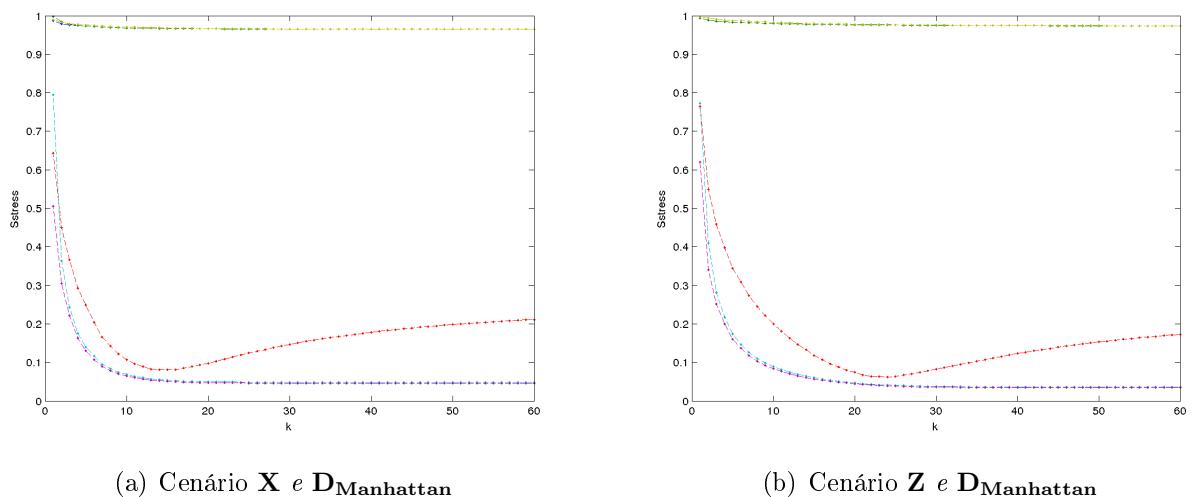
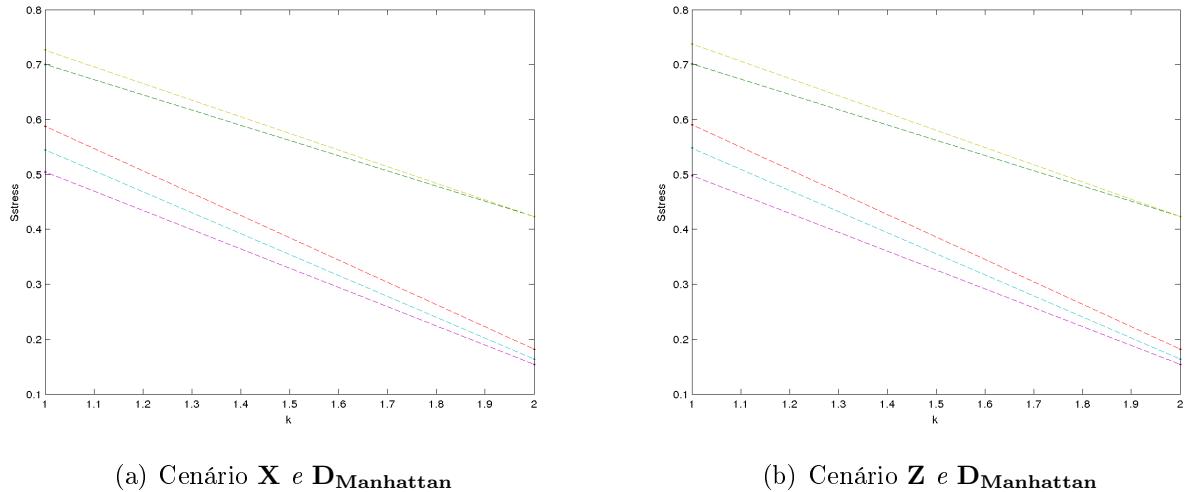
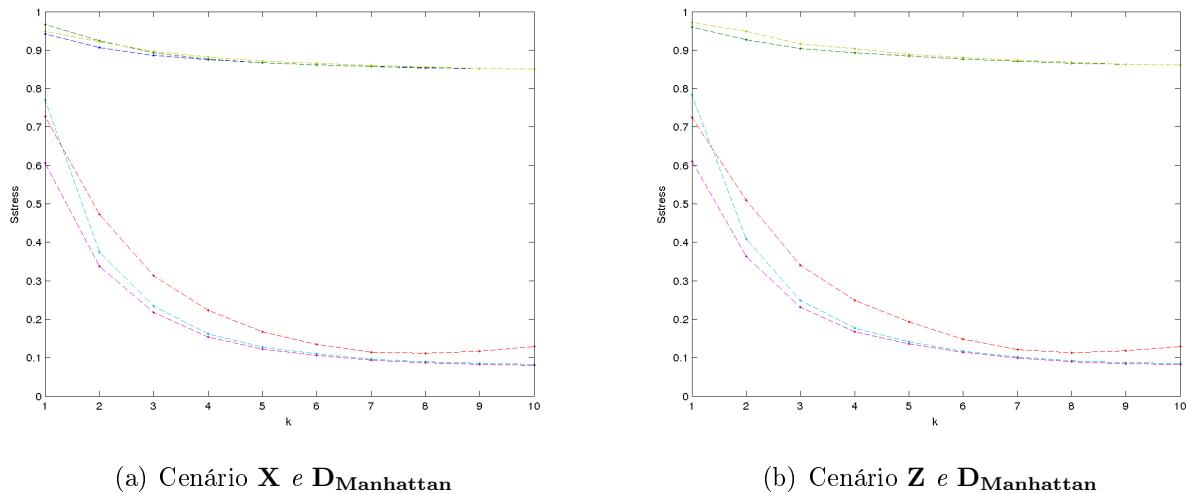
Figura B.45: *Sstress* (formulação alternativa) para base Imagens (Legenda Fig. B.1)Figura B.46: *Sstress* (formulação alternativa) para base Iris (Legenda Fig. B.1)Figura B.47: *Sstress* (formulação alternativa) para base Sonar (Legenda Fig. B.1)

Figura B.48: *Sstress* (formulação alternativa) para base Espiral (Legenda Fig. B.1)Figura B.49: *Sstress* (formulação alternativa) para base Vogal (Legenda Fig. B.1)Figura B.50: Legendas utilizada nos gráficos de acurácia do classificador nos cenários **X** e $\mathbf{D}_{\text{Euclidiana}}$ e **X** e $\mathbf{D}_{\text{Manhattan}}$

—	Dados originais (X)*
◆	PCA
✖	SVD
◆	CMDS - Euclidiana
◆	CMDS - Manhattan
◆	MDS Stress-1 - Euclidiana
◆	MDS Stress-1 - Manhattan
◆	MDS Sstress - Euclidiana
◆	MDS Sstress - Manhattan
◆	FastMap

Figura B.51: Legendas utilizada nos gráficos de acurácia do classificador nos cenários \mathbf{Z} e $\mathbf{D}_{\text{Euclidian}}a$ e \mathbf{Z} e $\mathbf{D}_{\text{Manhattan}}$

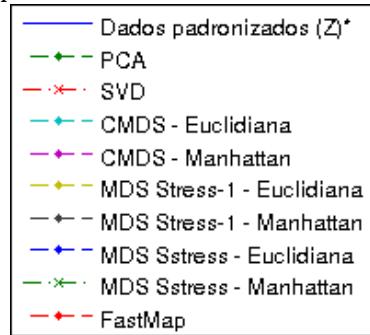


Figura B.52: Acurácia do classificador para base Triagem de Crédito

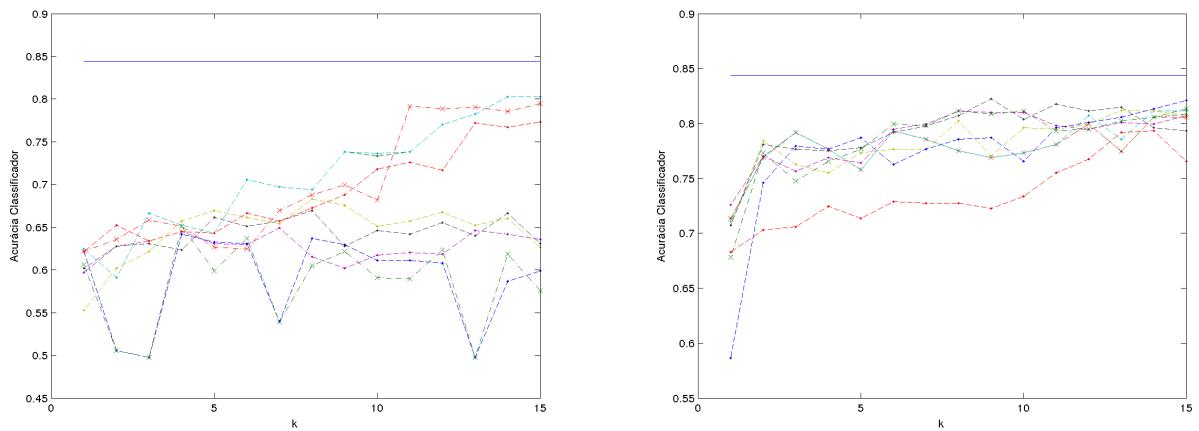


Figura B.53: Acurácia do classificador para base Imagens

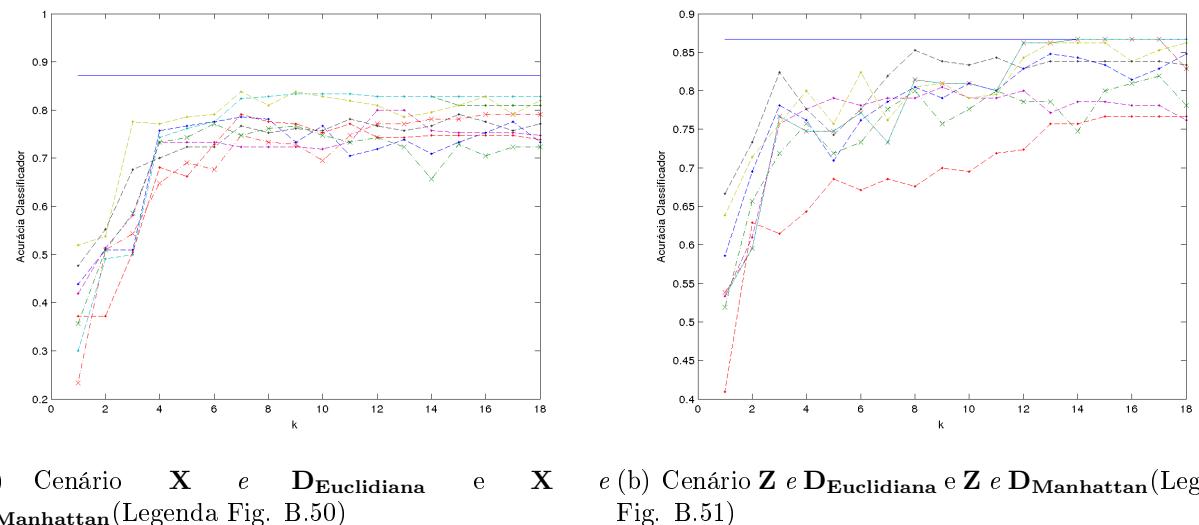


Figura B.54: Acurácia do classificador para base Iris

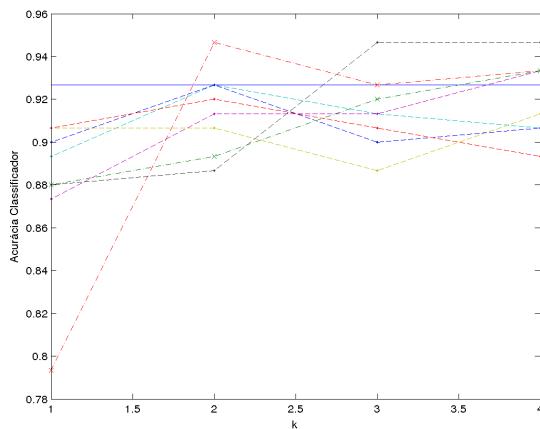
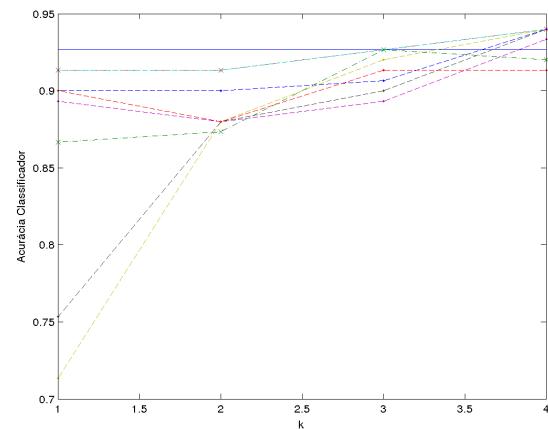
(a) Cenário \mathbf{X} e $\mathbf{D}_{\text{Euclidian}}_{} \text{ e } \mathbf{X} = \mathbf{Z}$ (Legenda Fig. B.50)(b) Cenário \mathbf{Z} e $\mathbf{D}_{\text{Euclidian}}_{} \text{ e } \mathbf{Z} = \mathbf{D}_{\text{Manhattan}}_{} \text{ (Legenda Fig. B.51)}$

Figura B.55: Acurácia do classificador para base Sonar

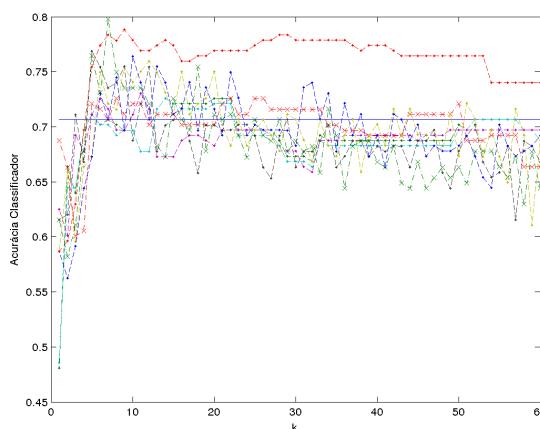
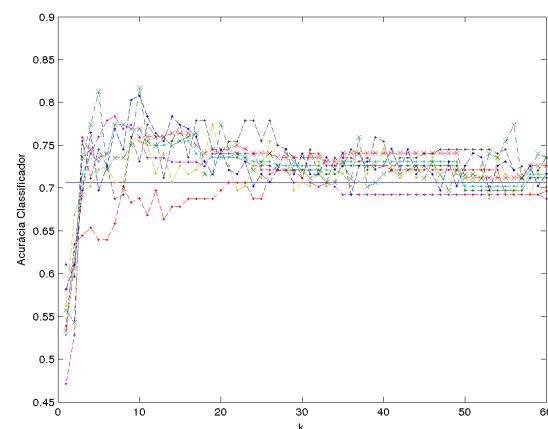
(a) Cenário \mathbf{X} e $\mathbf{D}_{\text{Euclidian}}_{} \text{ e } \mathbf{X} = \mathbf{Z}$ (Legenda Fig. B.50)(b) Cenário \mathbf{Z} e $\mathbf{D}_{\text{Euclidian}}_{} \text{ e } \mathbf{Z} = \mathbf{D}_{\text{Manhattan}} \text{ (Legenda Fig. B.51)}$

Figura B.56: Acurácia do classificador para base Espiral

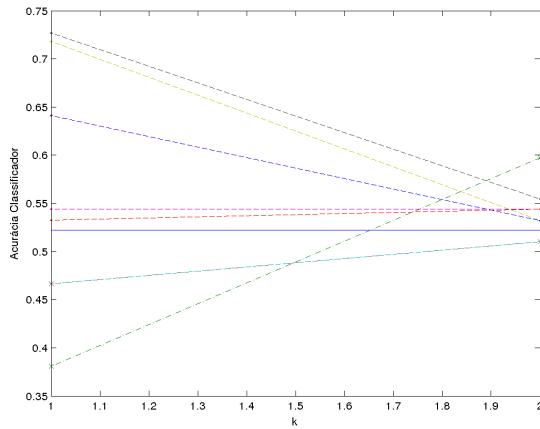
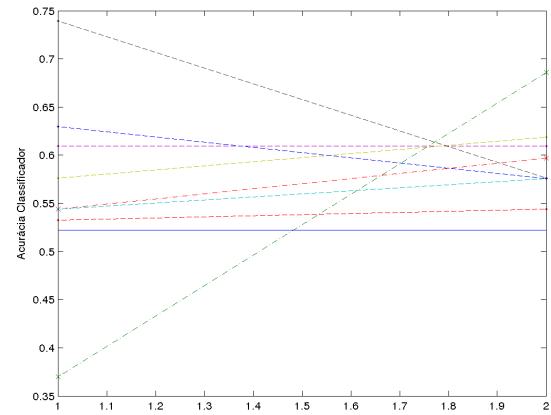
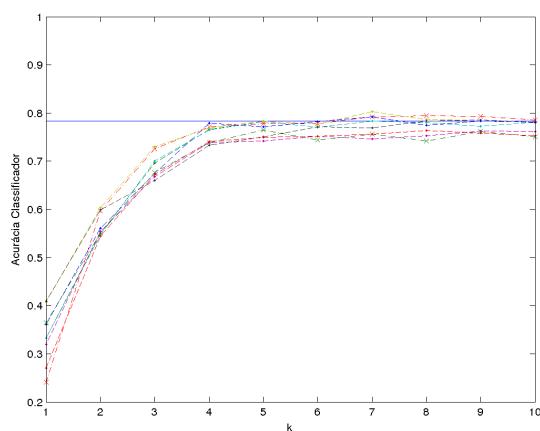
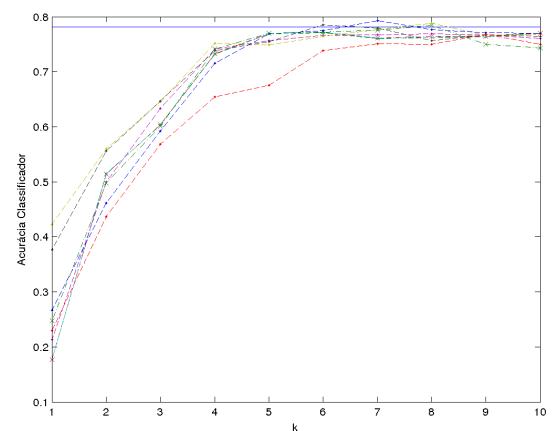
(a) Cenário \mathbf{X} e $\mathbf{D}_{\text{Euclidian}}_{} \text{ e } \mathbf{X} = \mathbf{D}_{\text{Manhattan}}_{} \text{ (Legenda Fig. B.50)}$ (b) Cenário \mathbf{Z} e $\mathbf{D}_{\text{Euclidian}}_{} \text{ e } \mathbf{Z} = \mathbf{D}_{\text{Manhattan}}_{} \text{ (Legenda Fig. B.51)}$

Figura B.57: Acurácia do classificador para base Vogal

(a) Cenário \mathbf{X} e $\mathbf{D}_{\text{Euclidian}}_{} \text{ e } \mathbf{X} = \mathbf{D}_{\text{Manhattan}}_{} \text{ (Legenda Fig. B.50)}$ (b) Cenário \mathbf{Z} e $\mathbf{D}_{\text{Euclidian}}_{} \text{ e } \mathbf{Z} = \mathbf{D}_{\text{Manhattan}}_{} \text{ (Legenda Fig. B.51)}$

C Visualização dos Dados

Este apêndice contém os *matrix plot* (gráficos de dispersão entre todas as dimensões) das matrizes de dados originais e padronizados para as três das seis bases utilizadas, são elas: Iris (C.1), Espiral (C.2) e Vogal (C.3), o das demais (Triagem de Crédito, Imagens e Sonar) foram gerados e analisados durante o desenvolvimento do trabalho, mas por se tratarem de dimensões muito grandes suas visualizações neste documento não é viável. E, em seguida, apresenta os gráficos de dispersão (*scatter plot*) para as dimensões visíveis das matrizes de dados obtidas por cada método ($\mathbf{Y} = \mathbf{Score}$) para todas as bases de dados.

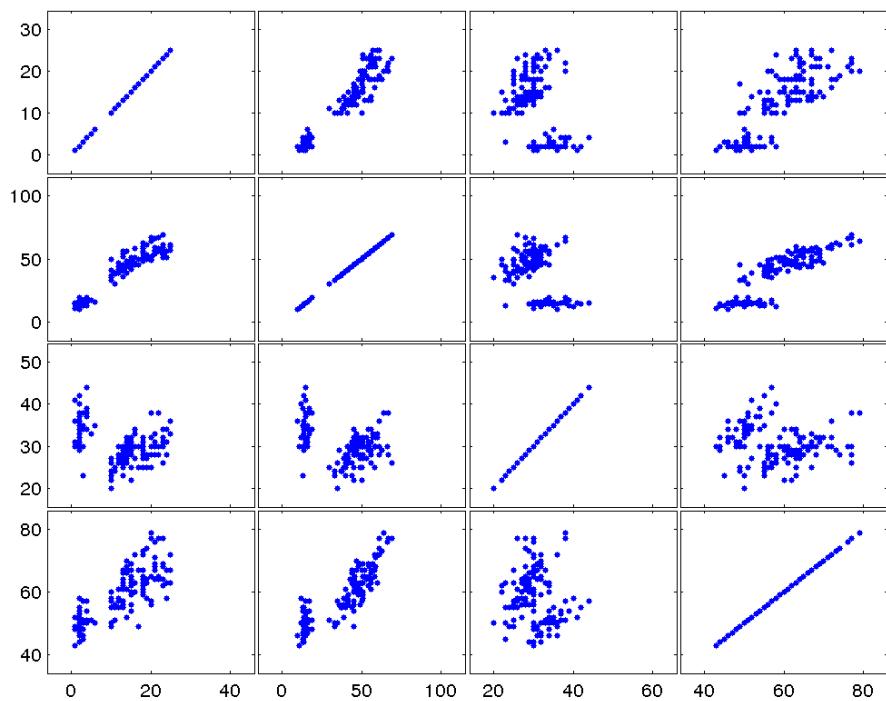
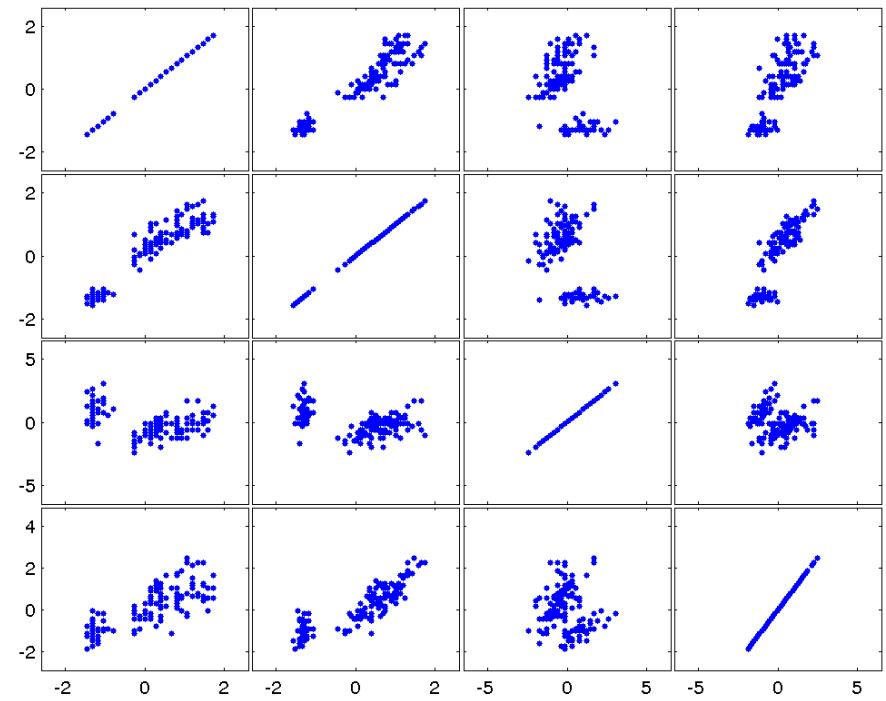
Figura C.1: *Matrix plot* para base Iris(a) Dados originais \mathbf{X} (b) Dados padronizados \mathbf{Z}

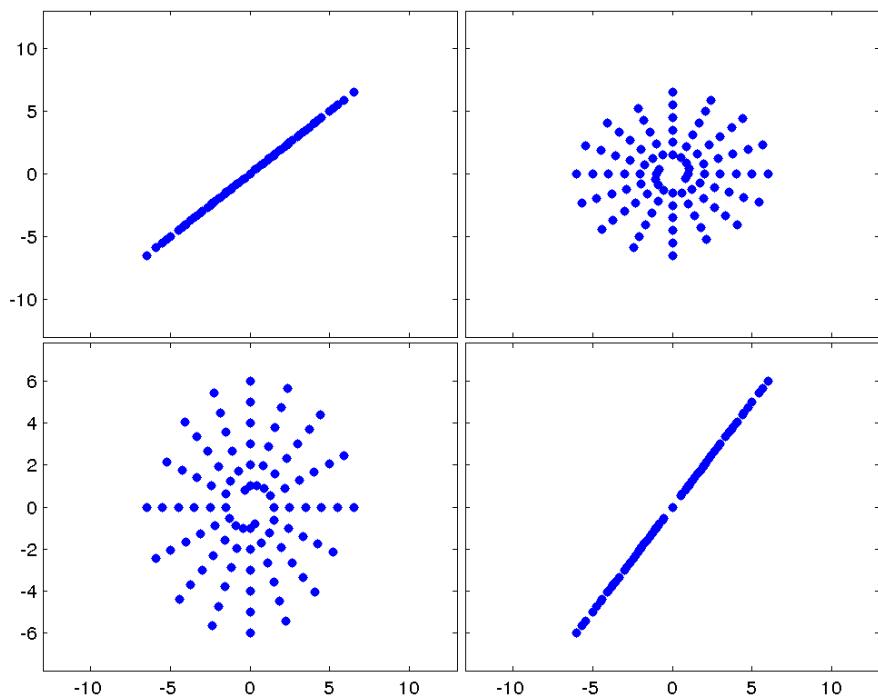
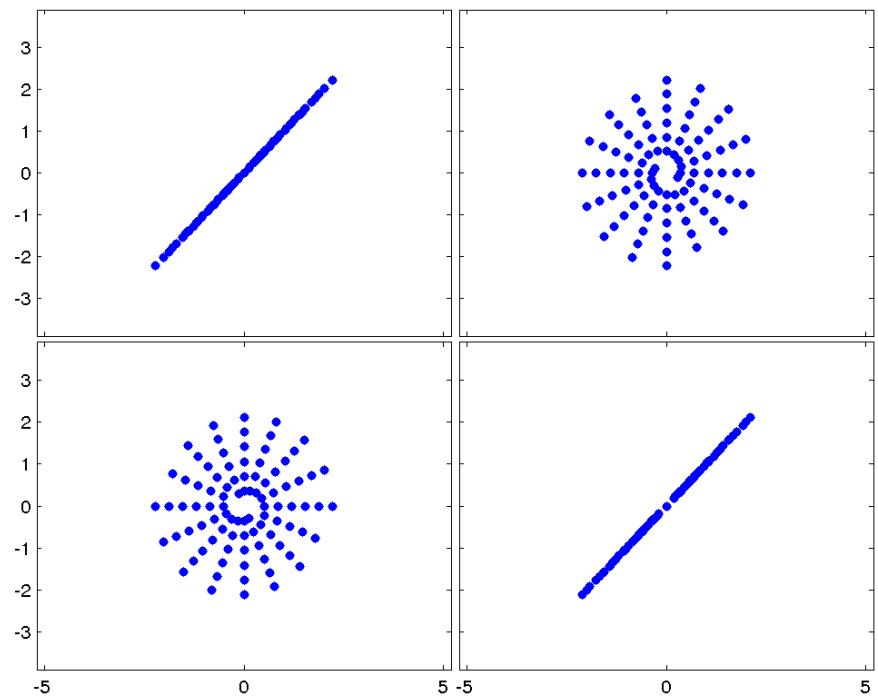
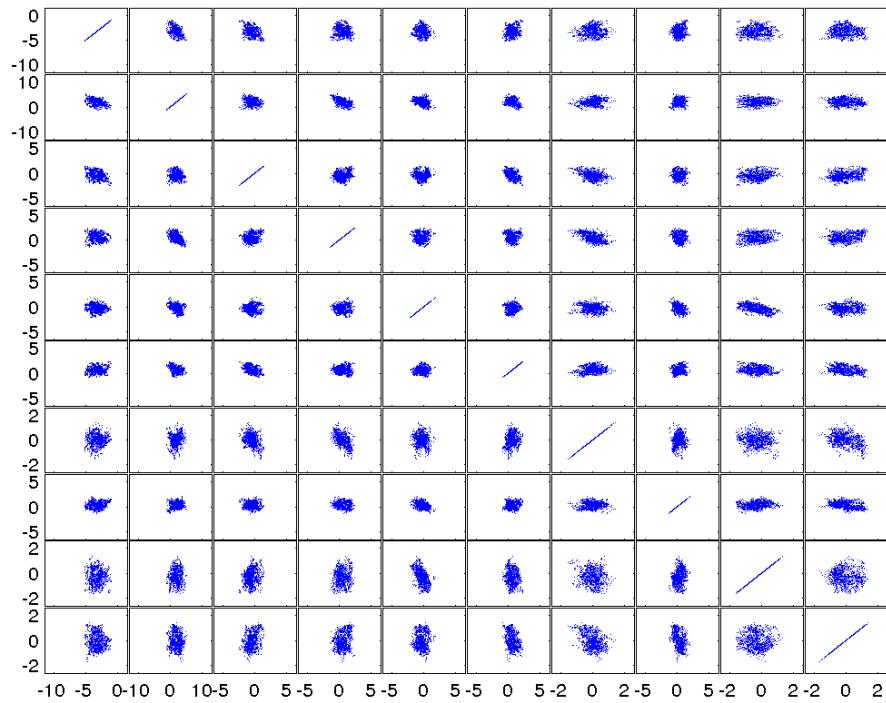
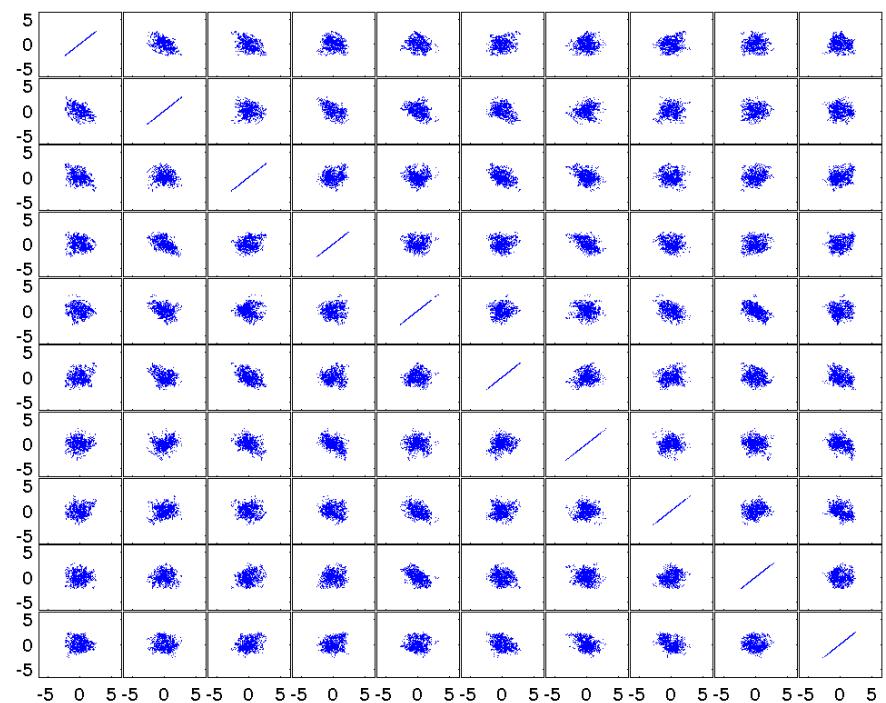
Figura C.2: *Matrix plot* para base Espiral(a) Dados originais \mathbf{X} (b) Dados padronizados \mathbf{Z}

Figura C.3: *Matrix plot* para base Vogal



(a) Dados originais \mathbf{X}



(b) Dados padronizados \mathbf{Z}

Figura C.4: Gráfico de dispersão para as representações unidimensionais a partir da base Triagem de Crédito não padronizada

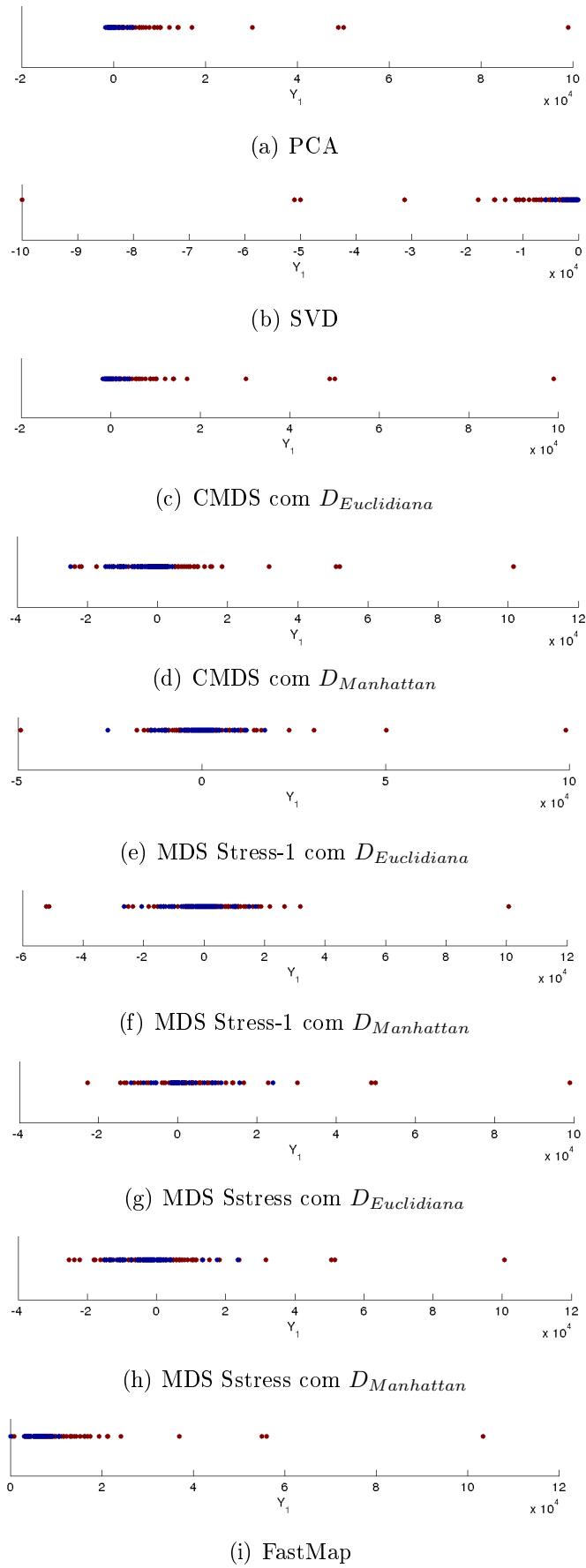


Figura C.5: Gráfico de dispersão para as representações unidimensionais a partir da base Imagens não padronizada

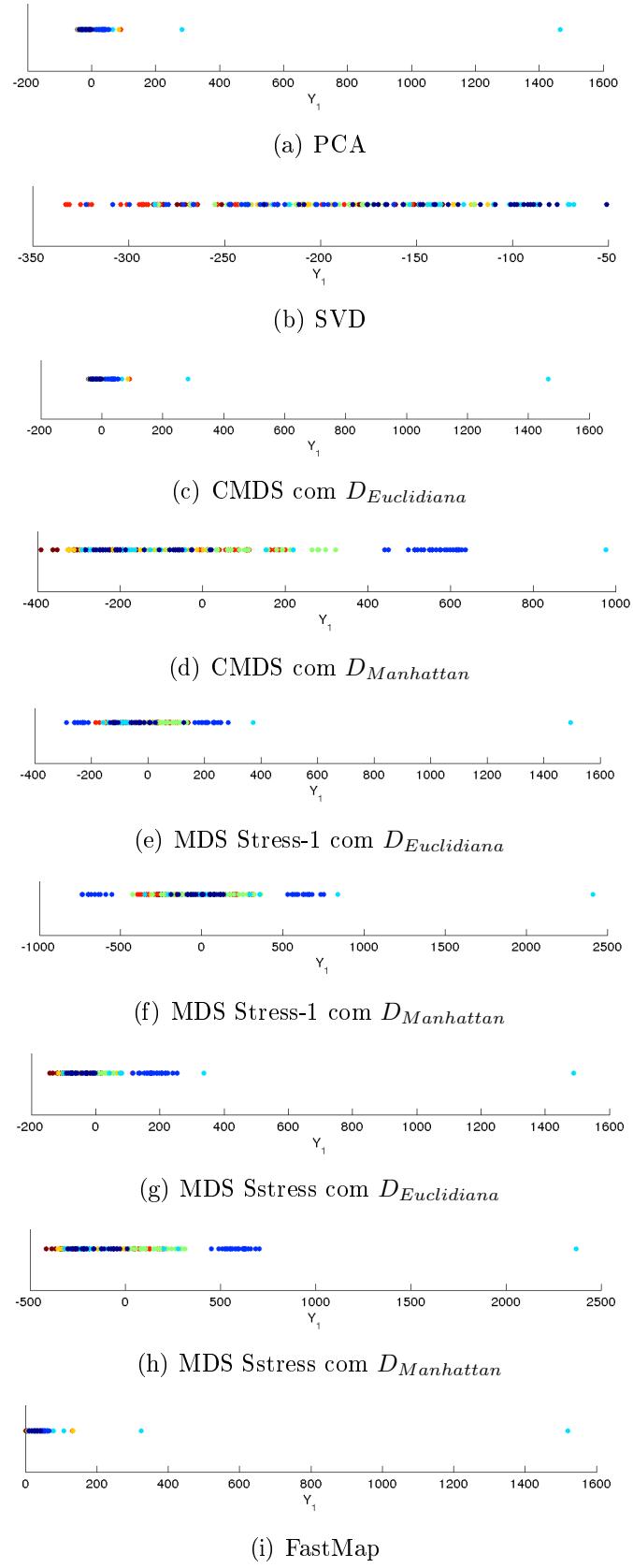


Figura C.6: Gráfico de dispersão para as representações unidimensionais a partir da base Iris não padronizada

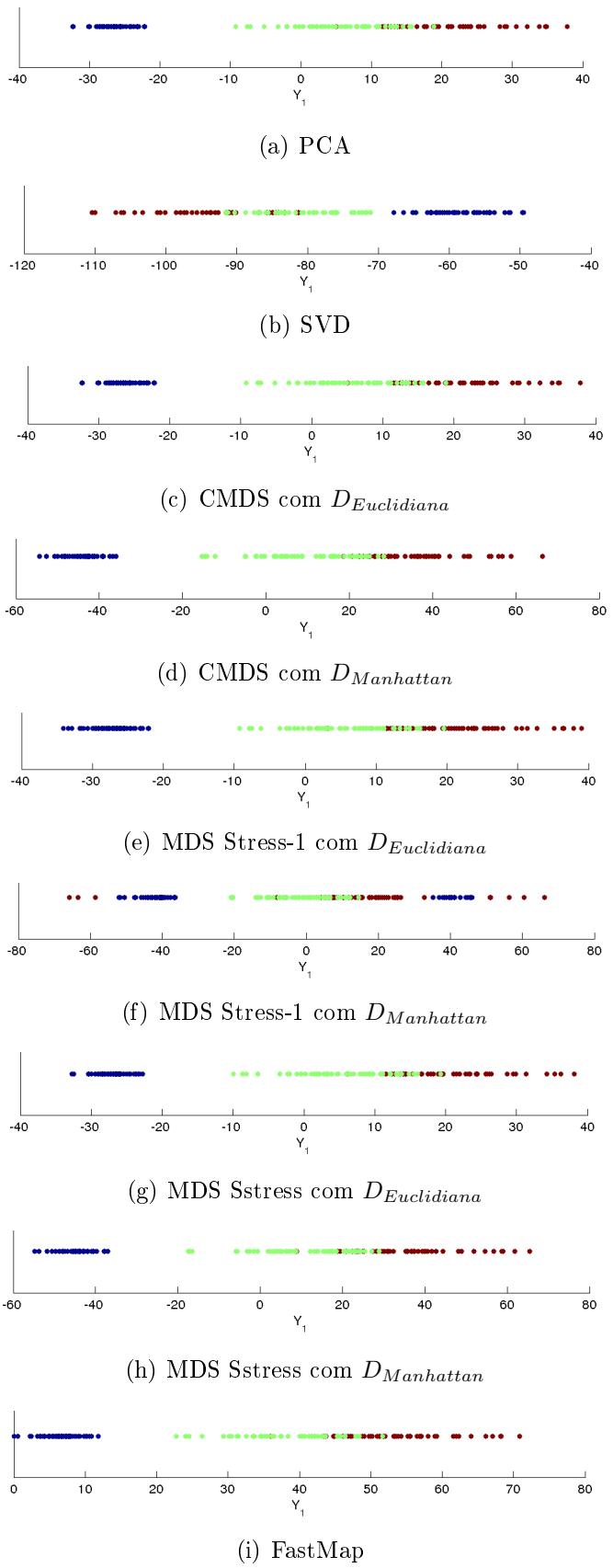


Figura C.7: Gráfico de dispersão para as representações unidimensionais a partir da base Sonar não padronizada

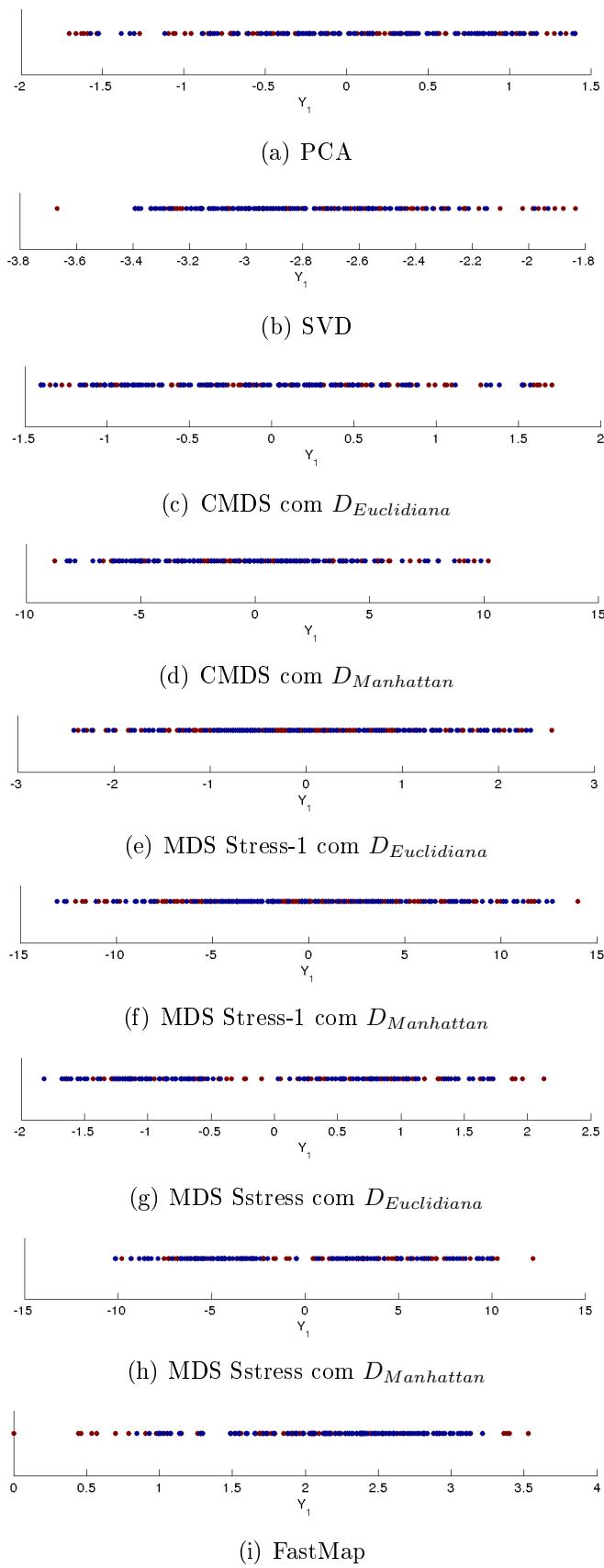
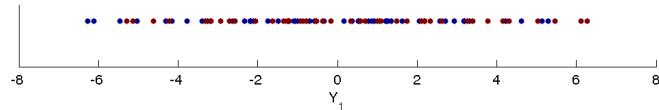
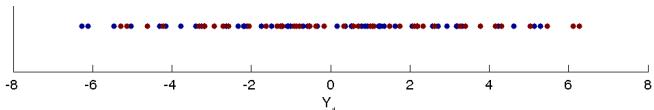


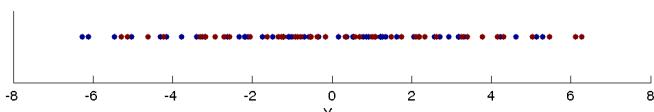
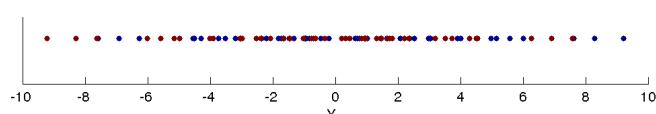
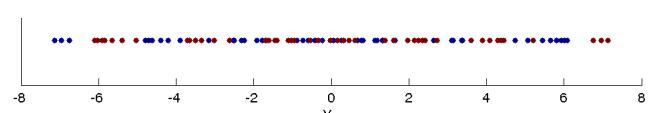
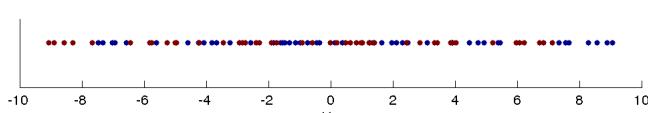
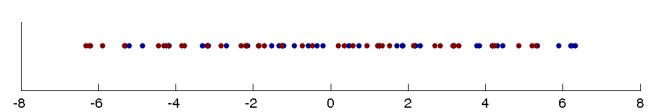
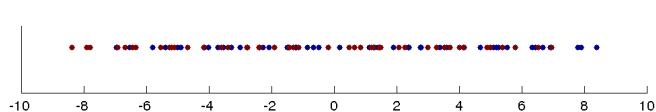
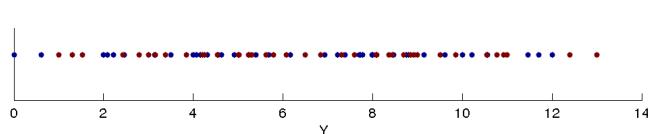
Figura C.8: Gráfico de dispersão para as representações unidimensionais a partir da base Espiral não padronizada



(a) PCA



(b) SVD

(c) CMDS com $D_{Euclidiana}$ (d) CMDS com $D_{Manhattan}$ (e) MDS Stress-1 com $D_{Euclidiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Euclidiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

Figura C.9: Gráfico de dispersão para as representações unidimensionais a partir da base Vocal não padronizada

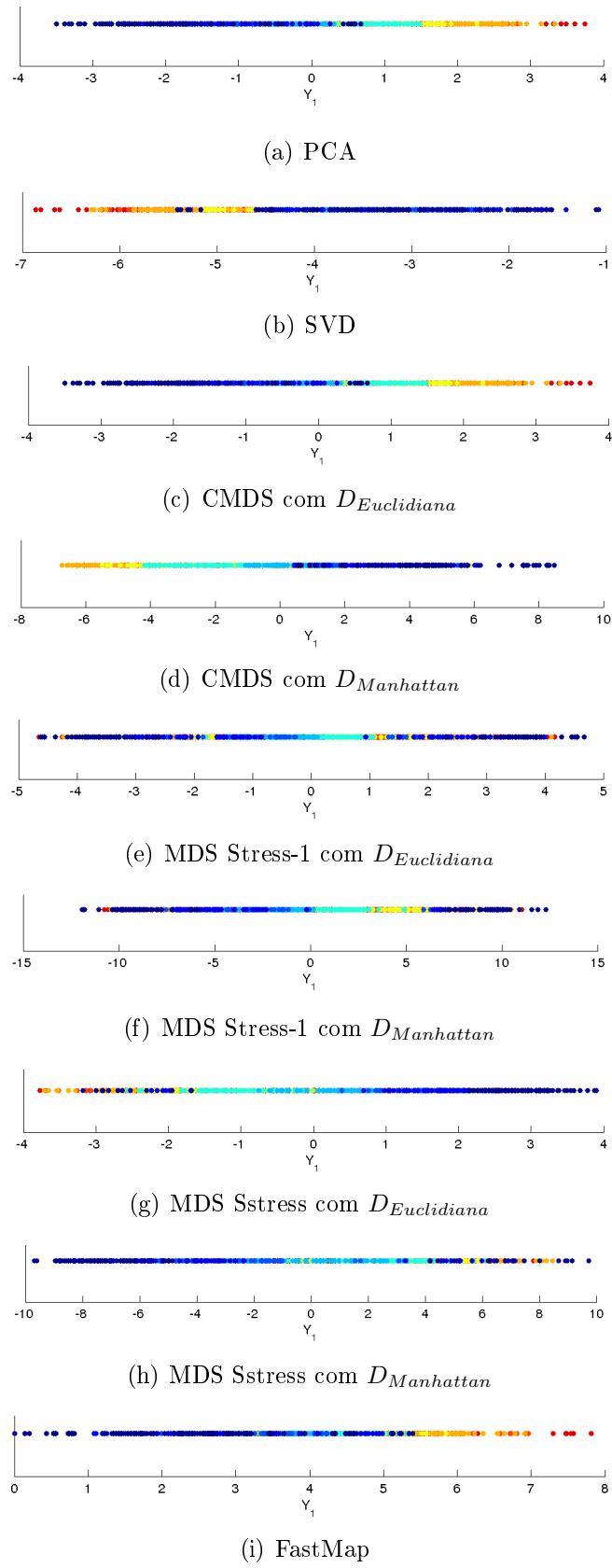


Figura C.10: Gráfico de dispersão para as representações unidimensionais a partir da base Triagem de Crédito padronizada

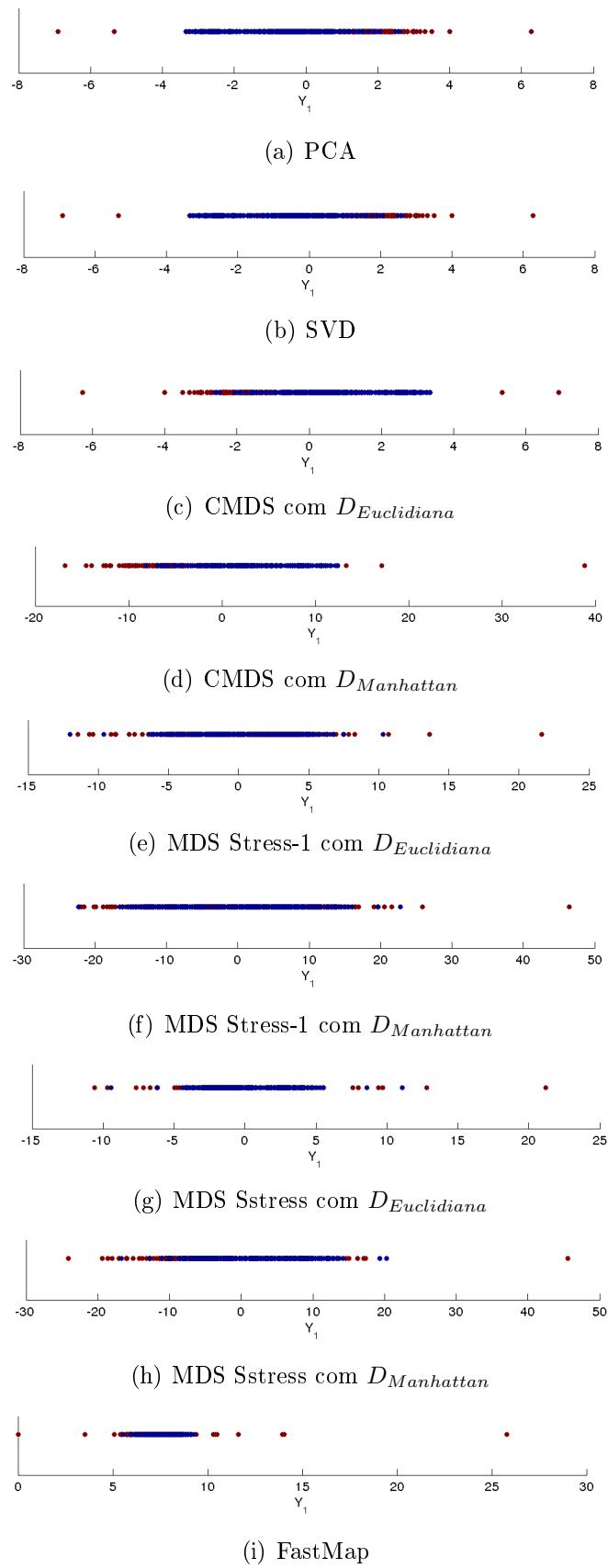
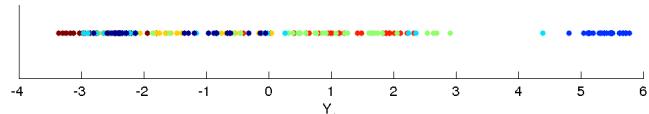
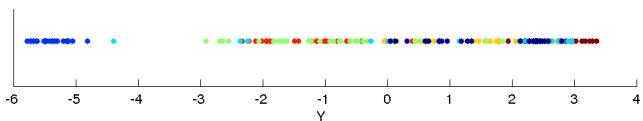


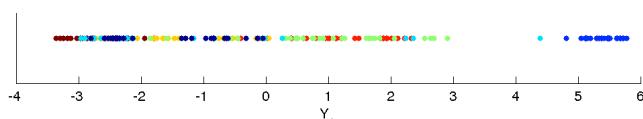
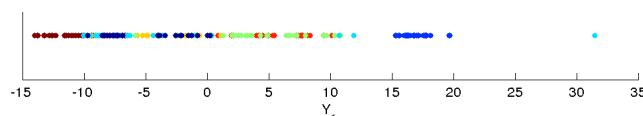
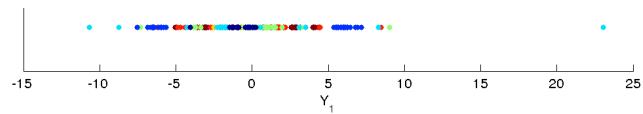
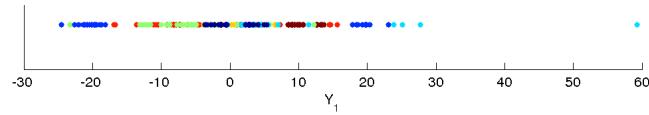
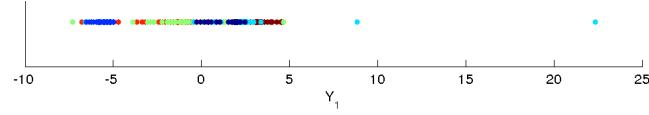
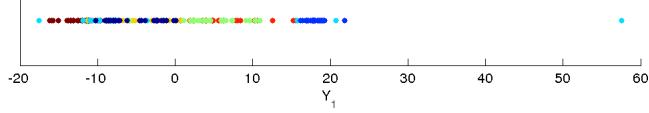
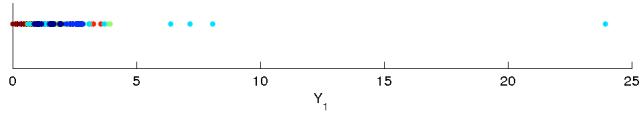
Figura C.11: Gráfico de dispersão para as representações unidimensionais a partir da base Imagens padronizada



(a) PCA



(b) SVD

(c) CMDS com $D_{Eucladiana}$ (d) CMDS com $D_{Manhattan}$ (e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

Figura C.12: Gráfico de dispersão para as representações unidimensionais a partir da base Iris padronizada

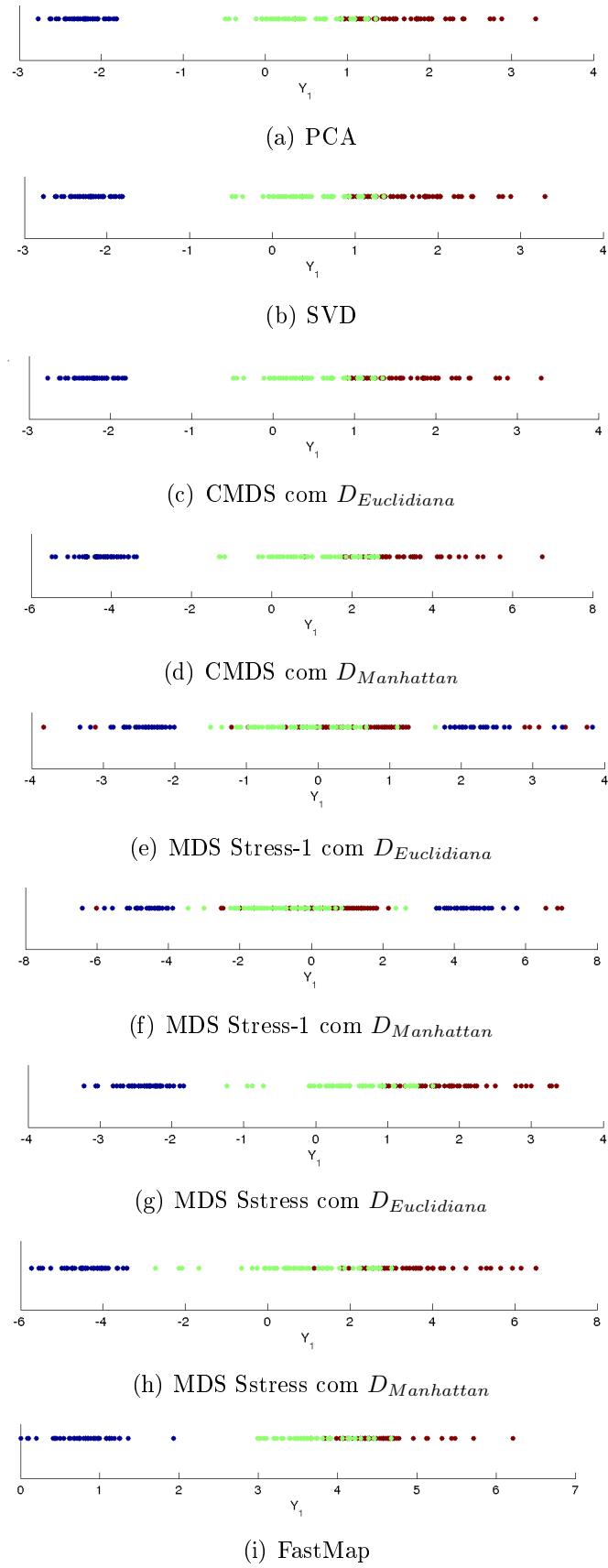


Figura C.13: Gráfico de dispersão para as representações unidimensionais a partir da base Sonar padronizada

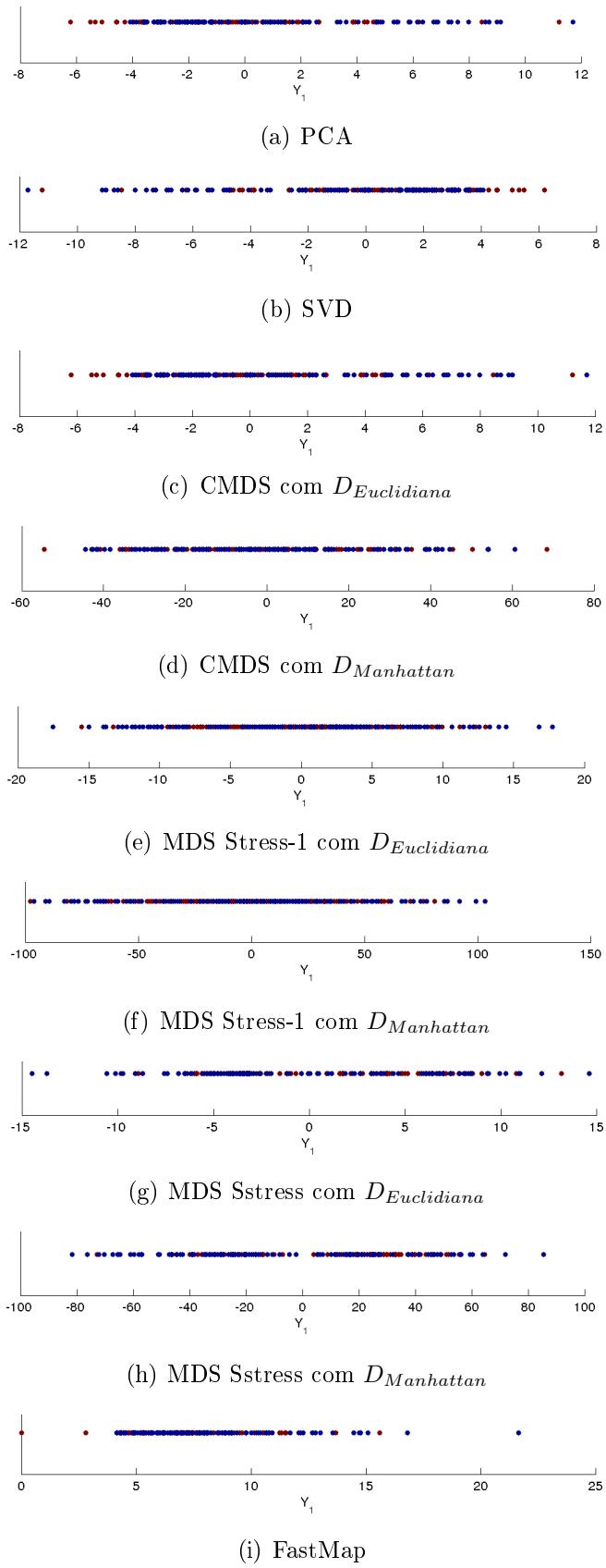


Figura C.14: Gráfico de dispersão para as representações unidimensionais a partir da base Espiral padronizada

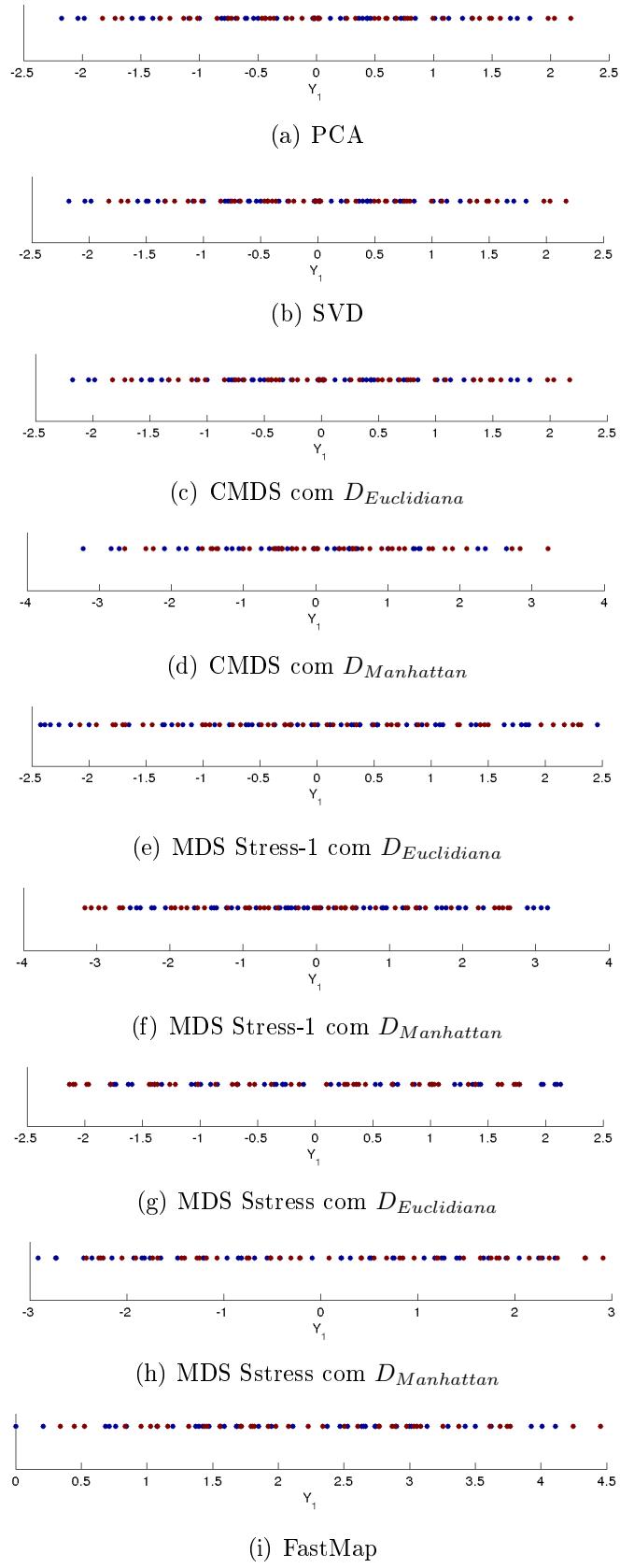


Figura C.15: Gráfico de dispersão para as representações unidimensionais a partir da base Vogal padronizada

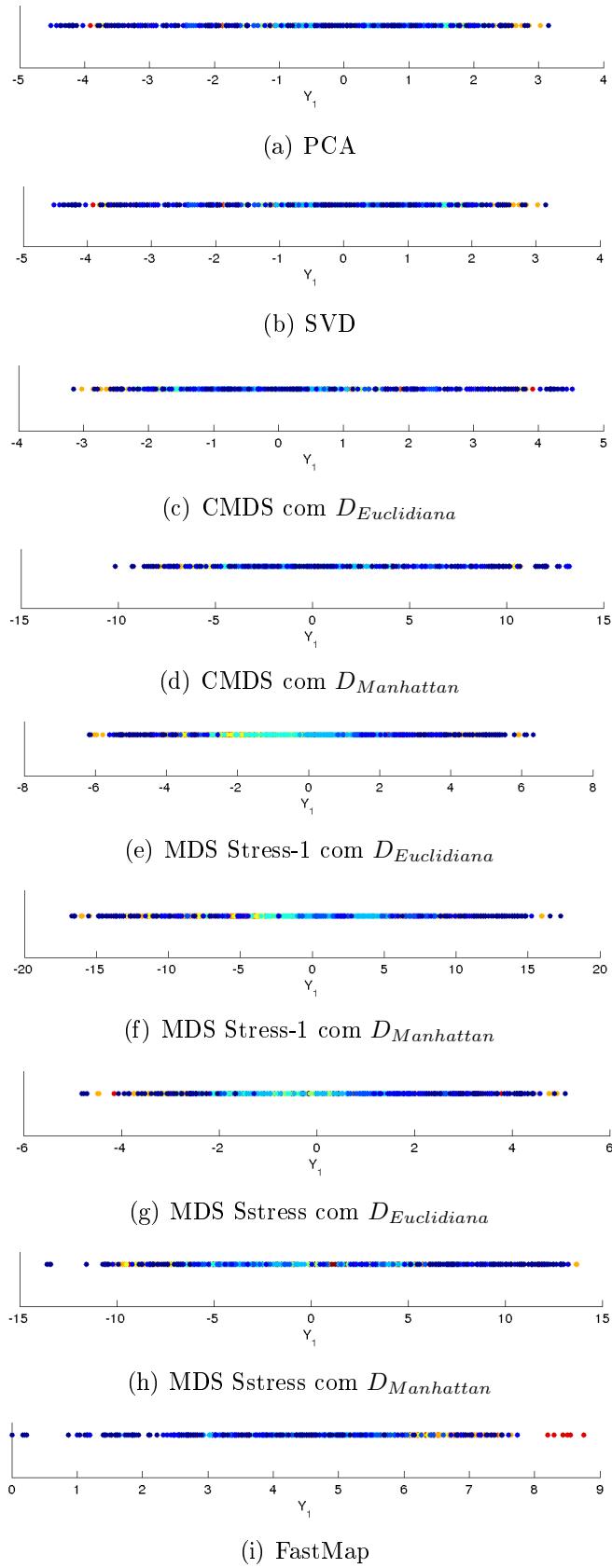
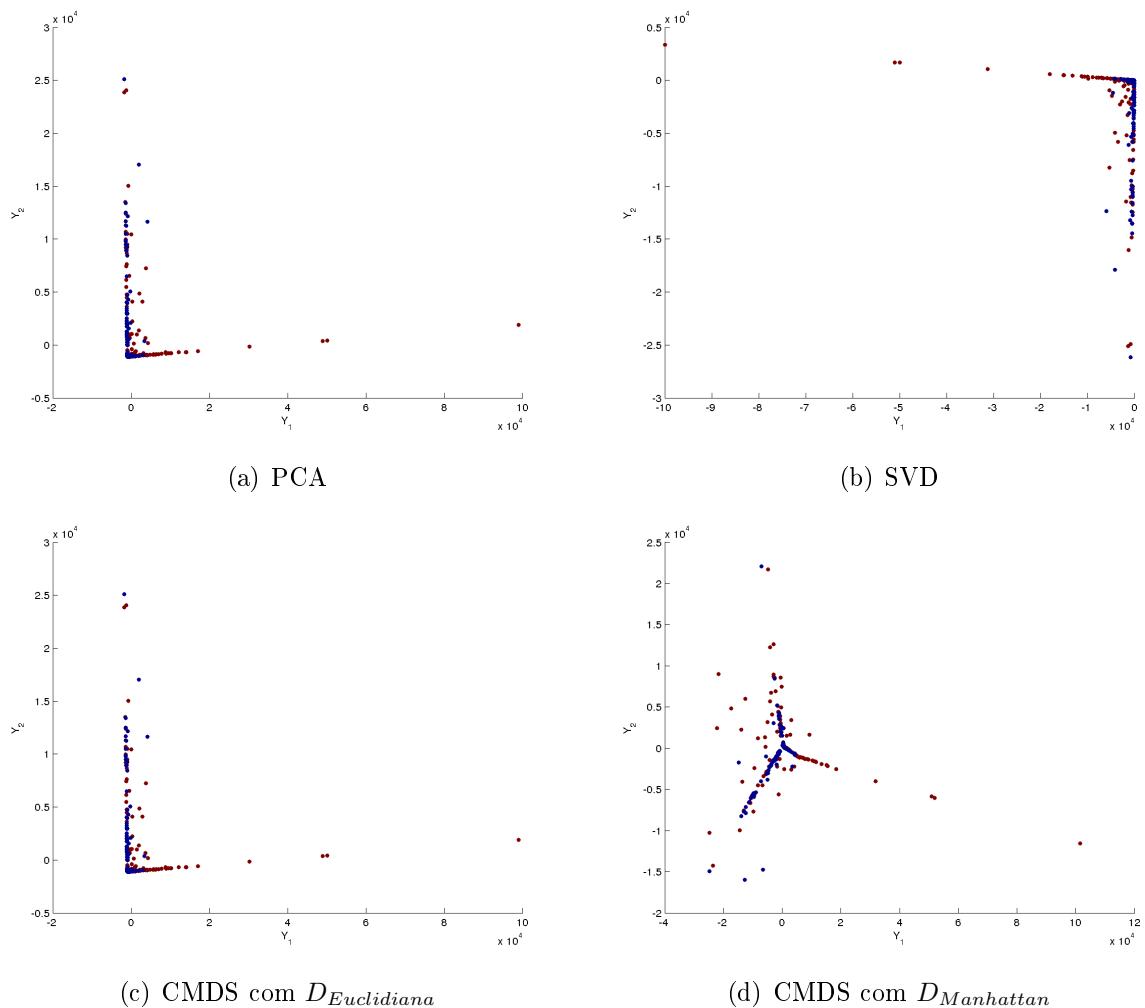
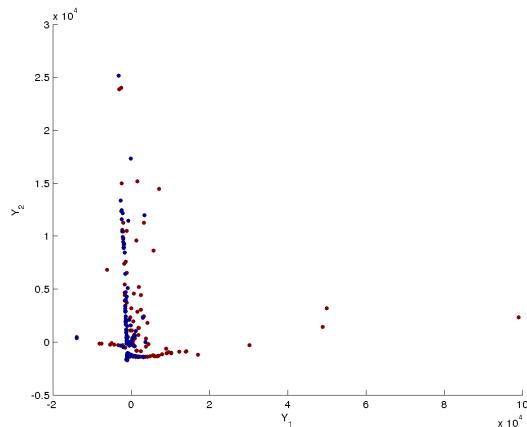
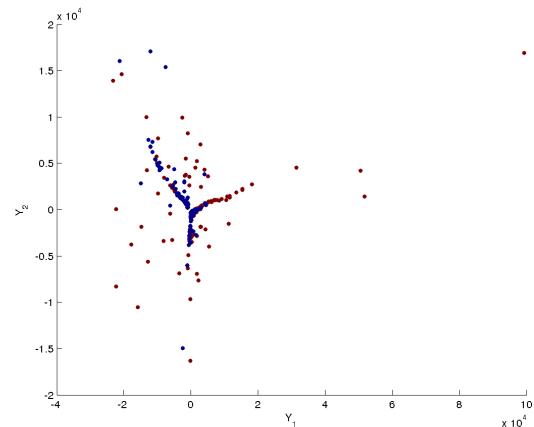
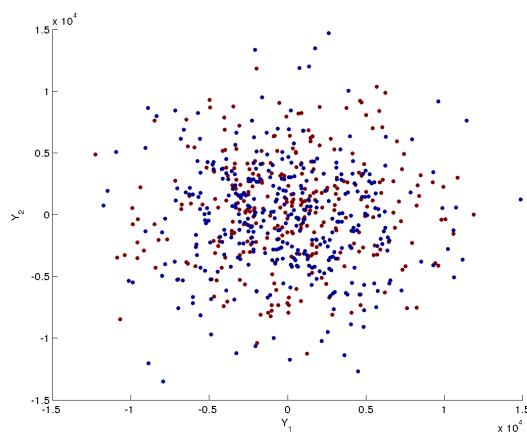
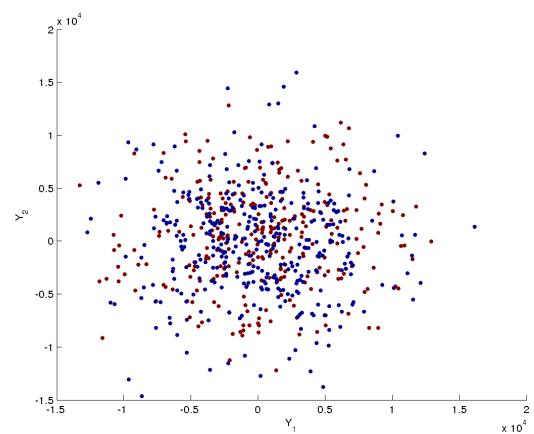
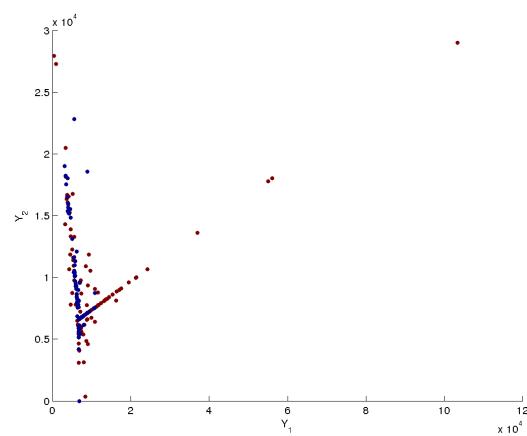


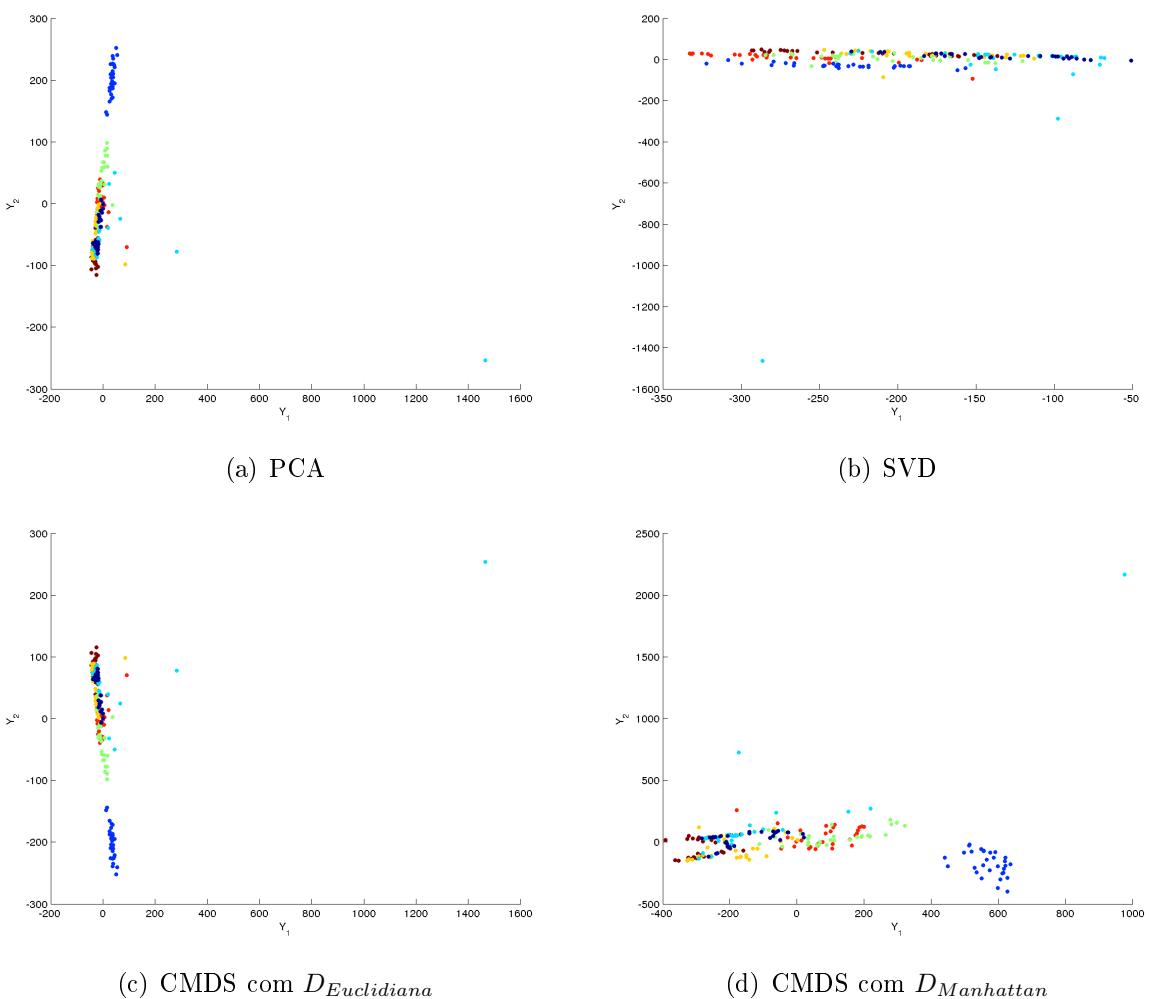
Figura C.16: Gráfico de dispersão para as representações bidimensionais a partir da base Triagem de Crédito não padronizada

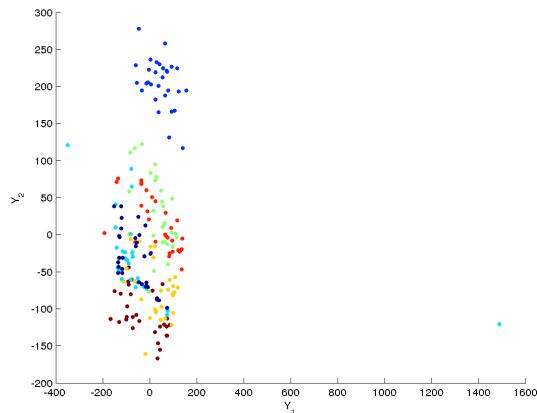
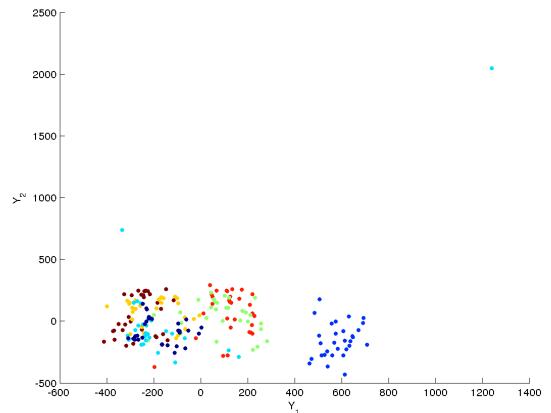
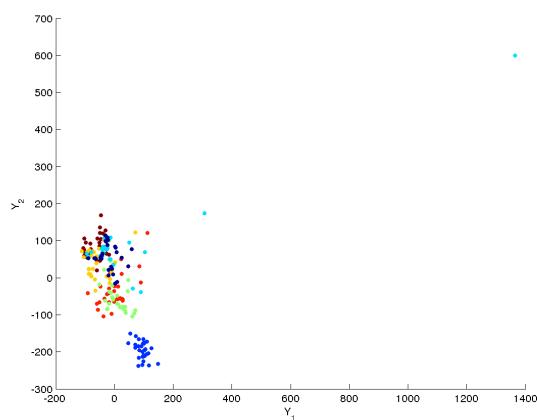
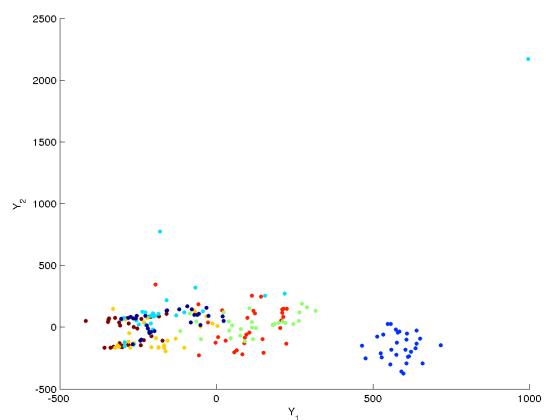
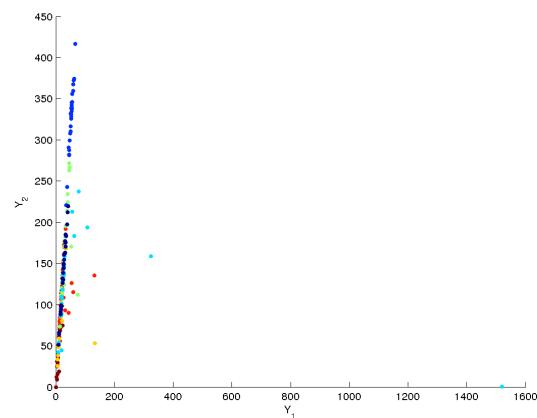


(e) MDS Stress-1 com $D_{Euclidian}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Euclidian}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

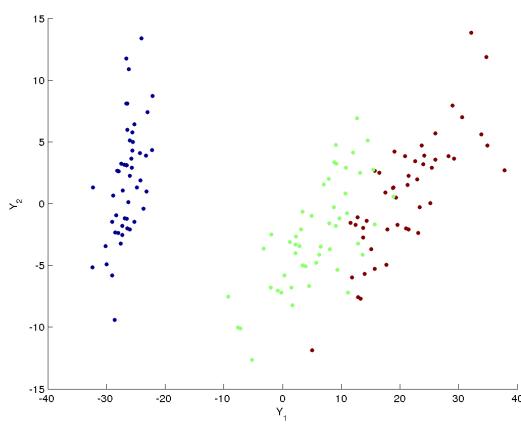
Figura C.17: Gráfico de dispersão para as representações bidimensionais a partir da base Imagens não padronizada



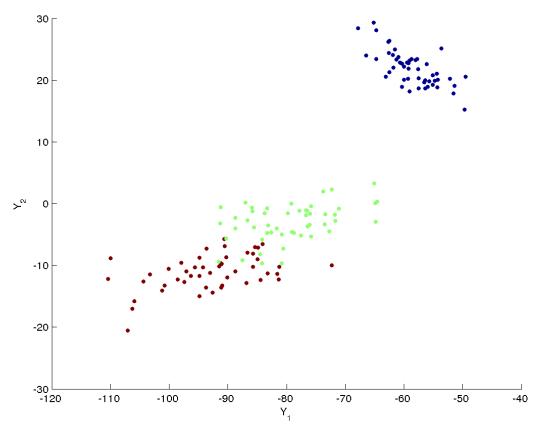
(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

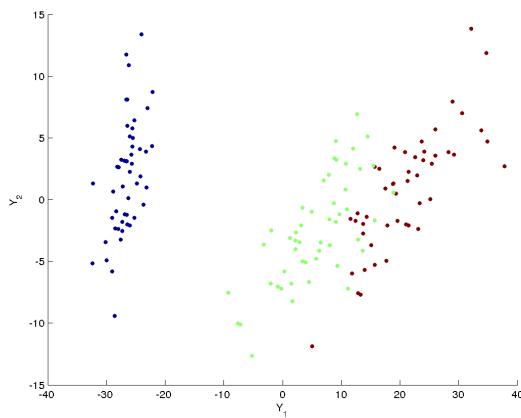
Figura C.18: Gráfico de dispersão para as representações bidimensionais a partir da base Iris não padronizada



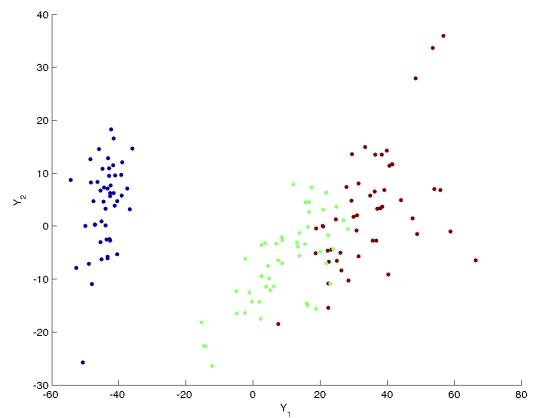
(a) PCA



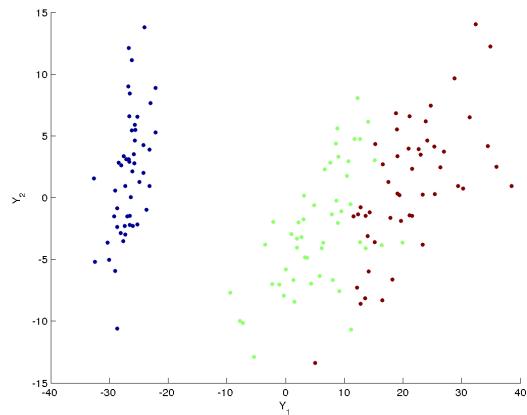
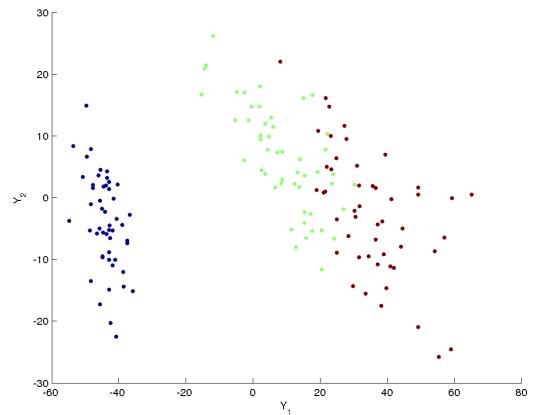
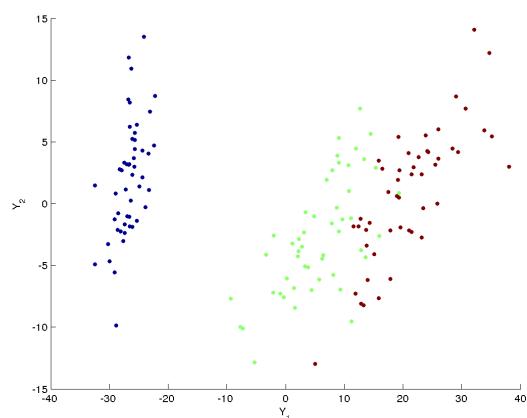
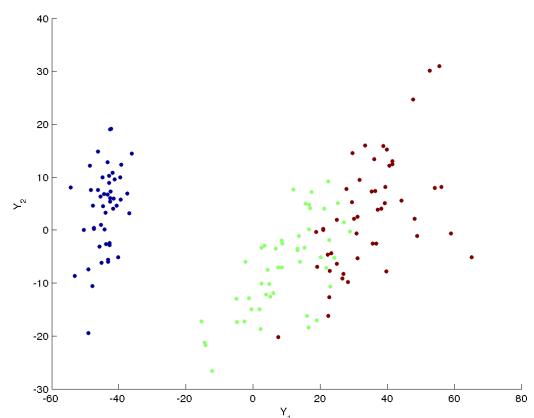
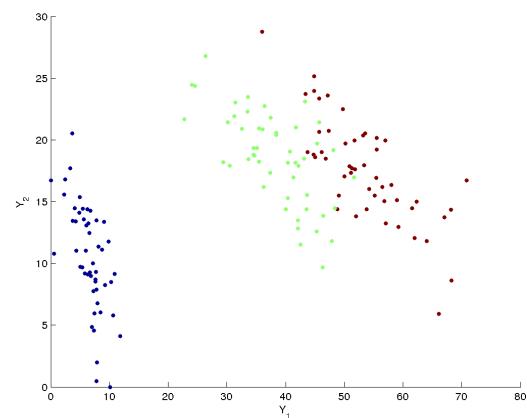
(b) SVD



(c) CMDS com $D_{Euclidian}$

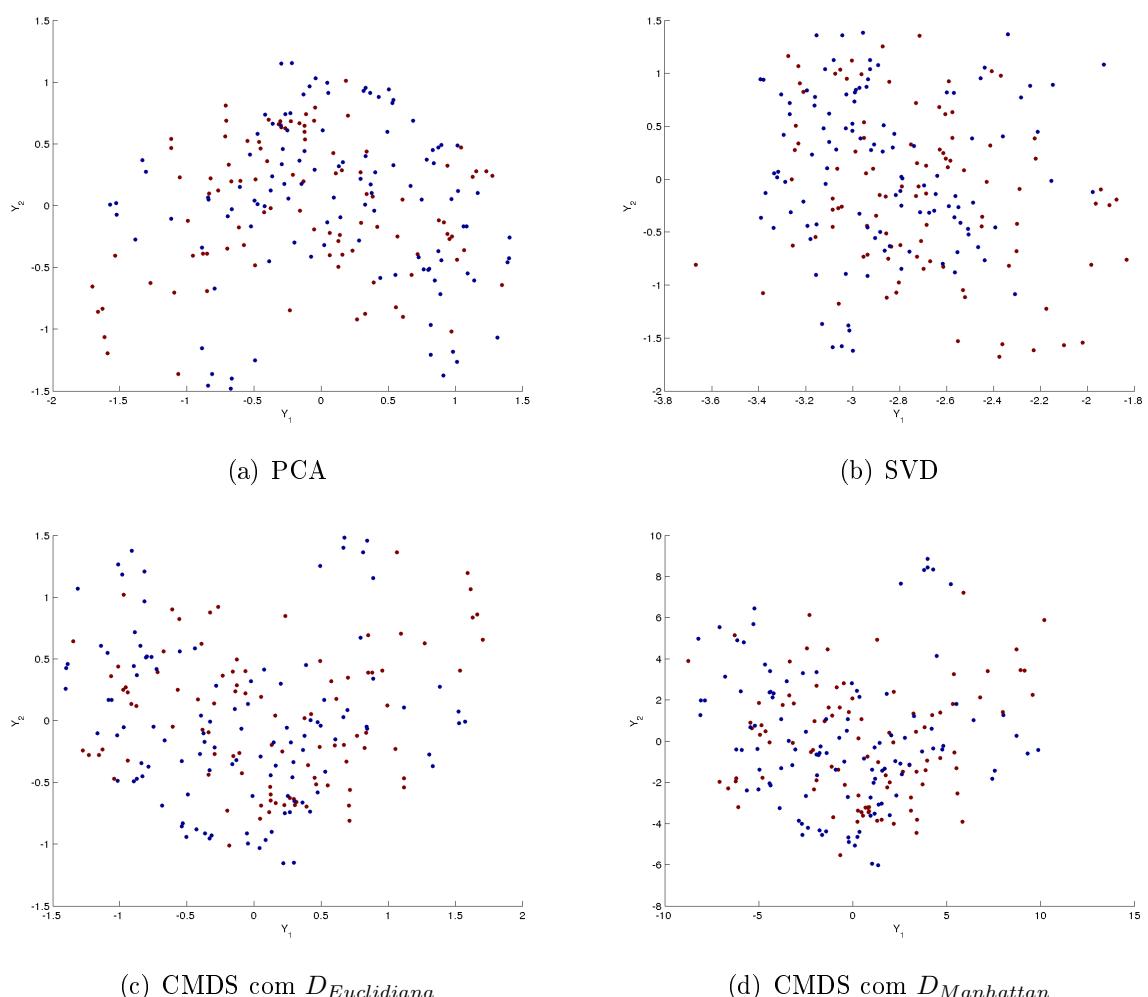


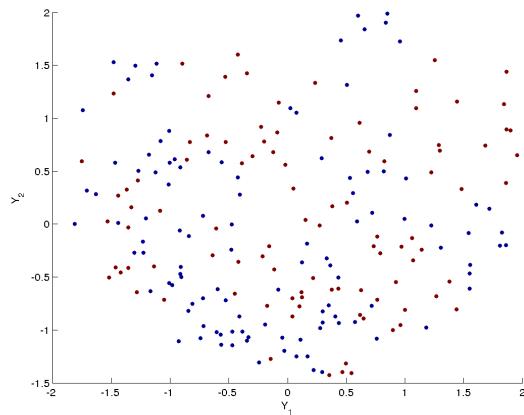
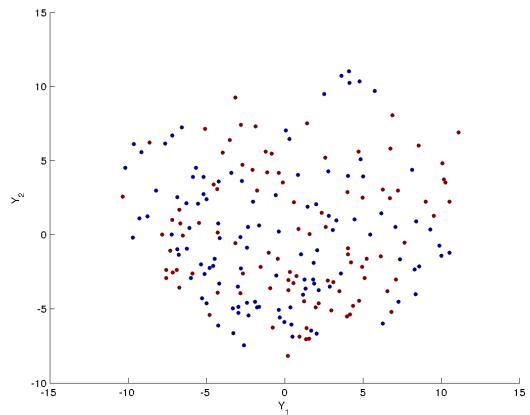
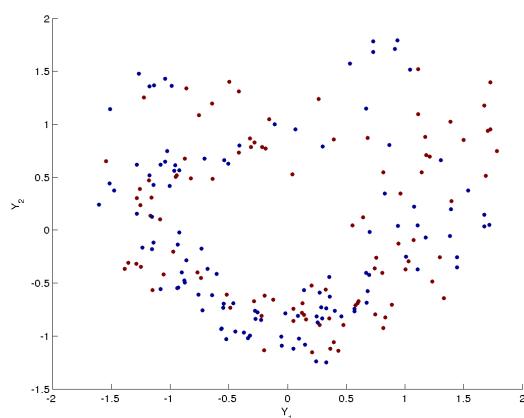
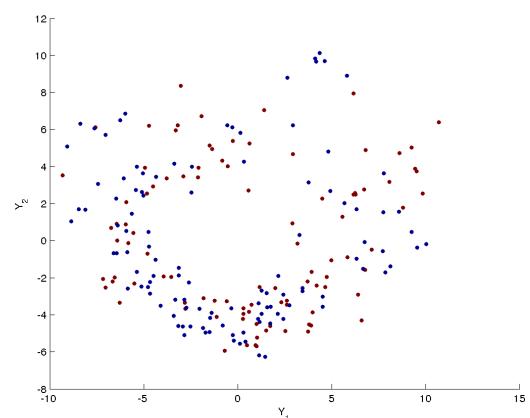
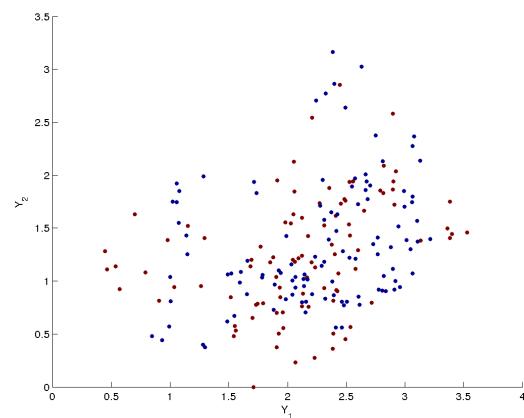
(d) CMDS com $D_{Manhattan}$

(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

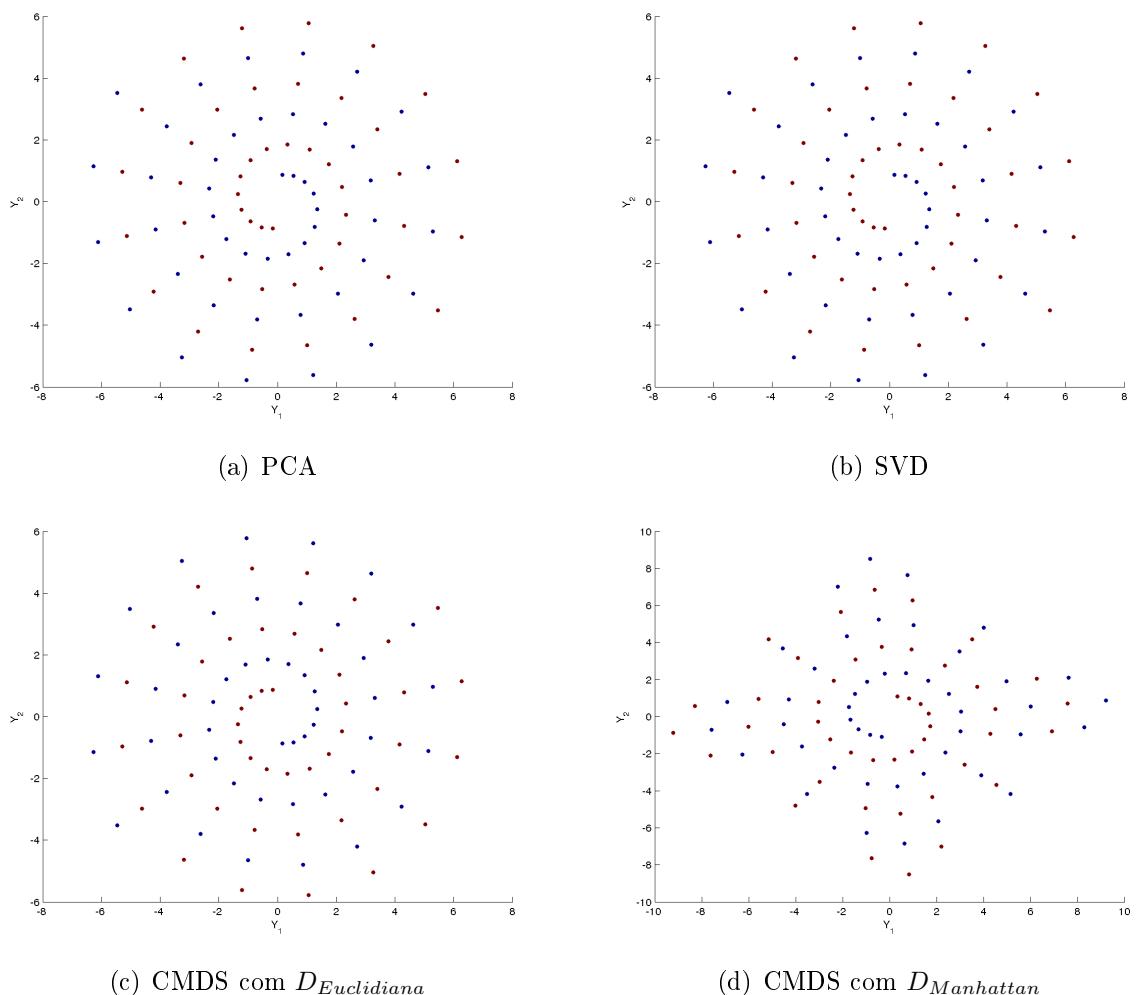
Figura C.19: Gráfico de dispersão para as representações bidimensionais a partir da base Sonar não padronizada

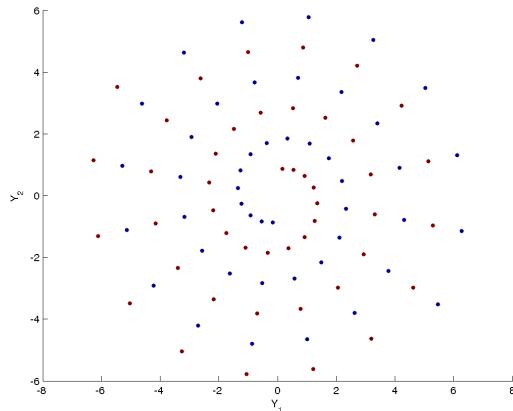
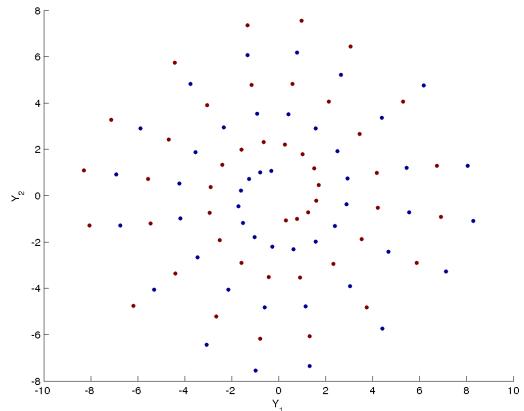
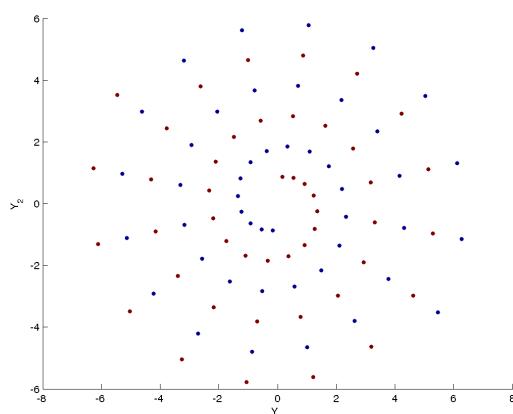
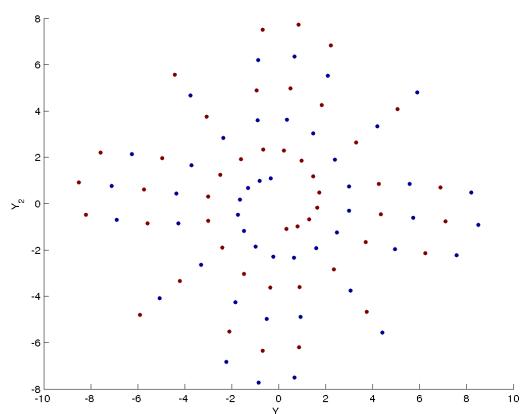
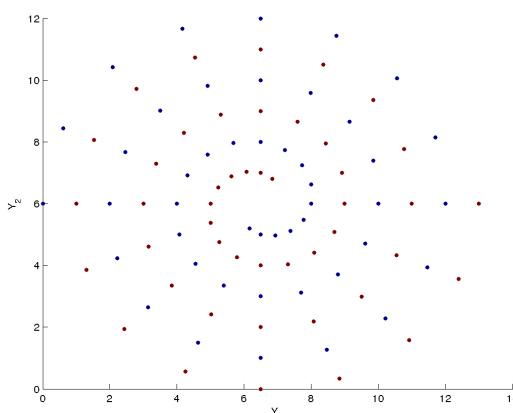


(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

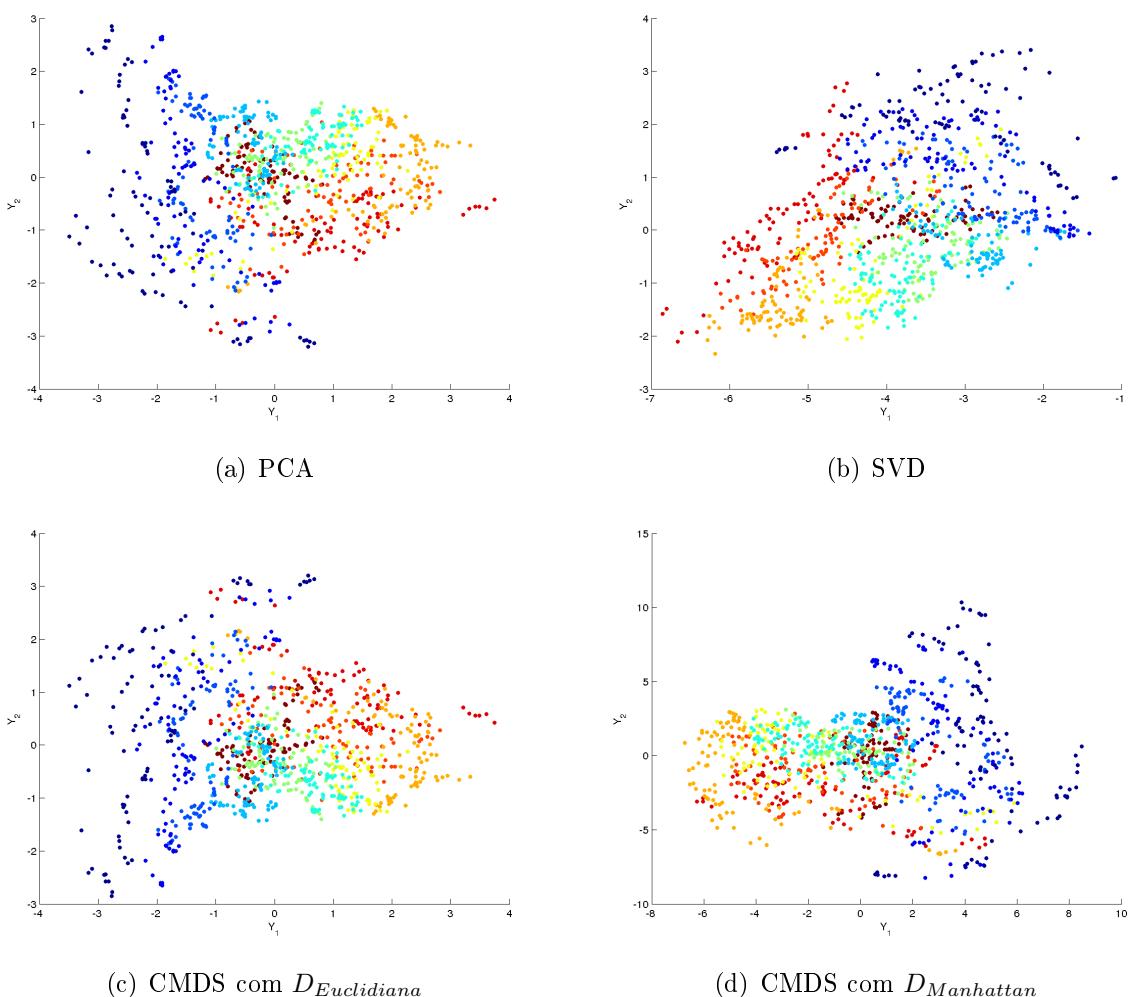
Figura C.20: Gráfico de dispersão para as representações bidimensionais a partir da base Espiral não padronizada

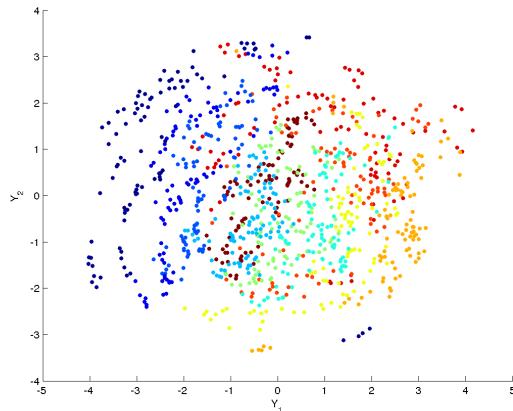
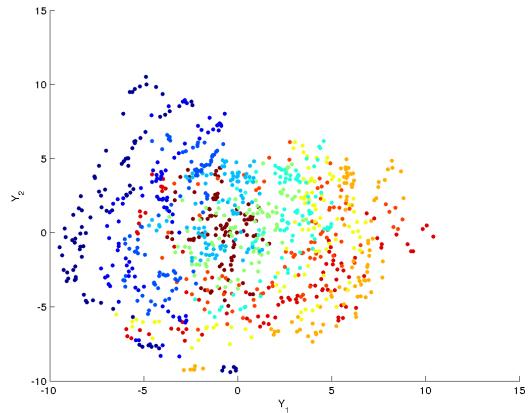
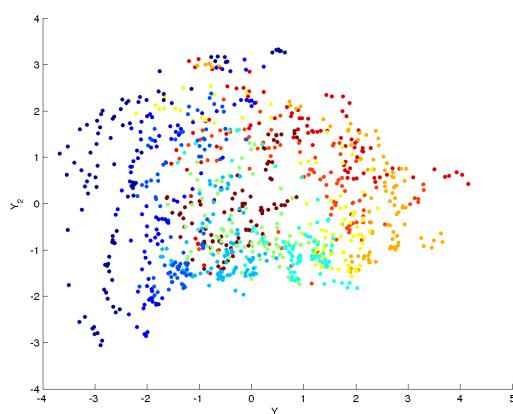
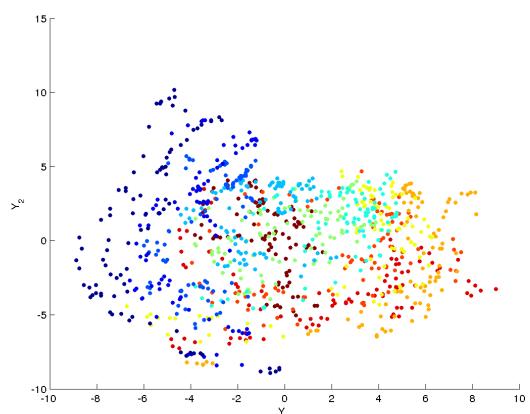
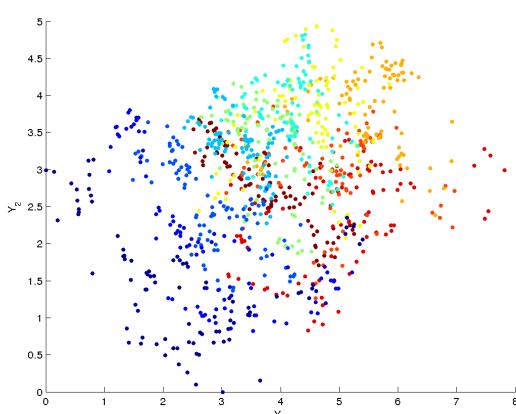


(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

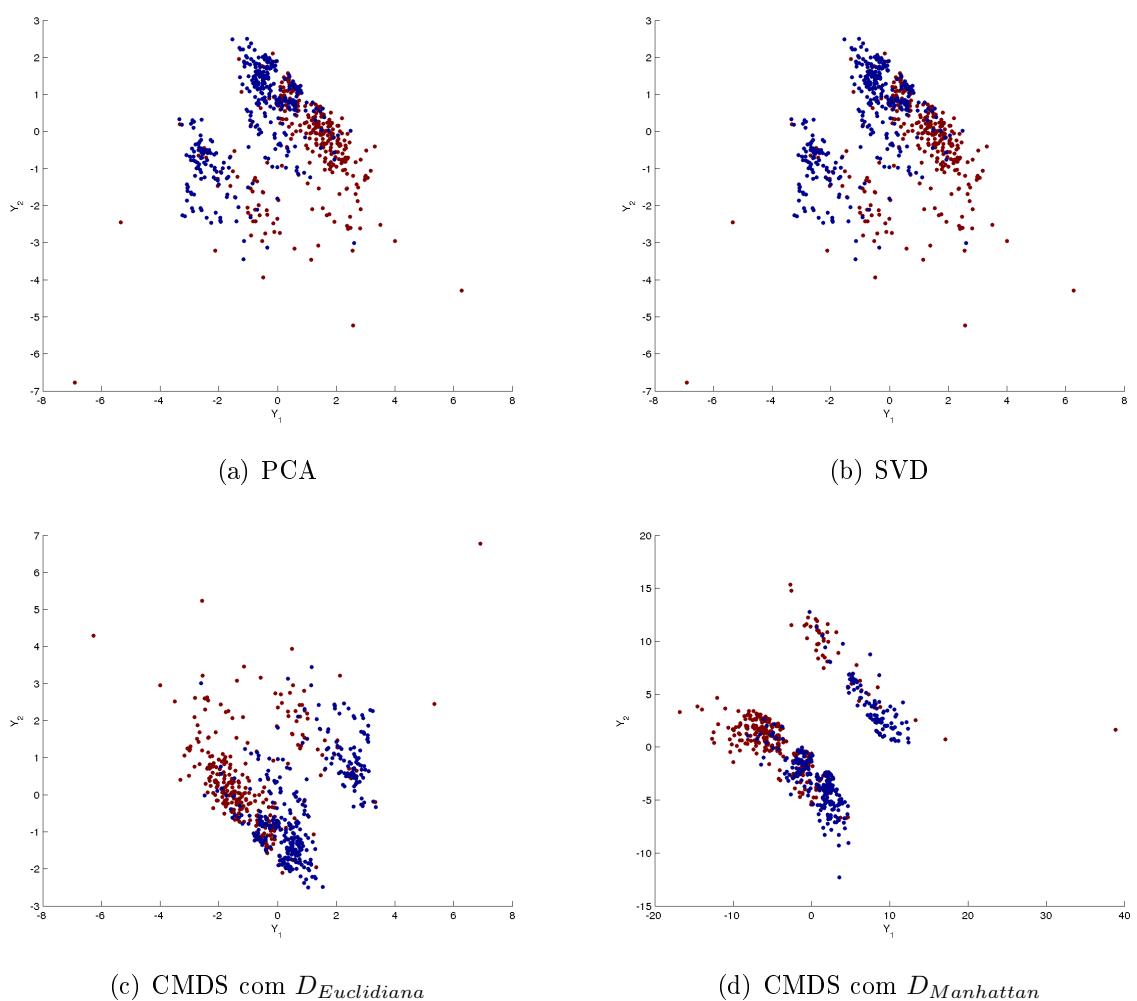
Figura C.21: Gráfico de dispersão para as representações bidimensionais a partir da base Vogal não padronizada

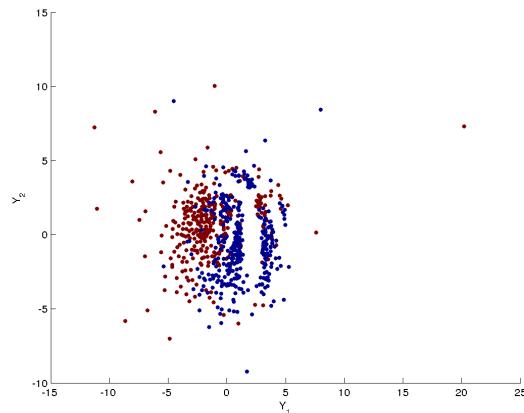
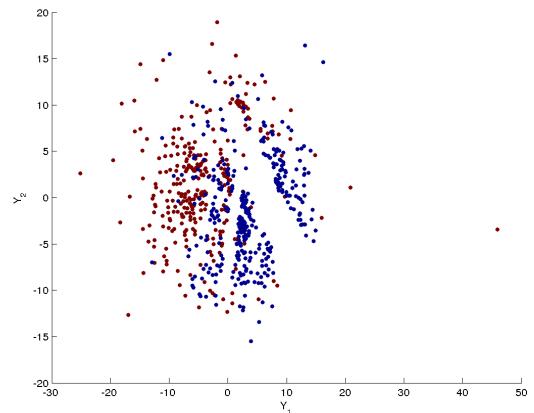
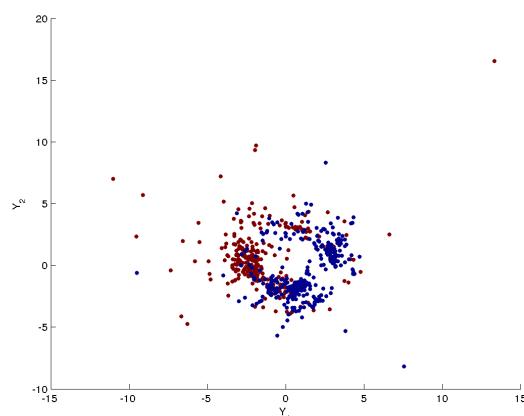
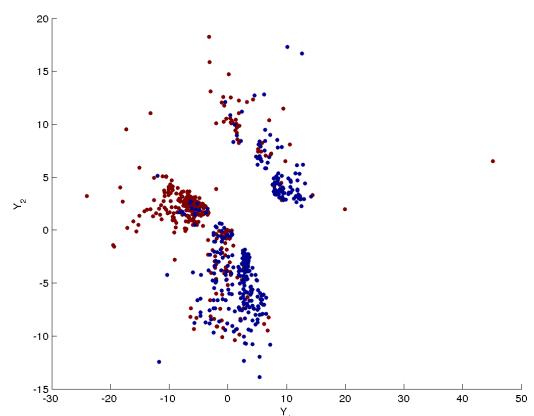
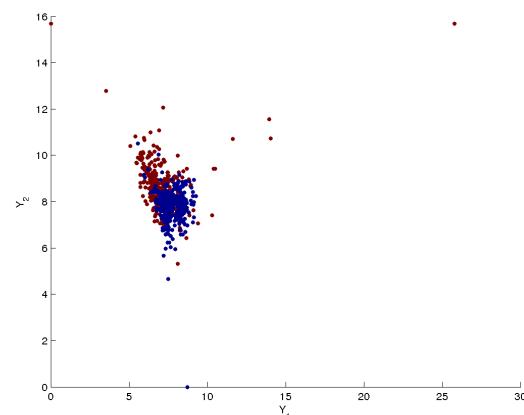


(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

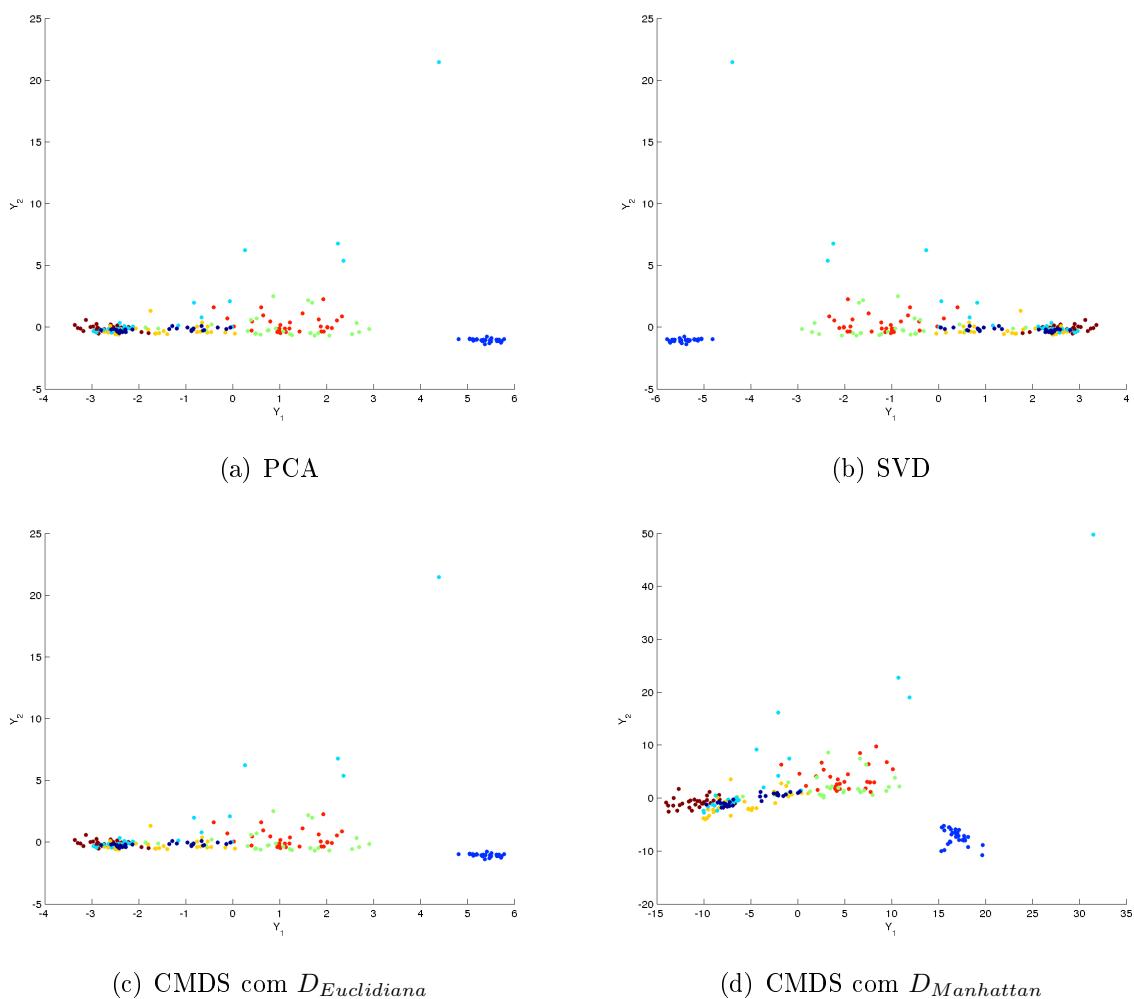
Figura C.22: Gráfico de dispersão para as representações bidimensionais a partir da base Triagem de Crédito padronizada

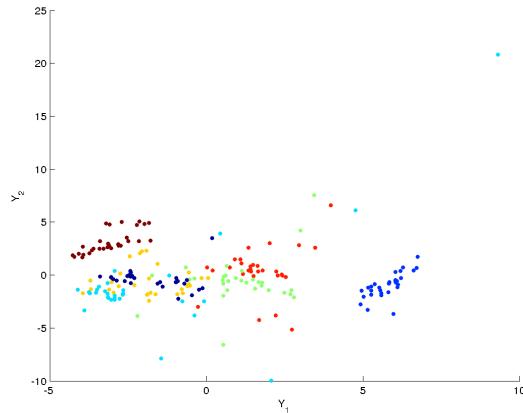
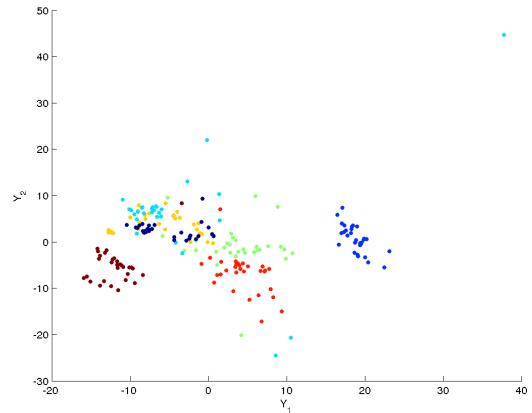
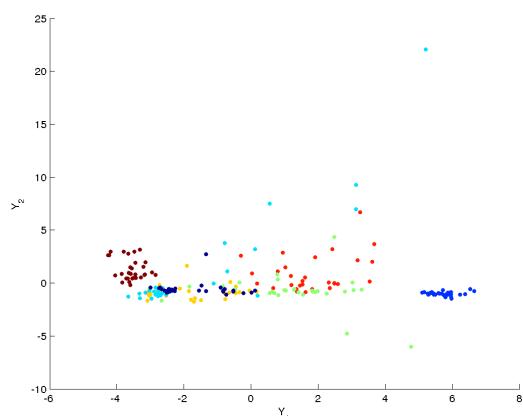
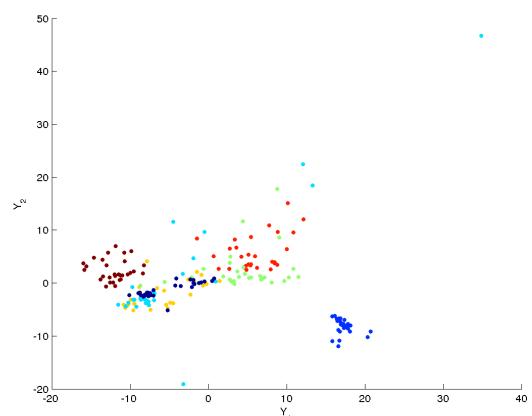
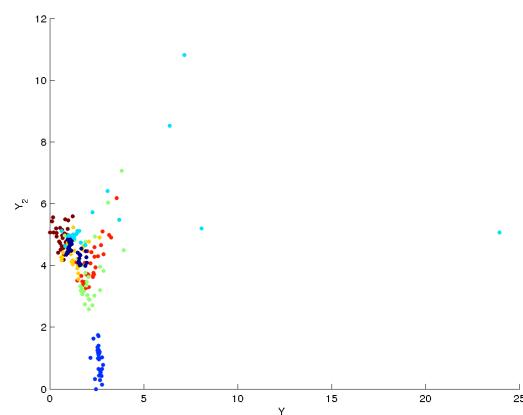


(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

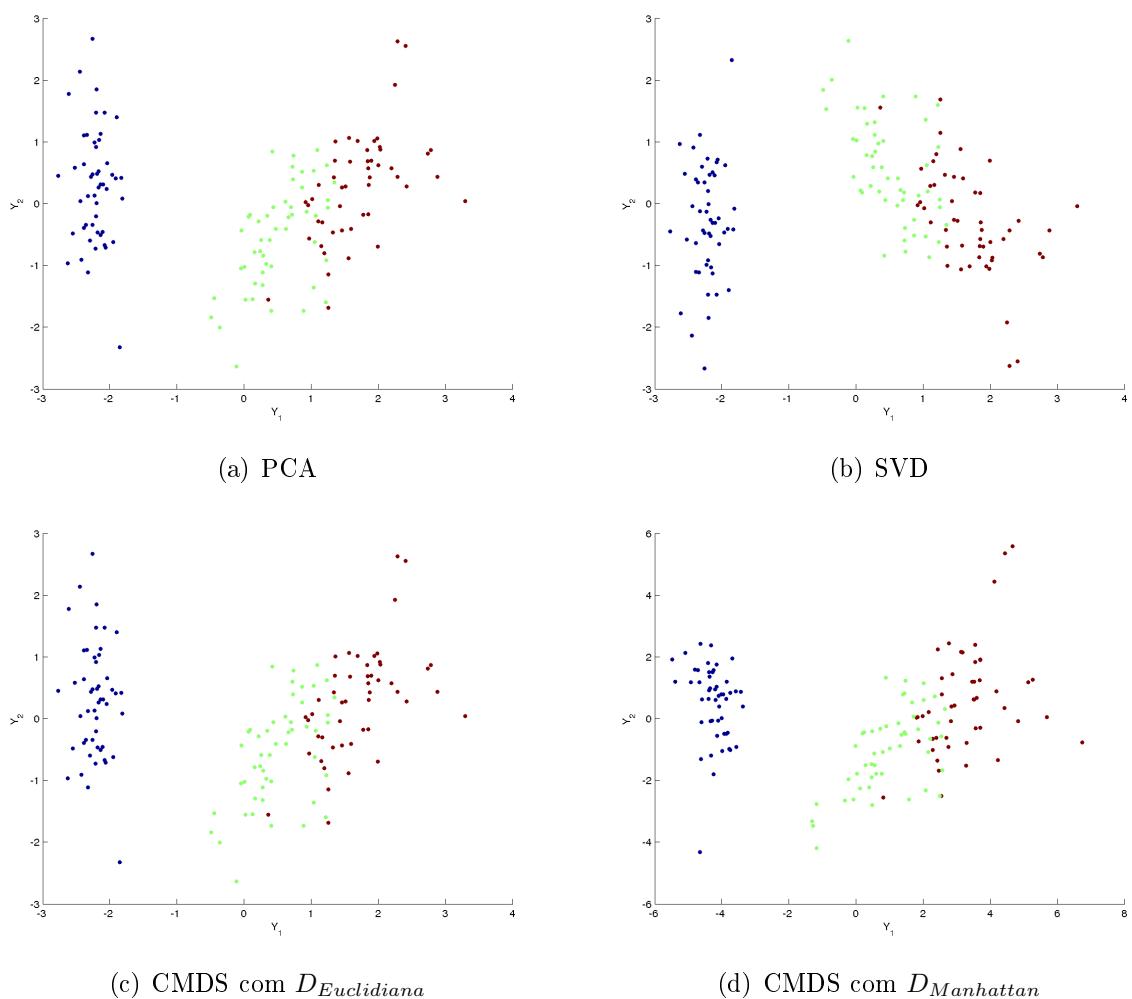
Figura C.23: Gráfico de dispersão para as representações bidimensionais a partir da base Imagens padronizada

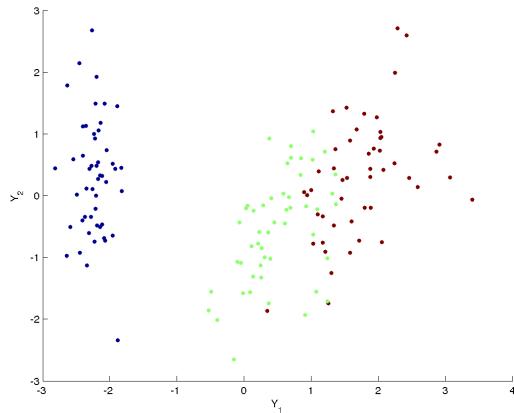
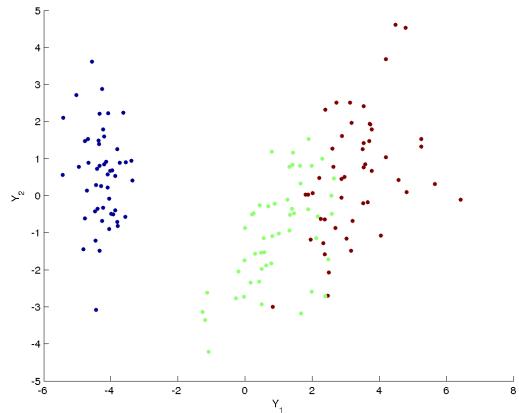
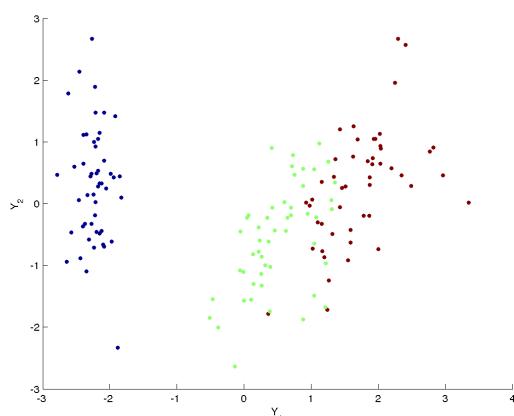
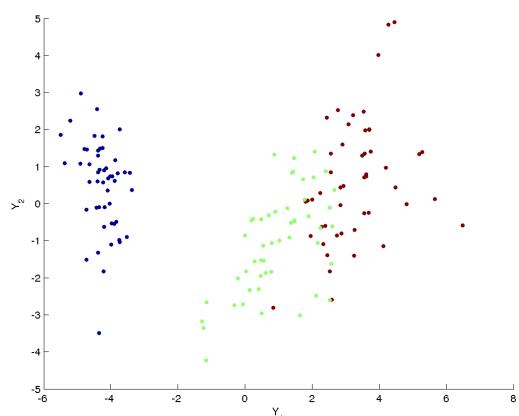
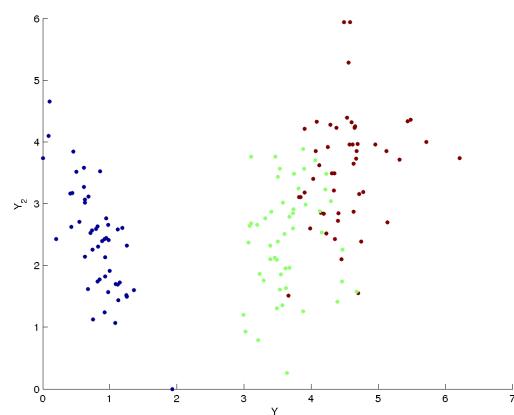


(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

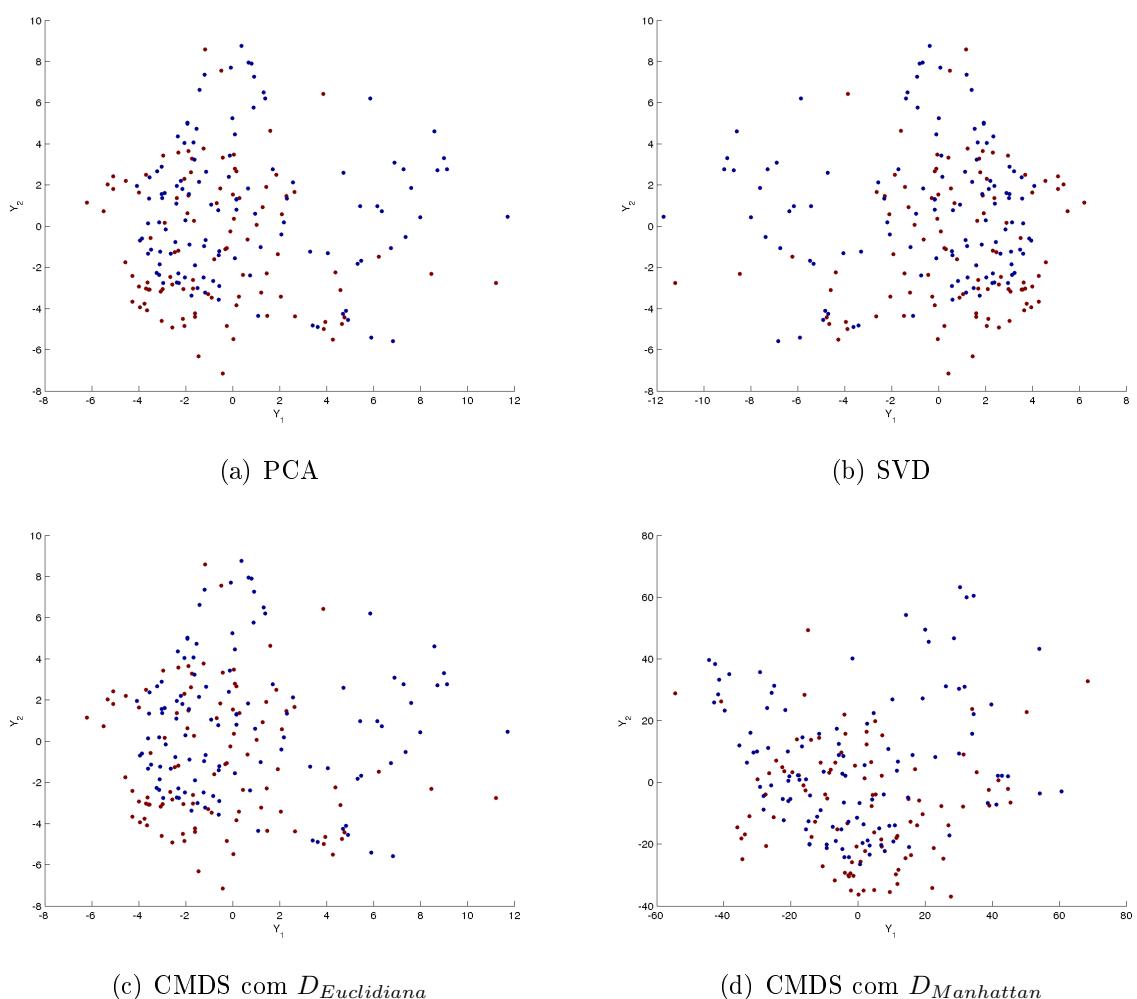
Figura C.24: Gráfico de dispersão para as representações bidimensionais a partir da base Iris padronizada

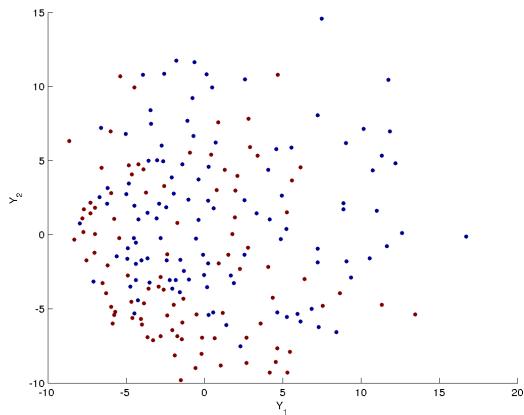
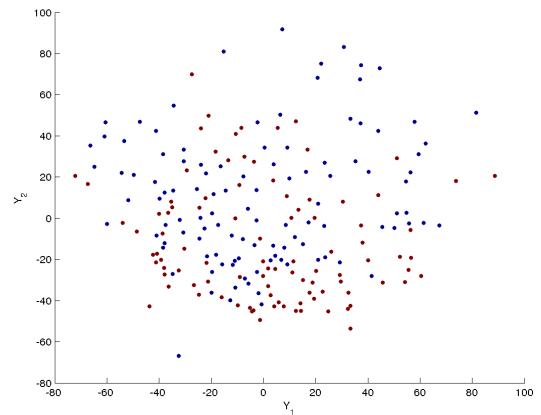
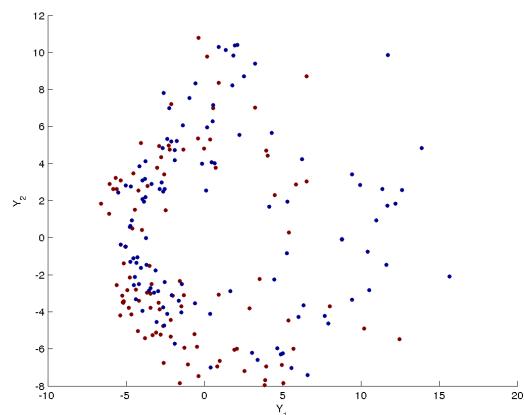
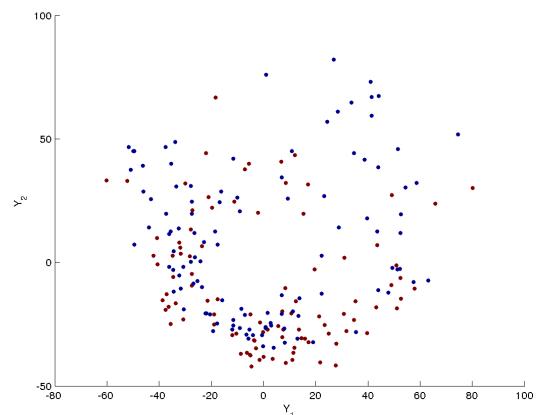
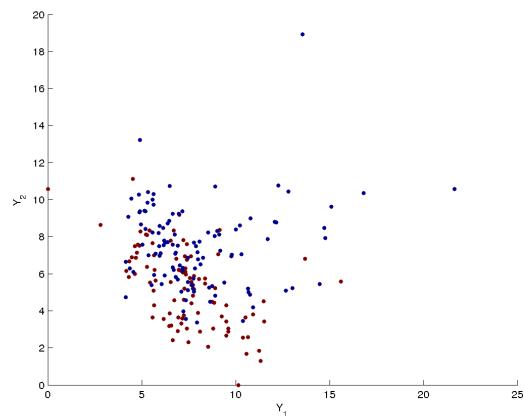


(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

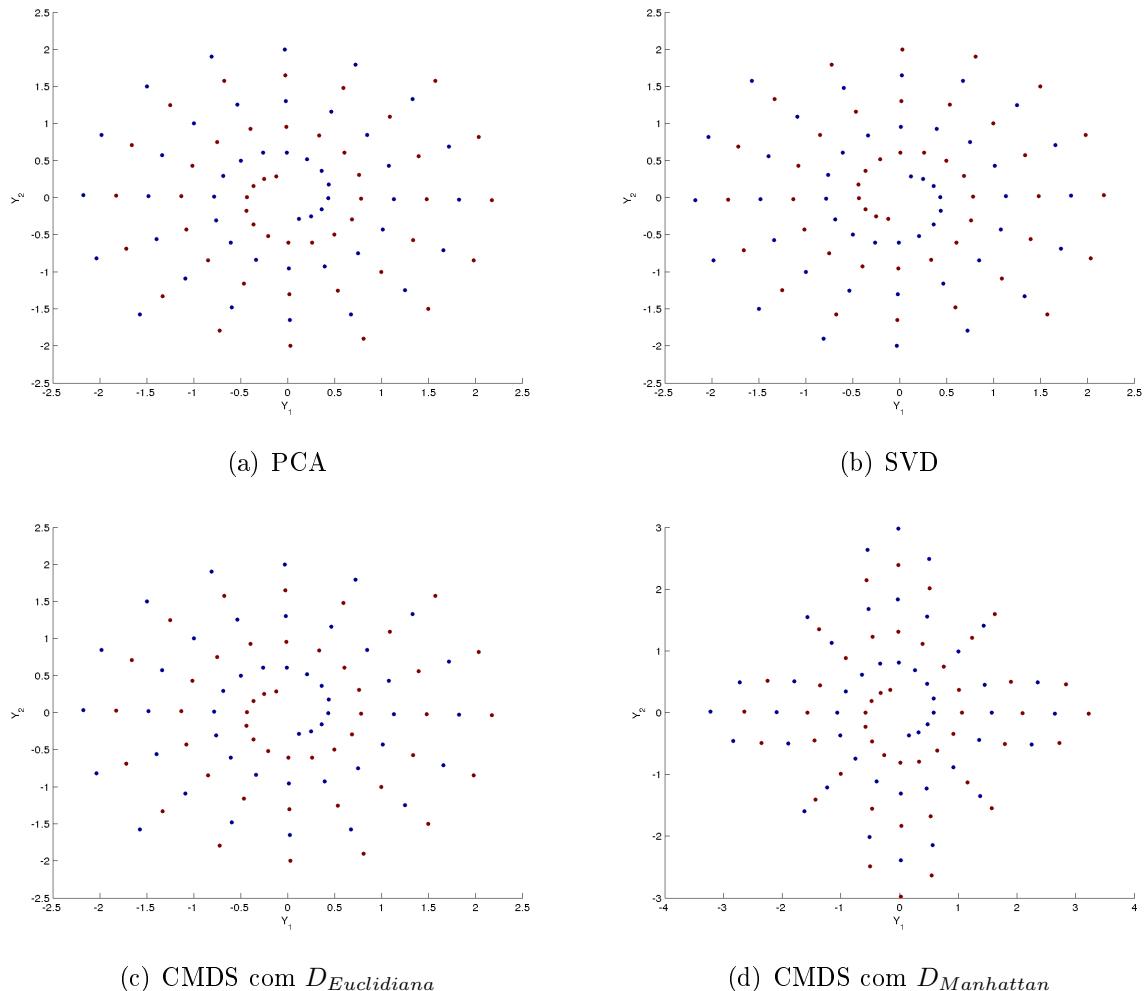
Figura C.25: Gráfico de dispersão para as representações bidimensionais a partir da base Sonar padronizada

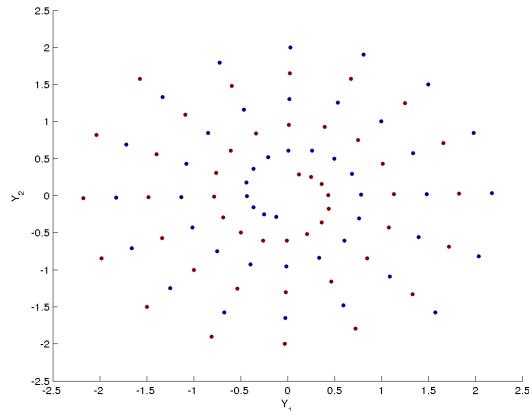
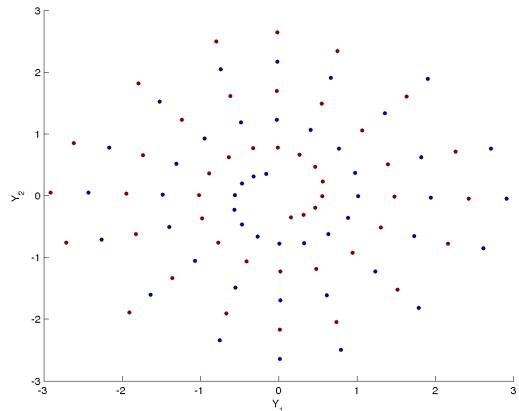
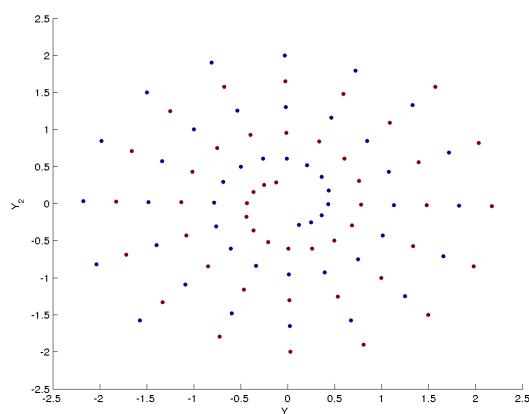
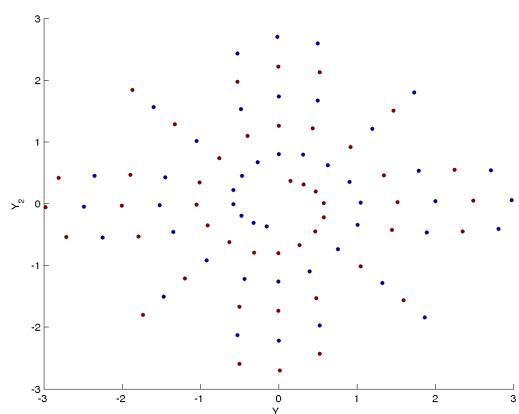
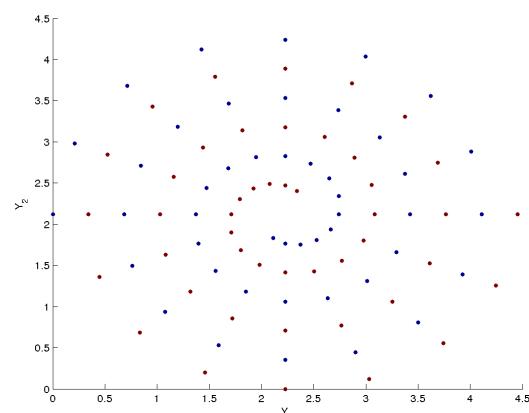


(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

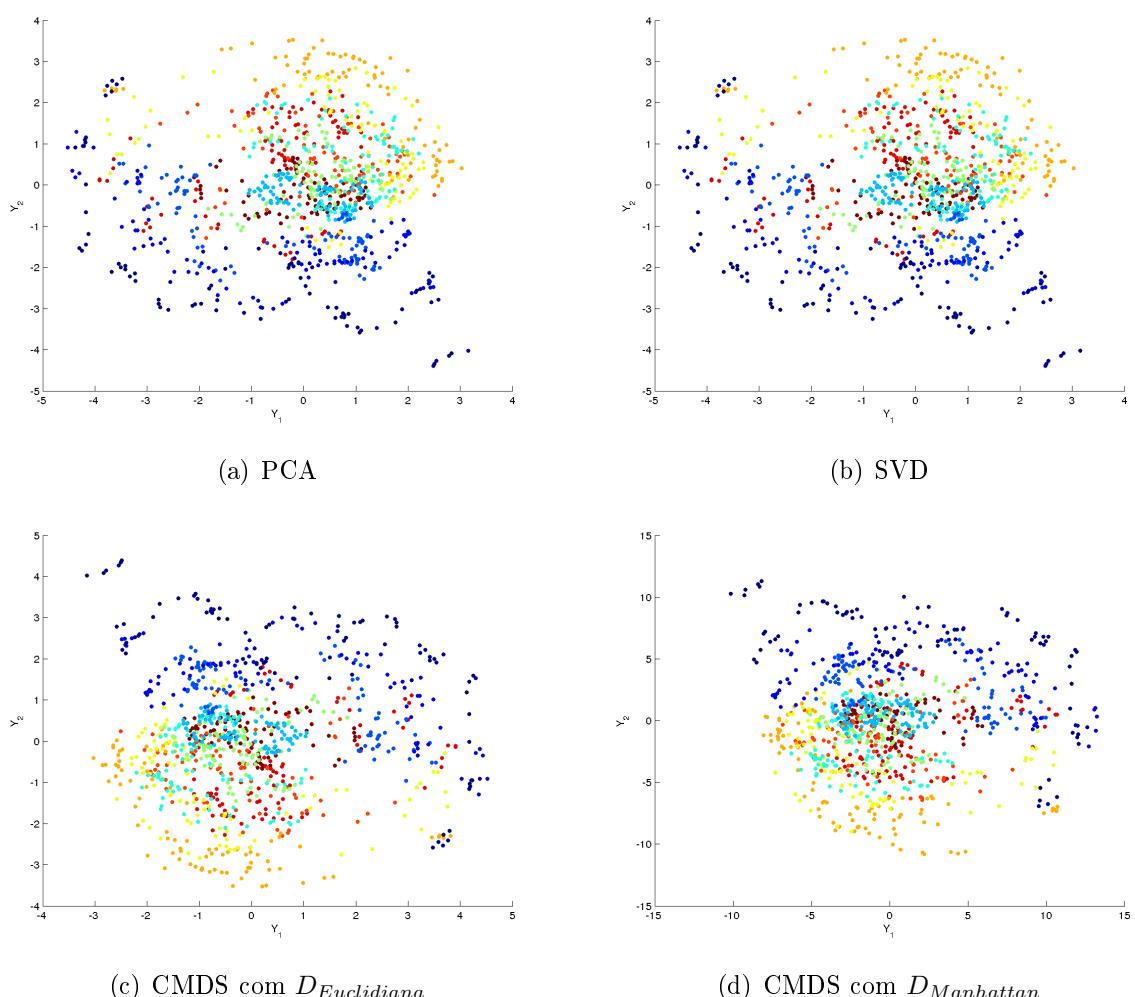
Figura C.26: Gráfico de dispersão para as representações bidimensionais a partir da base Espiral padronizada

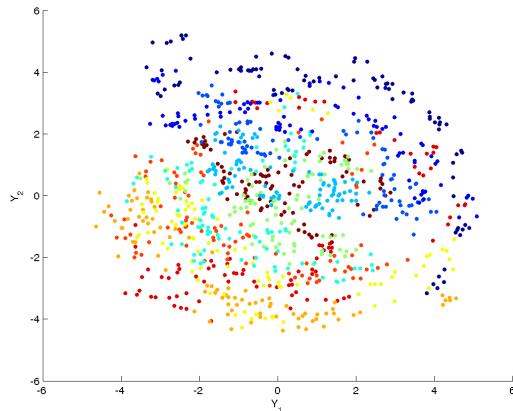
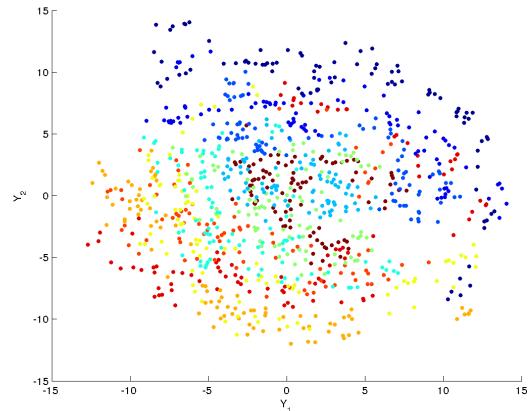
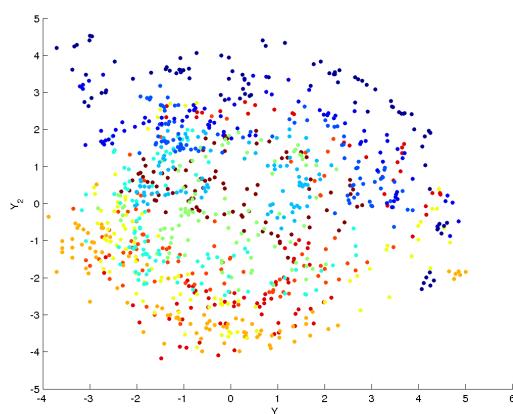
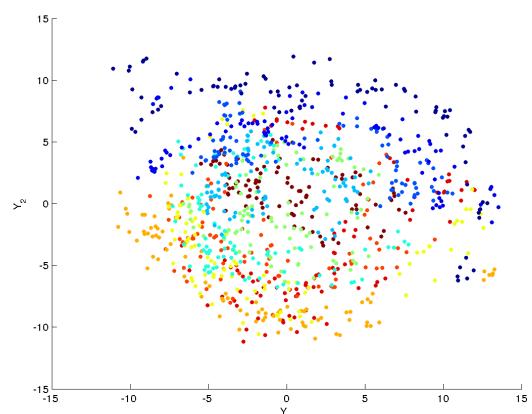
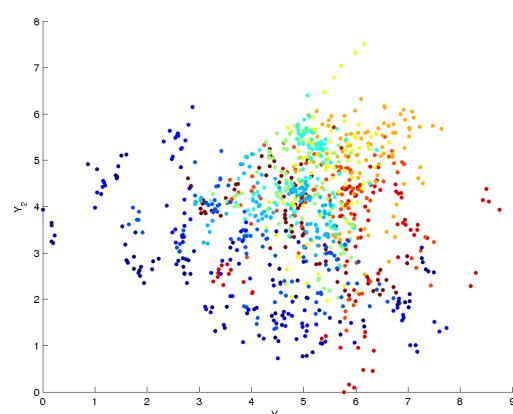


(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

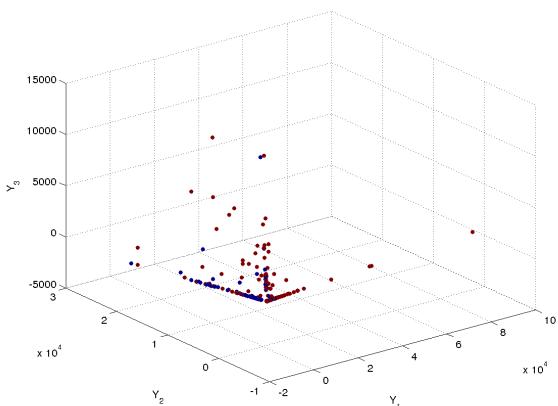
Figura C.27: Gráfico de dispersão para as representações bidimensionais a partir da base Vogal padronizada



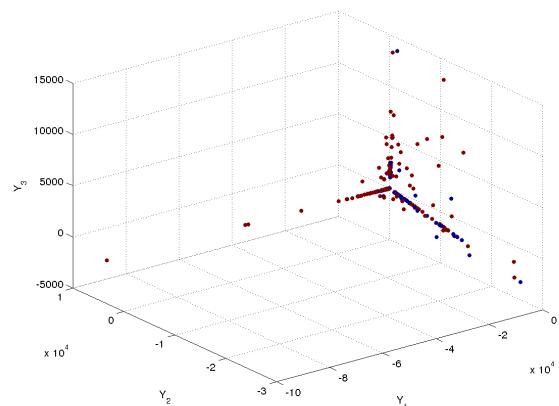
(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

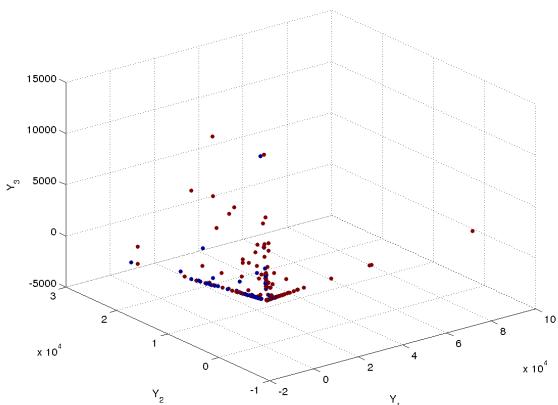
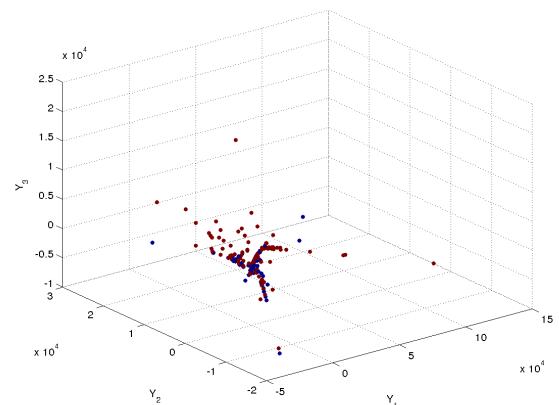
Figura C.28: Gráfico de dispersão para as representações tridimensionais a partir da base Triagem de Crédito não padronizada

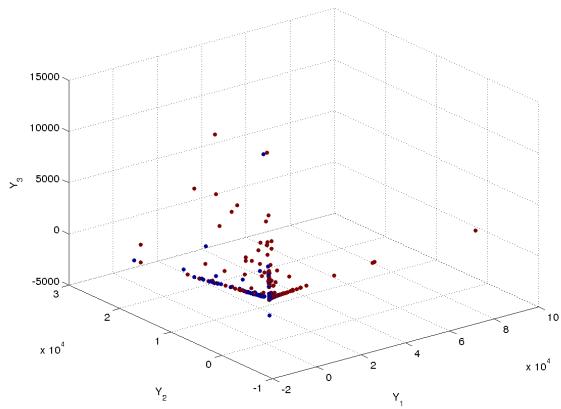
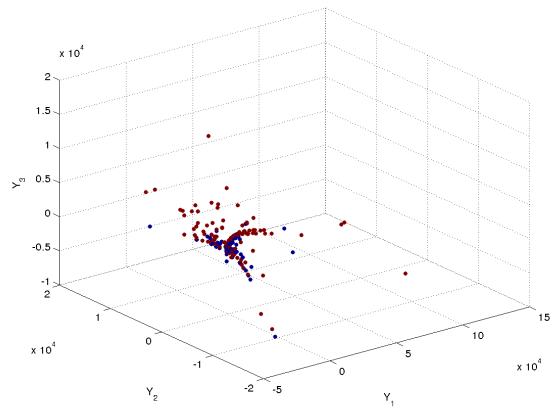
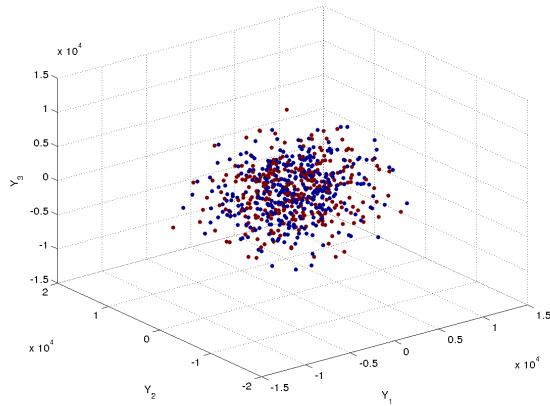
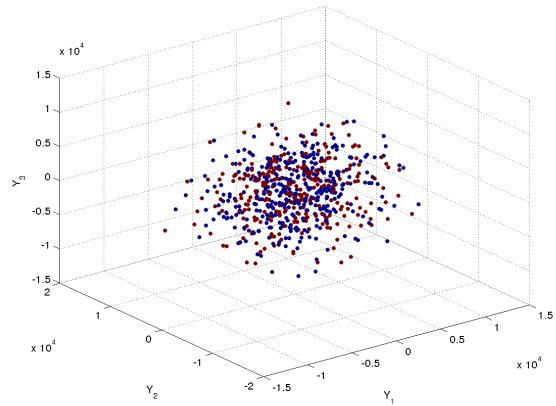
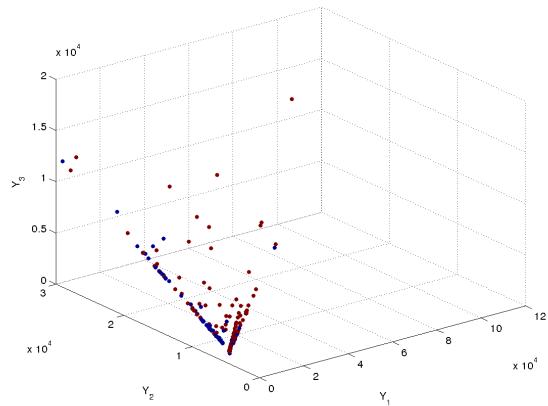


(a) PCA



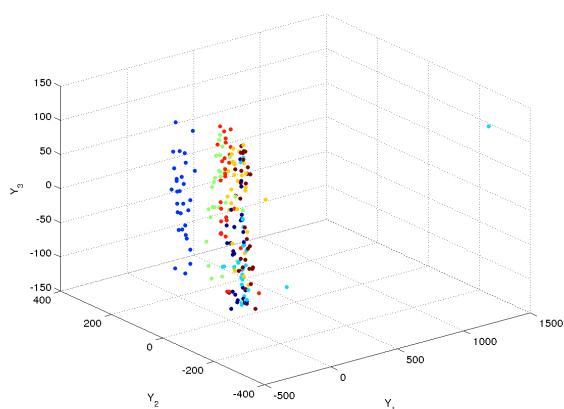
(b) SVD

(c) CMDS com $D_{Euclidiana}$ (d) CMDS com $D_{Manhattan}$

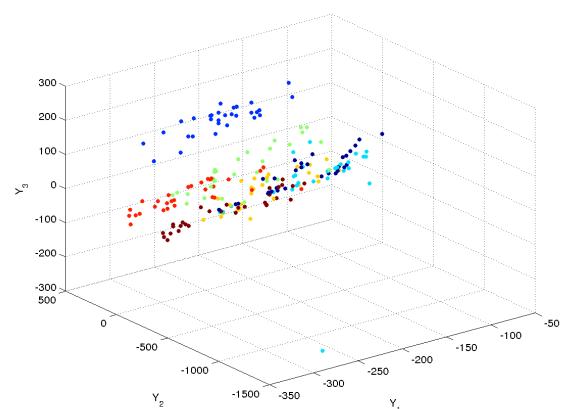
(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

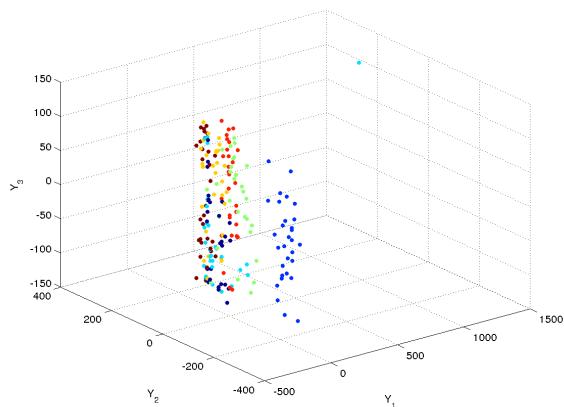
Figura C.29: Gráfico de dispersão para as representações tridimensionais a partir da base Imagens não padronizada



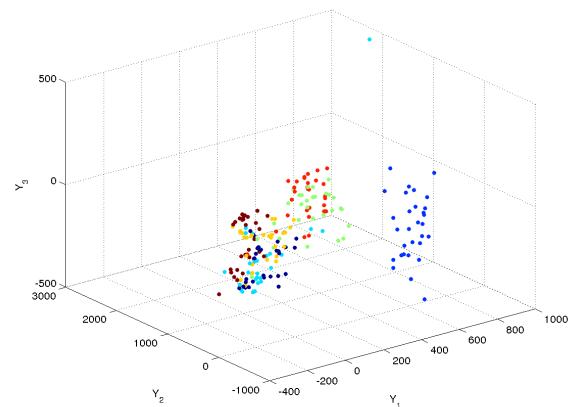
(a) PCA



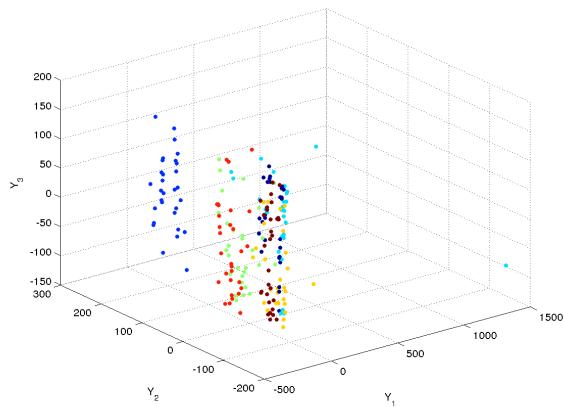
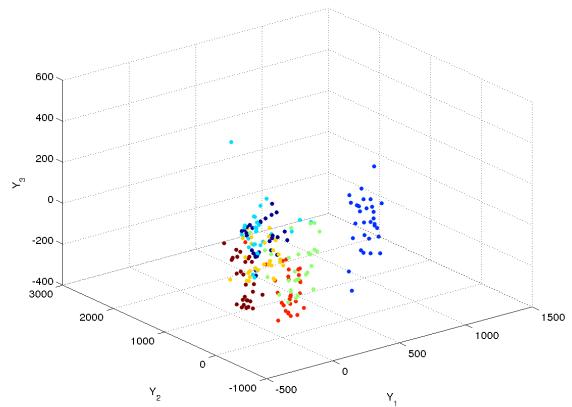
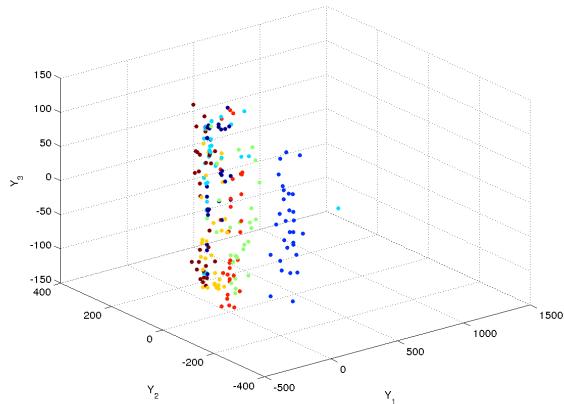
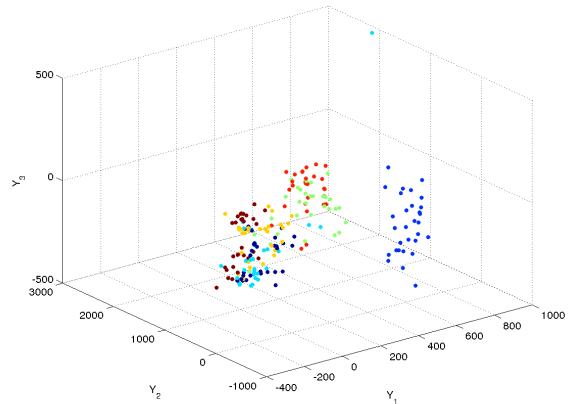
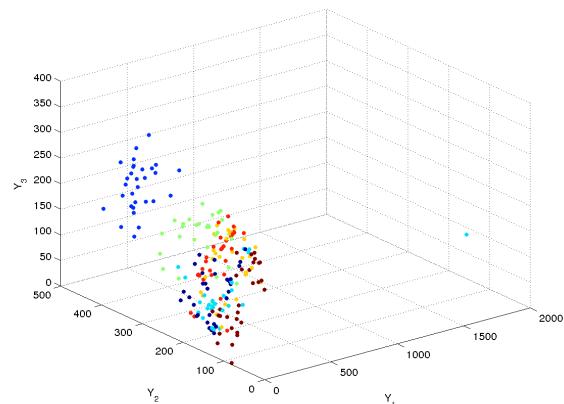
(b) SVD



(c) CMDS com $D_{Eucladiana}$

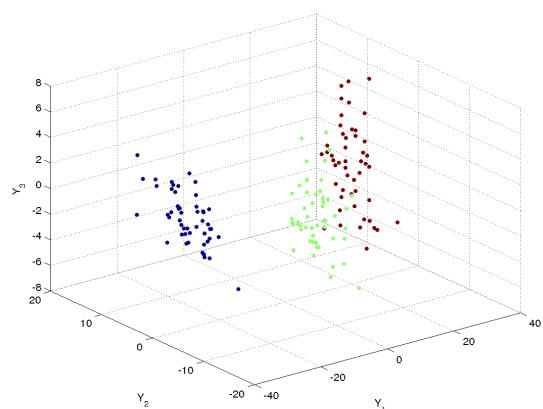


(d) CMDS com $D_{Manhattan}$

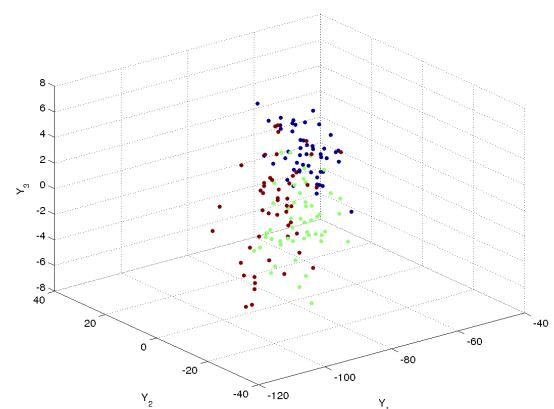
(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

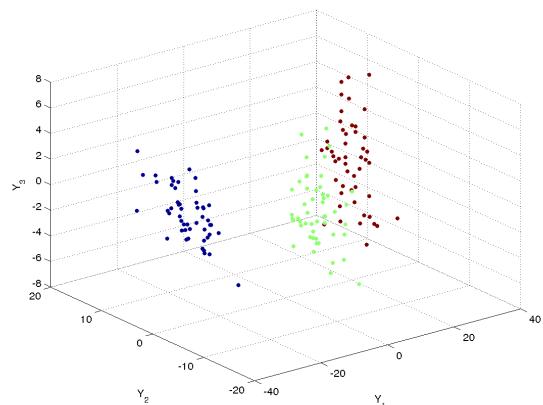
Figura C.30: Gráfico de dispersão para as representações tridimensionais a partir da base Iris não padronizada



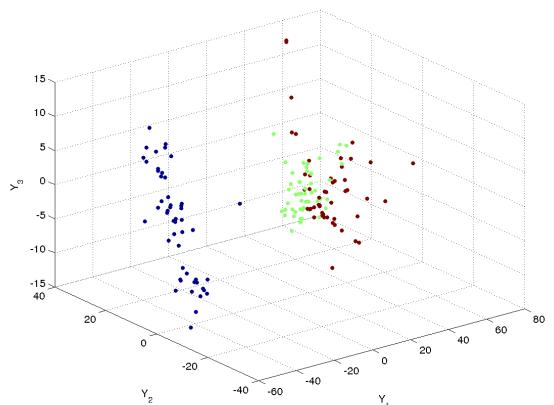
(a) PCA



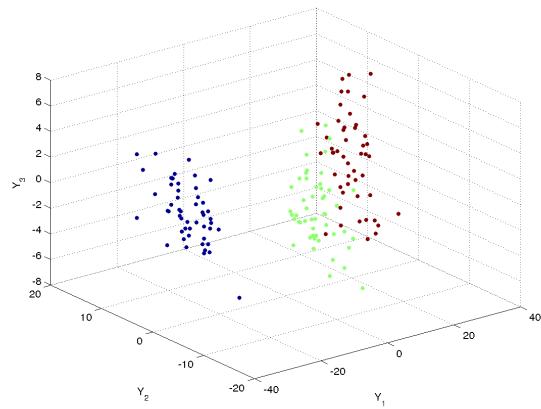
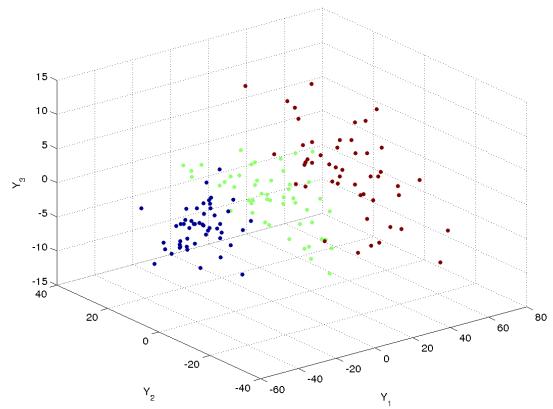
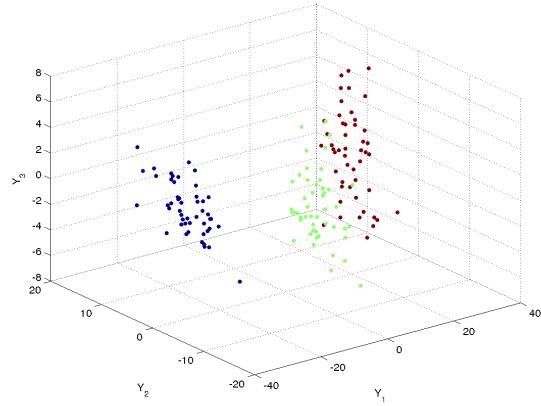
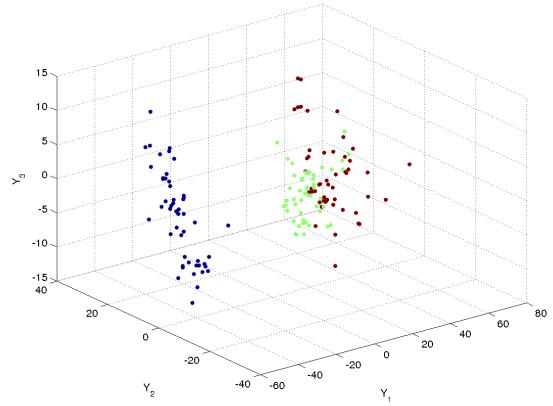
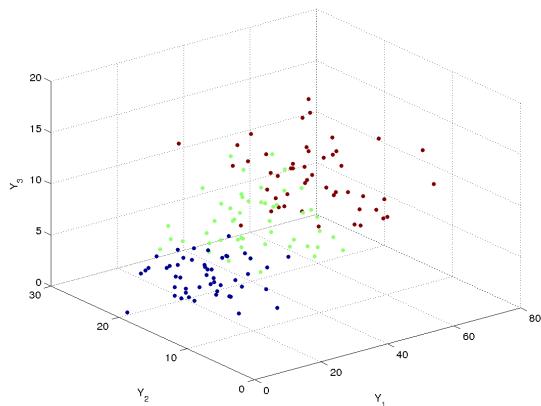
(b) SVD



(c) CMDS com $D_{Euclidiana}$

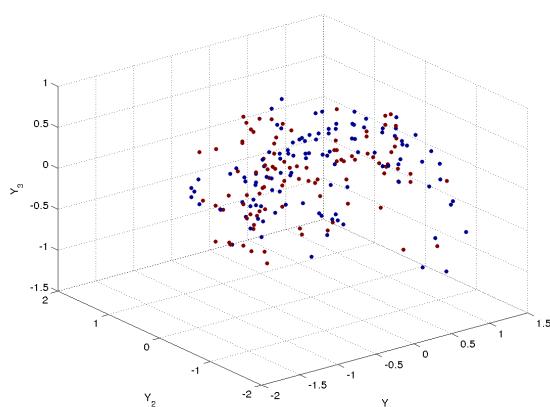


(d) CMDS com $D_{Manhattan}$

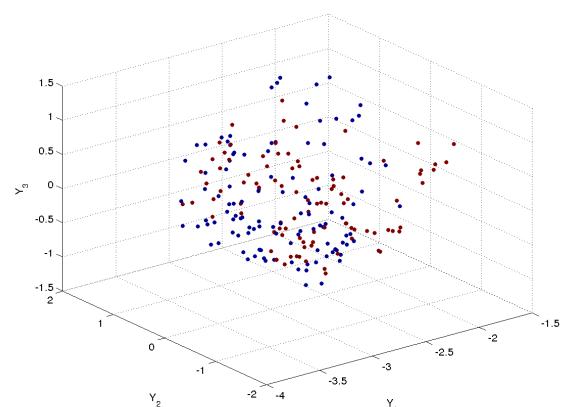
(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

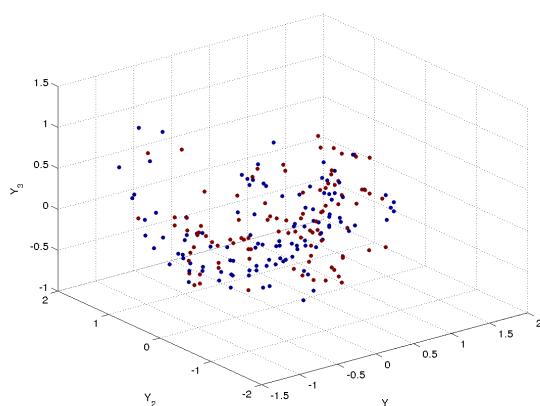
Figura C.31: Gráfico de dispersão para as representações tridimensionais a partir da base Sonar não padronizada



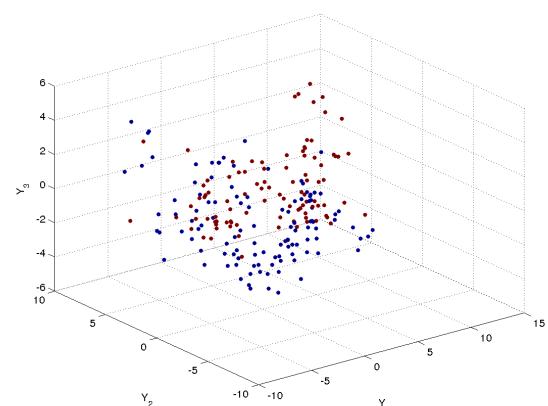
(a) PCA



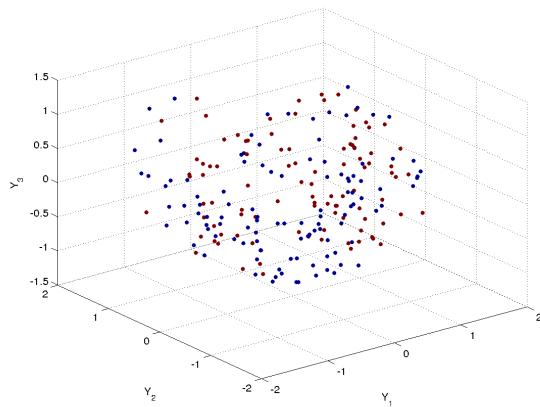
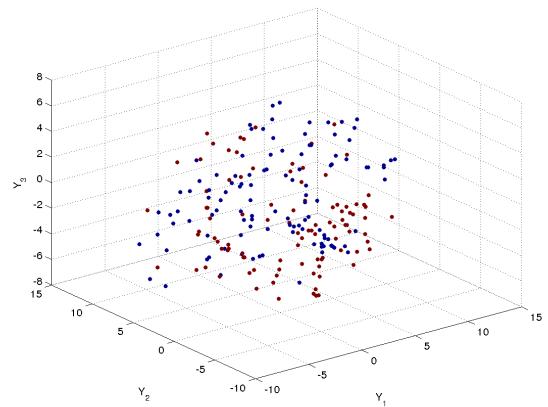
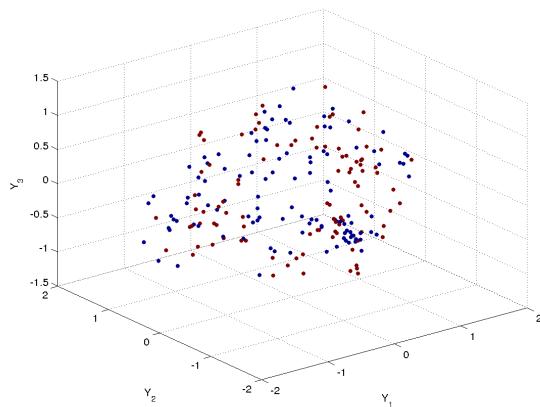
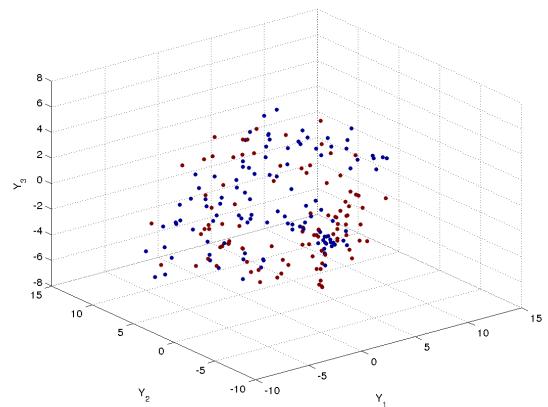
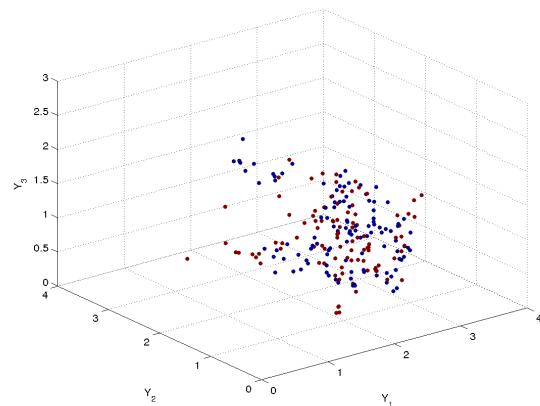
(b) SVD



(c) CMDS com $D_{Eucladiana}$

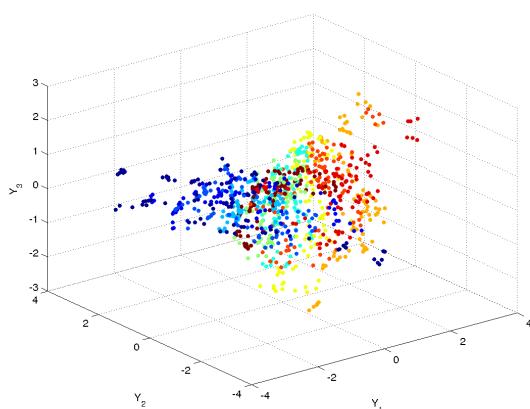


(d) CMDS com $D_{Manhattan}$

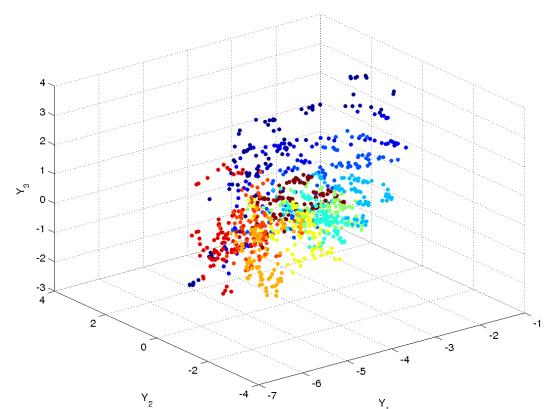
(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

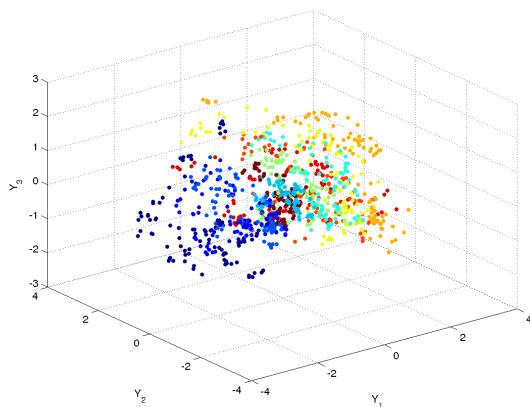
Figura C.32: Gráfico de dispersão para as representações tridimensionais a partir da base Vogal não padronizada



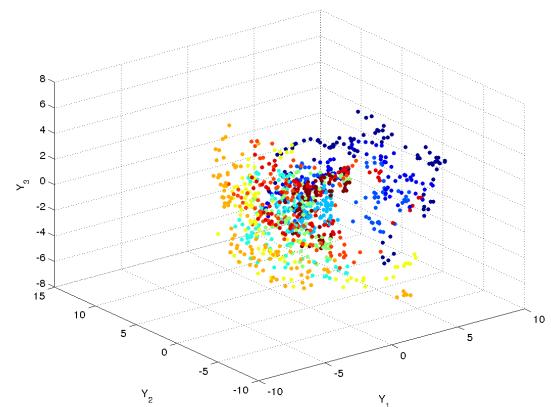
(a) PCA



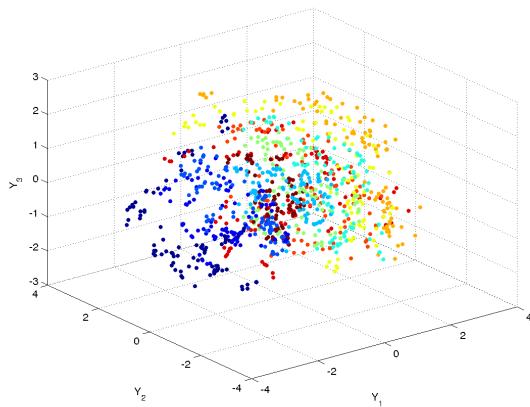
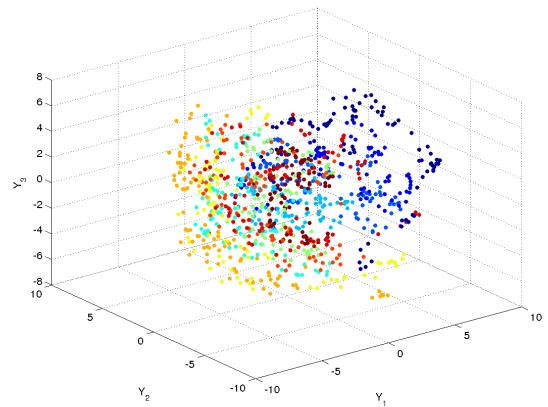
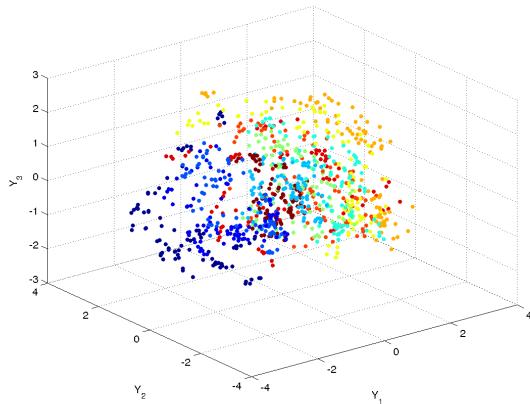
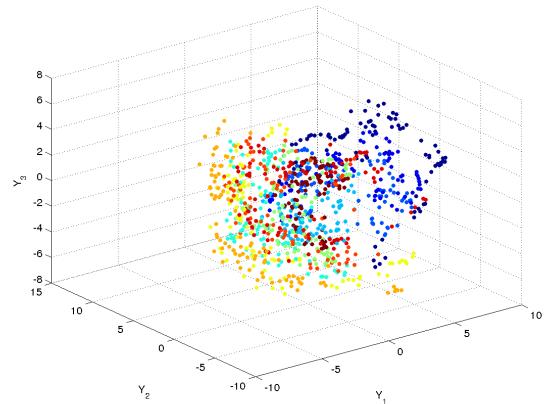
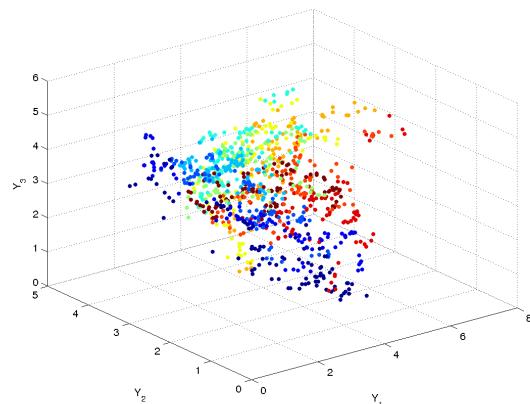
(b) SVD



(c) CMDS com $D_{Euclidiana}$

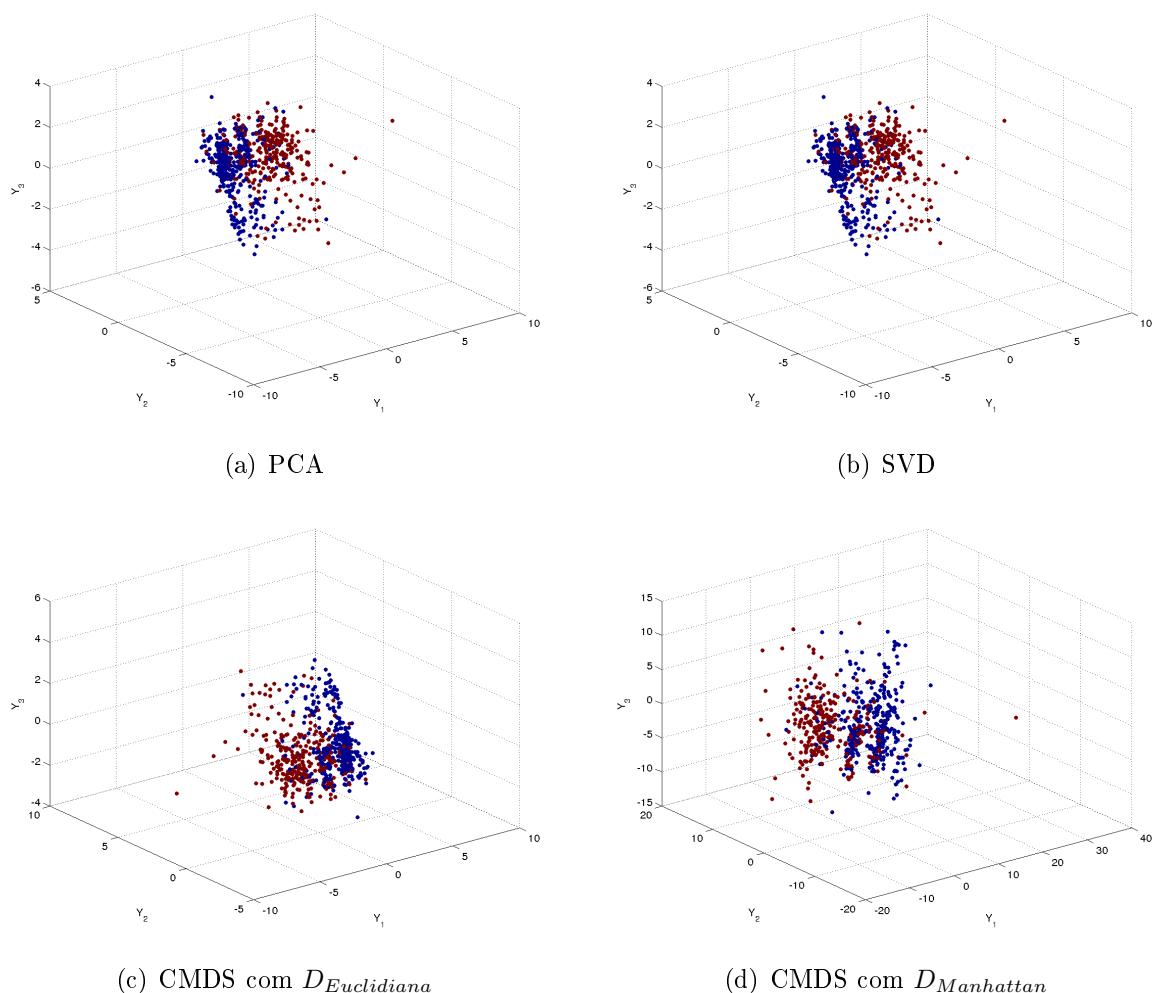


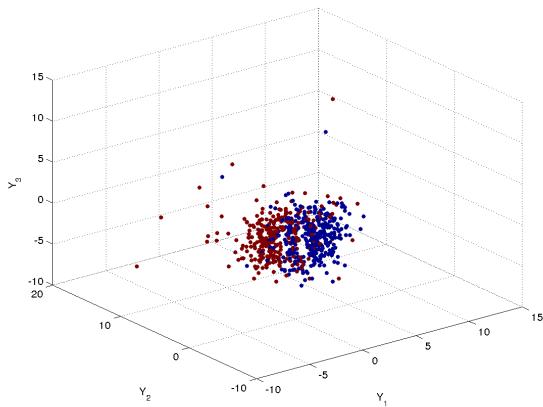
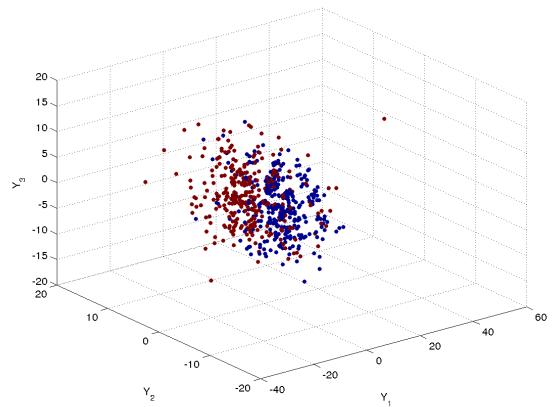
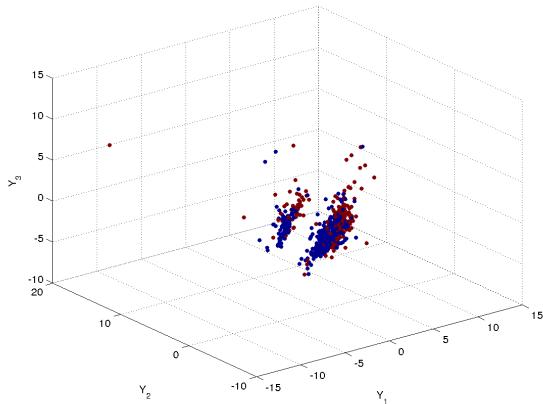
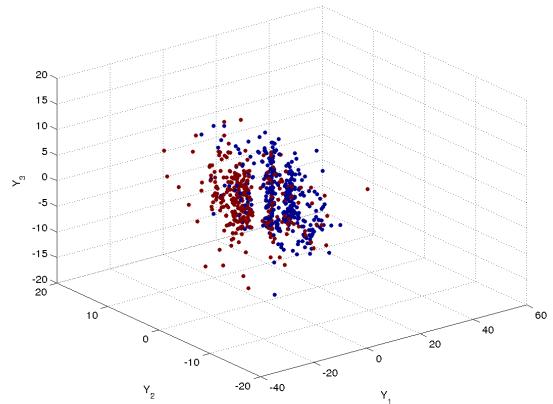
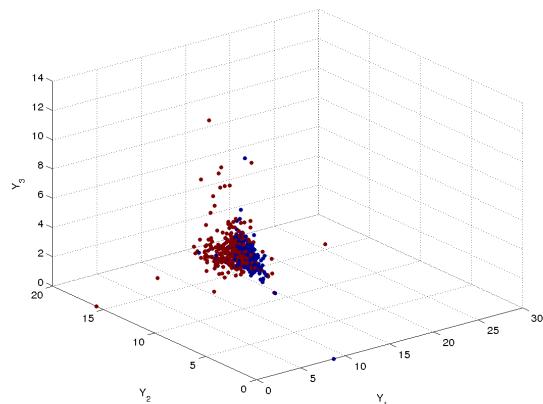
(d) CMDS com $D_{Manhattan}$

(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

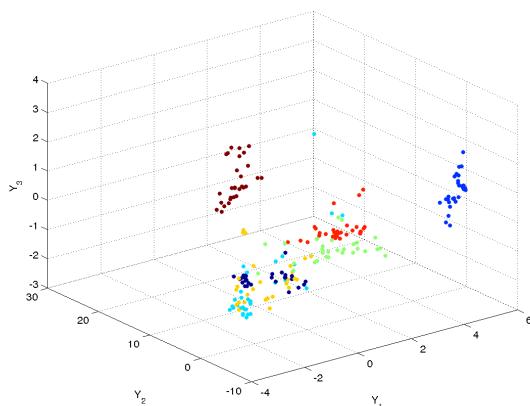
Figura C.33: Gráfico de dispersão para as representações tridimensionais a partir da base Triagem de Crédito padronizada



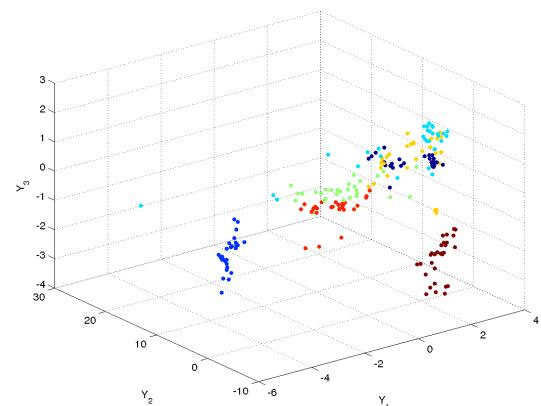
(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

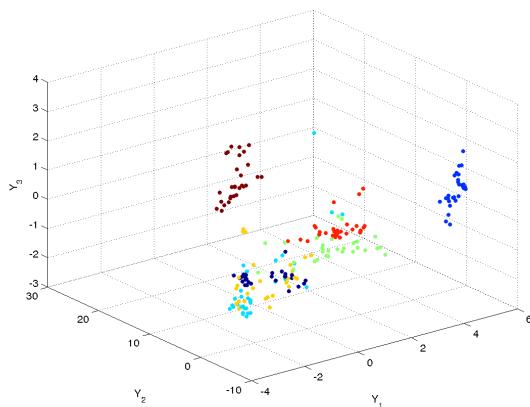
Figura C.34: Gráfico de dispersão para as representações tridimensionais a partir da base Imagens padronizada



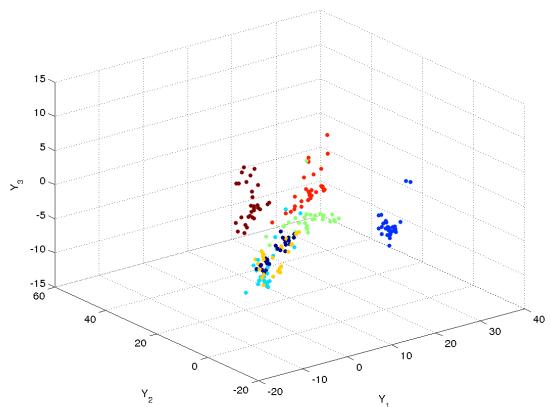
(a) PCA



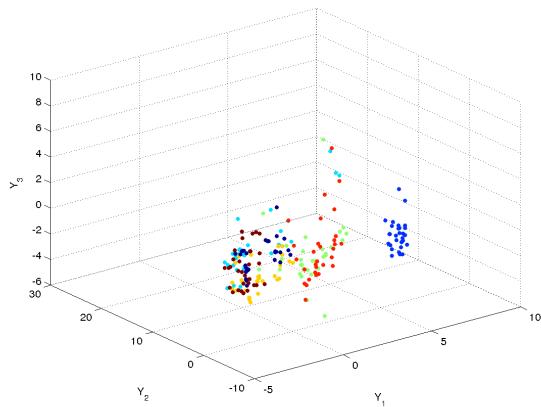
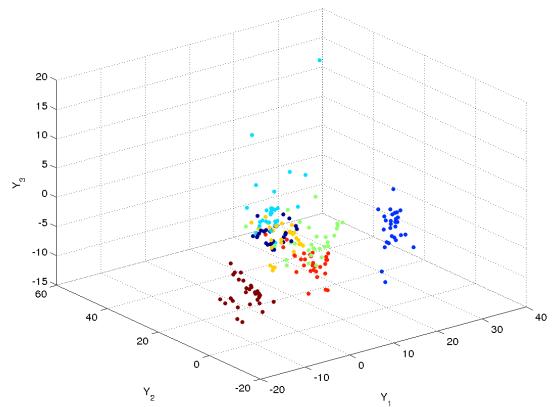
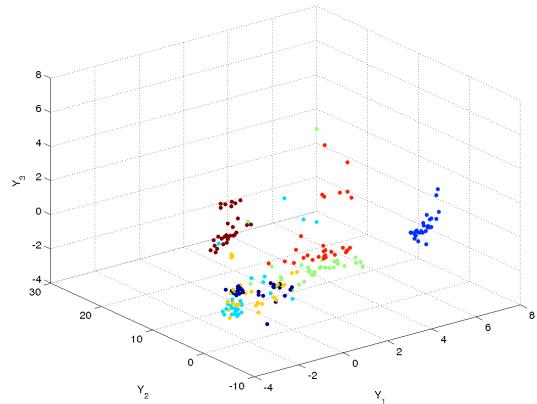
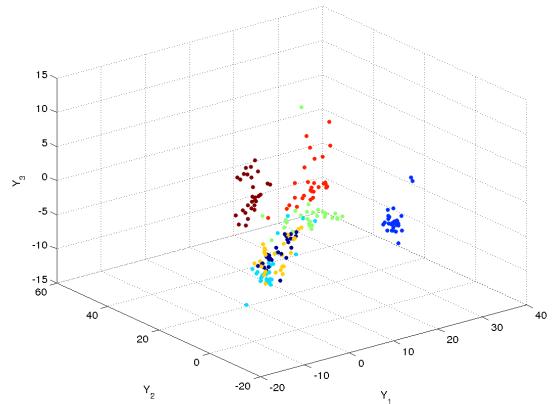
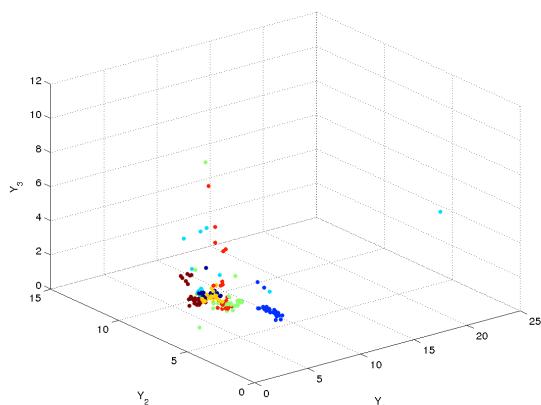
(b) SVD



(c) CMDS com $D_{Eucladiana}$

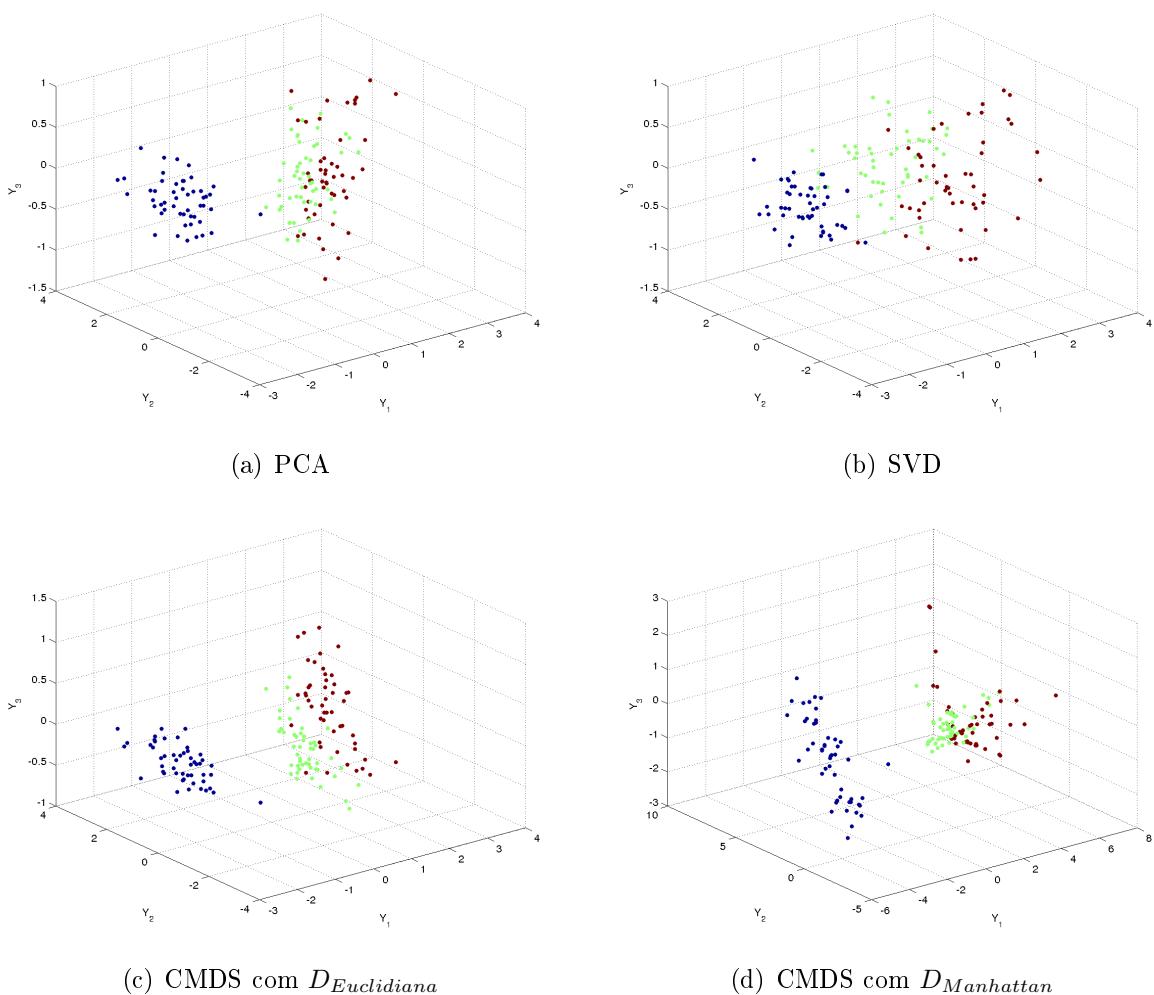


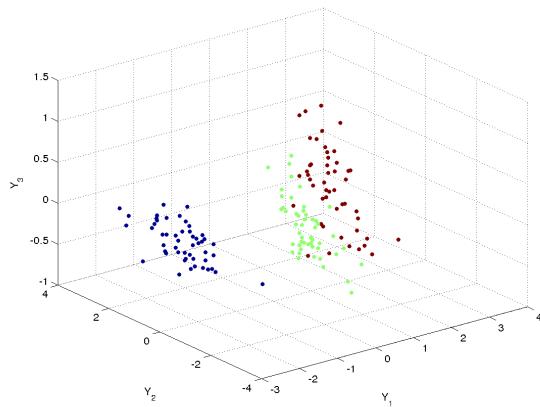
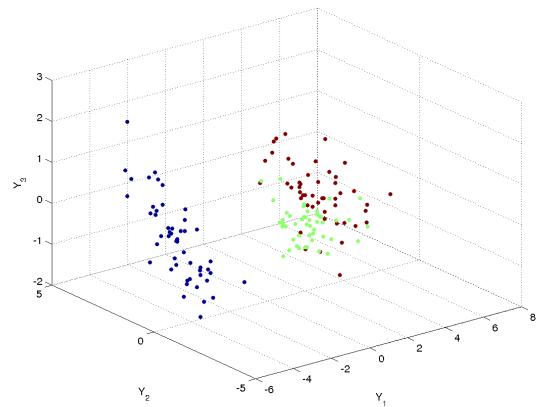
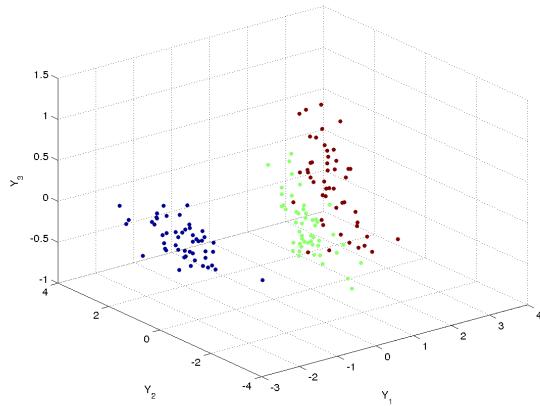
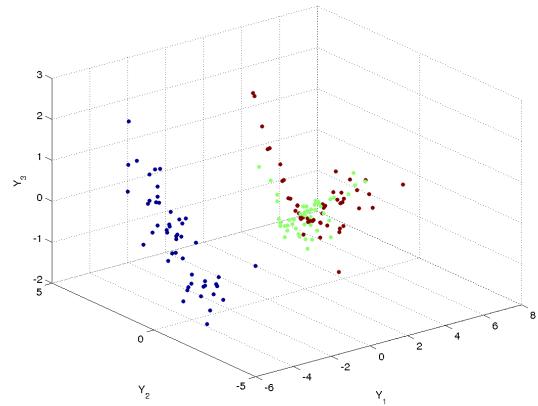
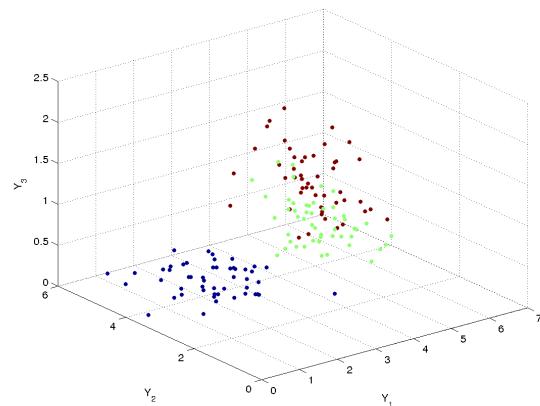
(d) CMDS com $D_{Manhattan}$

(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

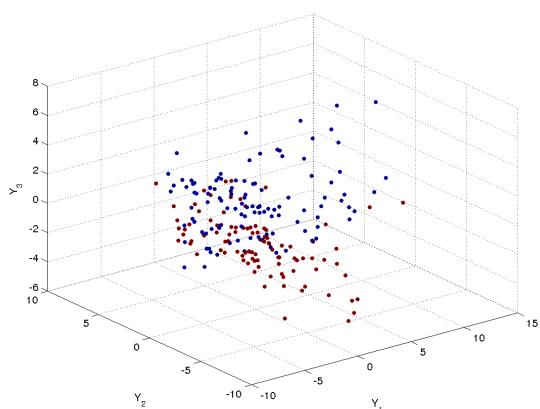
Figura C.35: Gráfico de dispersão para as representações tridimensionais a partir da base Iris padronizada



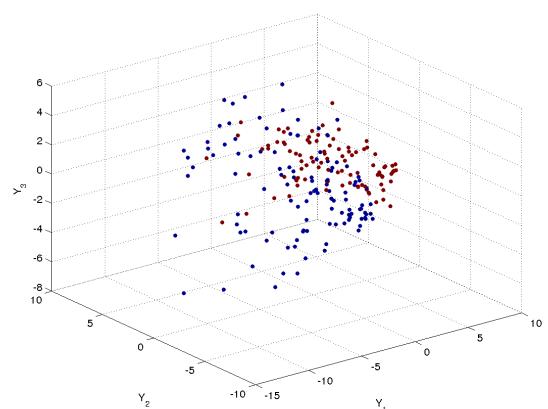
(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

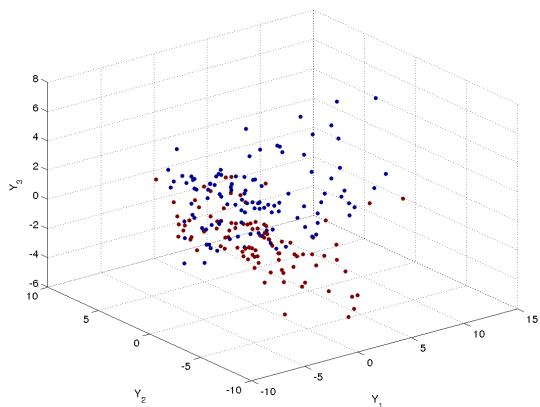
Figura C.36: Gráfico de dispersão para as representações tridimensionais a partir da base Sonar padronizada



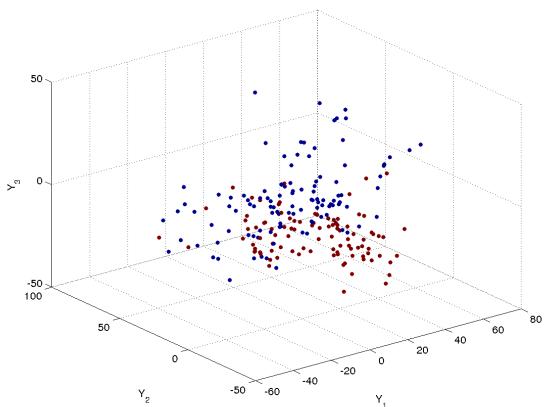
(a) PCA



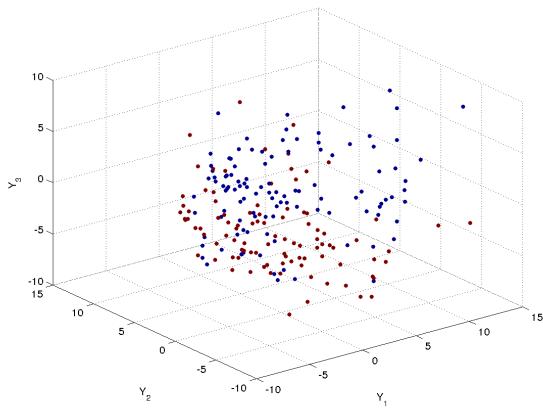
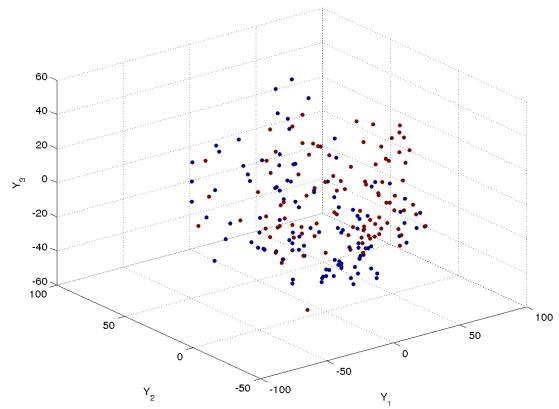
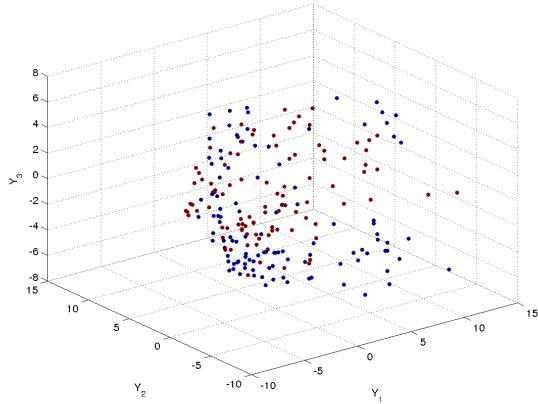
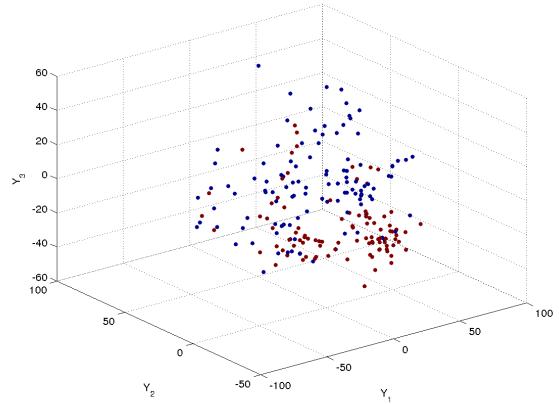
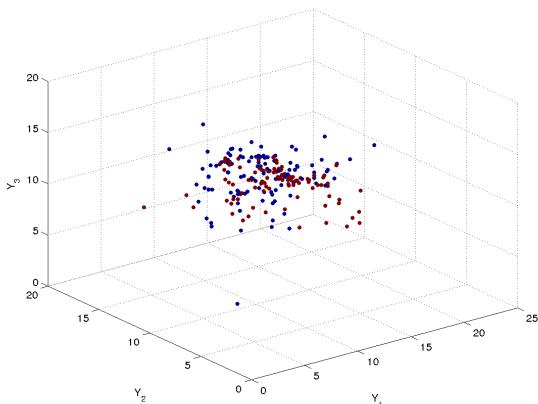
(b) SVD



(c) CMDS com $D_{Eucladiana}$

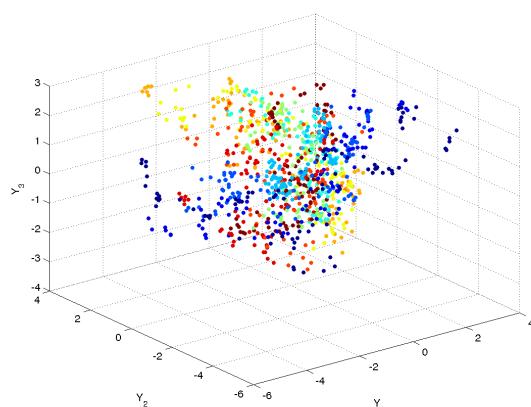


(d) CMDS com $D_{Manhattan}$

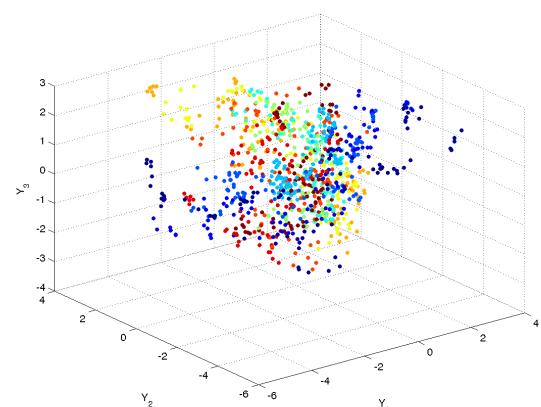
(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

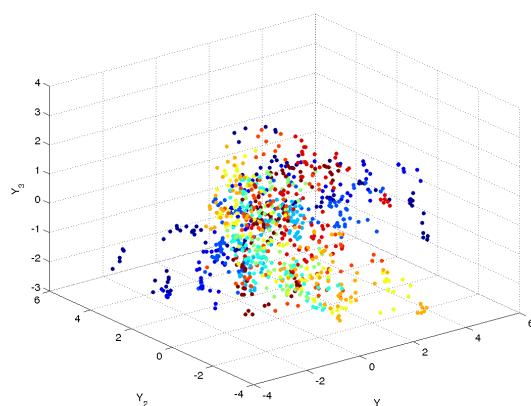
Figura C.37: Gráfico de dispersão para as representações tridimensionais a partir da base Vogal padronizada



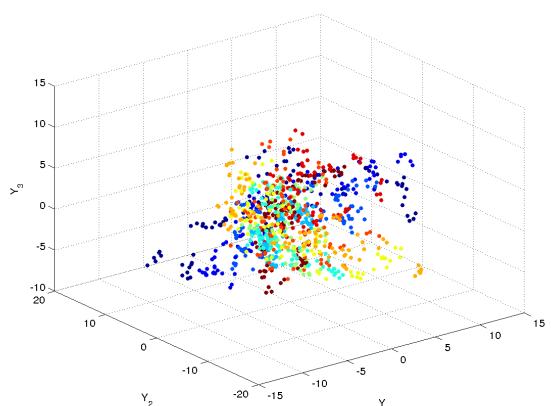
(a) PCA



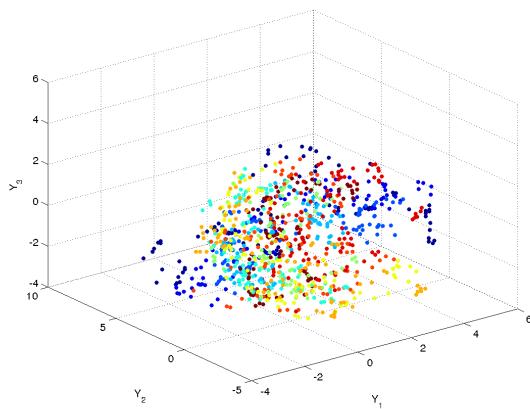
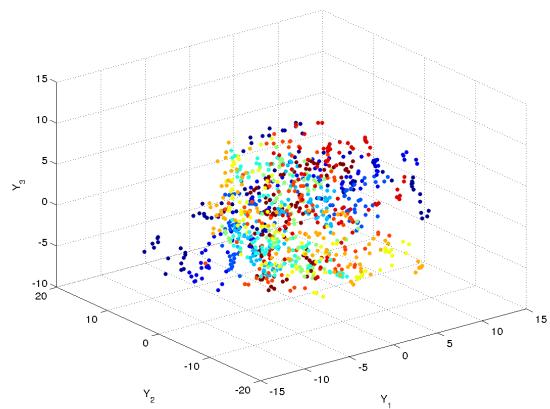
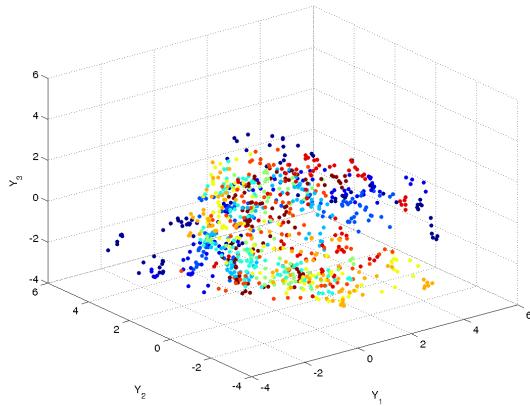
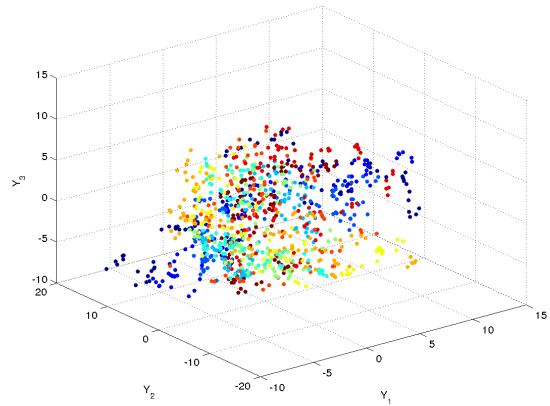
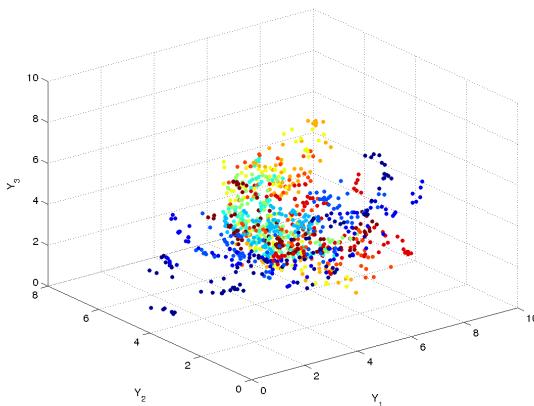
(b) SVD



(c) CMDS com $D_{Eucladiana}$



(d) CMDS com $D_{Manhattan}$

(e) MDS Stress-1 com $D_{Eucladiana}$ (f) MDS Stress-1 com $D_{Manhattan}$ (g) MDS Sstress com $D_{Eucladiana}$ (h) MDS Sstress com $D_{Manhattan}$ 

(i) FastMap

Referências Bibliográficas

- BORG, I.; GROENEN, P.J.F. Modern Multidimensional Scaling: Theory and Applications. Nova York, Nova York, EUA: Springer, 2005.
- BUJA, A. et al. Data Visualization with Multidimensional Scaling. *Journal of Computational and Graphical Statistics*, [S.l.: s.n.], 18 set. 2008.
- CADIMA, J. F. C. L. Redução de dimensionalidade através duma análise em componentes principais: um critério para o número de componentes a reter. [S.l.], [20 __]. Relatório Técnico - Departamento de Matemática, Instituto Superior de Agronomia.
- CANO, P. et al. On the Use of FastMap for Audio Retrieval and Browsing. In: Proceedings of 3rd Conference on Music Information Retrieval, Paris, França, 2002.
- COX, T. F.; COX, M.A.A. Multidimensional Scaling. 2. ed. Boca Raton, Florida, EUA: Chapman & Hall/CRC, 2001.
- CRUZ, C. J. D. W. Explorando conjuntos de dados volumétricos multidimensionais variantes no tempo usando projeções. São Carlos, SP, 2012. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) – ICMC, Universidade de São Paulo.
- CUNNINGHAM, P. Dimension Reduction. Dublin, Rep. Irlanda, 2007. Relatório Técnico - School of Computer Science and Informatics, University College Dublin.
- DING, C.; et al. Adaptive dimension reduction for clustering high dimensional data. In: Proceedings. 2002 IEEE International Conference on Data Mining, 2002. ICDM 2003., [S.l.]: IEEE, 147-154, 2002.
- ECKART, C.; YOUNG, G. Approximation of one matrix by another of lower rank. *Psychometrika*, 1, 211-218, 1963.
- FALOUTSOS C.; LIN K. FastMap: A Fast Algorithm for Indexing Data Mining and Visualization of Traditional and Multimedia Datasets. In: Proceedings of the Intl. Conf. on Management of Data (SIGMOD'95), p. 163-174, San Jose, California, EUA, 1995.
- FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMITH, P. From Data Mining to Knowledge Discovery in Databases. [S.l.] Al Magazine, Vol. 17, n. 3, p. 37-54, 1996a.
- FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMITH, P. Knowledge Discovery and Data Mining: Towards a Unifying Framework. In: Proc. of the KDD'96, 2nd Int'l Conf. on Knowledge Discovery and Data Mining, Portland, Oregon, EUA, p. 82-88, 1996b.
- FODOR, I. K. A survey of dimension reduction techniques. Technical report, Lawrence Livermore National Lab., California, EUA, 2002.

- GERING, D. Linear and Nonlinear Data Dimensionality Reduction. Cambridge, Massachusetts, EUA, 2002. Exame de Área (Doutorado em Inteligência Artificial) - Massachusetts Institute of Technology.
- GOWER, J. C. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53, 325-338, 1966.
- HAN, J.; KAMBER, M.; PEI, J. Data Mining – Concepts and Techniques. 3. ed. San Francisco, California, EUA: Morgan Kaufmann Publishers Inc., 2006.
- JOHNSON, R. A.; WICHERN, D. W. Applied Multivariate Statistical Analysis. 6. ed. Upper Saddle River, Nova Jersey, EUA: Prentice Hall Inc., 2007.
- KRUSKAL, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, [S.l.]: [s.n.], 27, 1-27, 1964a.
- KRUSKAL, J. B. Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, [S.l.]: [s.n.], 27, 1-27, 1964b.
- LÄUCHLI, P. Jordan-elimination und ausgleichung nach kleinsten quadraten. *Numerische Mathematik*, v.3, n.1, p. 226-240, 1961.
- MAATEN, L. J. P. van der. An Introduction to Dimensionality Reduction Using MatLab. Maastricht, Países Baixos, 2009. Relatório Técnico - Faculty of Humanities & Sciences, MICC/IKAT, Maastricht University.
- MAATEN, L. J. P. van der; POSTMA, E. O.; HERIK, H. J. van den. Dimensionality Reduction: A Comparative Review. Tilburg, Países Baixos, 2009. Relatório Técnico - Tilburg centre for Creative Computing, Tilburg University.
- MATHWORKS Inc., The. MatLab R2012a. Natick, Massachusetts, EUA, 2012.
- MOTTA, C. G. L. da. Metodologia para Mineração de Regras de Associação Multiníveis Incluindo Pré e Pós-Processamento. Rio de Janeiro, RJ, 2010. Tese (Doutorado em Engenharia Civil) - COPPE, Universidade Federal do Rio de Janeiro.
- MOTTA, C. G. L. da. Sistema Inteligente para Avaliação de Riscos em Vias de Transporte Terrestre. Rio de Janeiro, RJ, 2004. Dissertação (Mestrado em Engenharia Civil) - COPPE, Universidade Federal do Rio de Janeiro.
- MOTTA, C. G. L. da. Mineração de Dados: Capítulo 1 – Introdução. Material de Aula, Disciplina Mineração de Dados (DCC 127). Juiz de Fora, MG: Departamento de Ciência da Computação, Universidade Federal de Juiz de Fora, 2012a.
- MOTTA, C. G. L. da. Mineração de Dados: Capítulo 2 – Preparação dos Dados, Parte 1. Material de Aula, Disciplina Mineração de Dados (DCC 127). Juiz de Fora, MG: Departamento de Ciência da Computação, Universidade Federal de Juiz de Fora, 2012b.
- NAVAB, N. 3D Computer Vision Script Draft - Appendix. Material de Aula, 3D Computer Vision. Munique, Baviera, Alemanha: Institut für Informatik, Technische Universität München, p. 48-51, 2014.
- NEWMAN, C. L.; BLAKE, D. J.; MERZ, C. J. UCI Repository of machine learning databases. Irvine, California, EUA, 1998. Relatório Técnico - Dept. of Information and Computer Sciences, University of California.

- NG, A. Dimensionality Reduction. Material de Aula, Machine Learning. Standford, California, EUA: Standford University e Coursera Inc., 2013.
- PAULOVICH, F. V. et al. Least square projection: A fast high-precision multidimensional projection technique and its application to document mapping. *IEEE Transactions on Visualization and Computer Graphics* 14, p. 564-575, 2008.
- REZENDE, S. O. et al. Mineração de Dados, In: REZENDE, S. O. (Org.). Sistemas Inteligentes: Fundamentos e Aplicações. Barueri, SP: Editora Manole Ltda., 2003. Cap. 12, p. 307-336.
- ROWEIS, S. T.; SAUL, L. K. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science* 290, 5500, 2323–2326, 2010. SBC. Grandes Desafios da Pesquisa em Computação no Brasil – 2006 – 2016 – Relatório sobre o Seminário realizado em 8 e 9 de maio de 2006. São Paulo, SP: Sociedade Brasileira de Computação, 2006.
- SHEPARD, R. N. The analysis of proximities: multidimensional scaling with an unknown distance function. *Psychometrika*, [S.l.]: [s.n.], 27, 125-139, 1962.
- SHLENS, J. A tutorial on principal component analysis. [S.l.], 2003. Relatório Técnico - [s.n.]
- TAN, P.-N.; Steinbach, M. ; Kumar, V. Introduction to Data Mining. Boston, MA, EUA: Addison-Wesley Longman Publishing Co., Inc., 2005.
- TAKANE, Y.; YOUNG, F. W.; LEEUW, J. D. Non-metric Individual Differences Multidimensional Scaling: Alternating Least Squares with Optimal Scaling Features. *Psychometrika*, 42, 7-67, 1977.
- TORGESON, W. S. Multidimensional scaling: Theory and method. *Psychometrika*, [S.l.]: [s.n.], 17, 401-419, 1952.
- TORGESON, W. S. Multidimensional scaling of similarity. Theory and methods of scaling. Nova York, Nova York, EUA: Wiley, 1958.
- TZENG, J.; LU, H.; LI, W. Multidimensional scaling for large genomic data sets. *BMC Bioinformatics*, [S.l.]: BioMed Central, 9, 1, 1-17, 4 abr. 2008.
- VARGAS, A.. Metáforas visuais alternativas para layouts gerados por projeções multidimensionais. São Carlos, SP, 2012. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) – ICMC, Universidade de São Paulo.
- YOUNG, M. P.; HOUSEHOLDER, A. S. Discussion of a set of points in terms of their mutual distances. *Psychometrika*, [S.l.]: [s.n.], 3, 19-22, 1938.
- ZHANG, Z.; ZHA, H.. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *Journal of Shanghai University (English Edition)*, [Xangai, China]: Shanghai University e Springer, 8, 4, 406-424, 24 dez. 2004.