

Rosilda Benício de Souza

Uma metodologia inspirada quanticamente para  
aprimoramento do poder de discriminação de dados

RECIFE-PE - FEV/2015



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO  
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

## Uma metodologia inspirada quanticamente para aprimoramento do poder de discriminação de dados

Tese apresentada ao Programa de Pós-Graduação em Biometria e Estatística Aplicada como exigência parcial à obtenção do título de Doutor.

**Área de Concentração: Modelagem Estatística e Computacional**

Orientador: Prof. Dr. Tiago Alessandro Espínola Ferreira

RECIFE-PE - FEV/2015.

Ficha catalográfica

S729u Souza, Rosilda Benício de  
Uma metodologia inspirada quanticamente para  
aprimoramento do poder de discriminação de dados /  
Rosilda Benício de Souza. – Recife, 2015.  
87 f. : il.

Orientador: Tiago Alessandro Espínola Ferreira.  
Tese (Doutorado em Biometria e Estatística Aplicada) –  
Universidade Federal Rural de Pernambuco, Departamento  
de Estatística e Informática, Recife, 2015.  
Referências.

1. Classificação 2. Teorema de Cover 3. Computação  
quântica. 4. Poder de discriminação. 5. Sistemas  
inteligentes I. Ferreira, Tiago Alessandro Espínola,  
orientador II. Título

CDD 310

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO  
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

UMA METODOLOGIA INSPIRADA QUANTICAMENTE PARA  
APRIMORAMENTO DO PODER DE DISCRIMINAÇÃO DE DADOS

Rosilda Benício de Souza

Tese julgada adequada para obtenção do título de Doutor em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 23/02/2015 pela Comissão Examinadora.

Orientador:



Prof. Dr. Tiago Alessandro Espínola Ferreira  
Universidade Federal Rural de Pernambuco

Banca Examinadora:



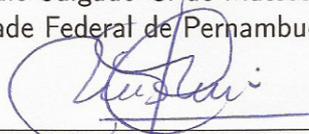
Prof. Dr. Wilson Rosa de Oliveira Junior  
Universidade Federal Rural de Pernambuco



Prof. Dr. Francisco Madeiro B. Junior  
Universidade Católica de Pernambuco



Prof. Dr. Paulo Salgado G. de Mattos Neto  
Universidade Federal de Pernambuco



Profa. Dra. Marina Paola Cipolletti  
Universidad Nacional del Sur, Bahía Blanca  
Argentina

Dedico este trabalho à minha família, em especial a minha mãe, meu exemplo, meu refúgio, minha fortaleza.

## Agradecimentos

Primeiramente a Deus, pelo dom da vida, por sua presença constante, pela saúde e determinação para alcançar meus objetivos e pelas pessoas boas que me rodeiam.

Da Universidade Federal Rural de Pernambuco, agradeço primeiramente ao Professor Tiago Alessandro Espínola Ferreira, pela orientação, conhecimentos transmitidos, incentivos, compreensão e apoio em momentos difíceis. Aos professores do PPGBEA, em especial a Profa. Tatijana Stosic e ao Prof. Wilson Rosa. E a Marco Santos, agradeço pela atenção que direta ou indiretamente tornaram esta fase mais simples de ser vivida.

A Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE), pelo apoio financeiro.

Aos Professores membros da banca examinadora: Wilson Rosa de Oliveira Junior, Francisco Madeiro B. Junior, Paulo Salgado G. de Mattos Neto e Marina Paola Cipolletti, pelas contribuições valiosas para melhoria desse trabalho.

Aos amigos e colegas do PPGBEA e PPGIA, pelas trocas de conhecimentos e boa convivência: em especial a Samuel, Cícero Carlos, Pedro Ferreira, Rivelino, Djalma, Diego, Joseilme, Lázaro, Dâmocles, Gabriel Rivas, Leila, Thaize, Hérica, Lêda Valéria, Evelyn, Gabriela, Ricardo Oliveira e Fillipo Regis.

Aos colegas do DECAT-UFS pela compreensão, contribuições e atenção em cada solicitação realizada por mim.

A Emeson Santana pela contribuição essencial para realização desse trabalho.

Aos amigos que estiveram mais presentes e fizeram meus dias melhores nessa cidade: Rodrigo Silva, Patrícia Alcântara e Amanda Menezes, pela amizade e companheirismo em todas as horas. Cícero Pinheiro, Elias Neto e Thaís Cerdeira, pela boa e feliz convivência.

Agradeço a minha família pelo amor, atenção, confiança, incentivos e ensinamentos de vida. Meus pais Adocil e Maria, minhas irmãs Ivanilda e Ana, meus irmãos Rivaldo e Arnaldo, cunhados Cicero, Raimundo Cândido e Ciene, meus doces sobrinhos Naiani, Rauany, Kelvin e Brian.

Ao meu esposo, Francisco Damião, pelo companheirismo, compreensão e apoio na fase

de conclusão do curso. Pelos momentos felizes e turbulentos que dividimos. E pela família especial em que me fez parte dela.

Aos Professores Evandro Carlos Ferreira e Pedro Ferreira de Lima que foram responsáveis por minha continuidade na vida acadêmica.

A todos que de alguma forma contribuíram para a realização deste trabalho.

## Resumo

Neste trabalho é proposta uma metodologia para a resolução do problema de classificação. Inspirada na Mecânica Quântica e utilizando técnicas de Inteligência Artificial, o método consiste em aprimorar o poder de discriminação de dados, baseado no teorema de Cover. Seja  $\mathcal{P}$  um problema de classificação não linearmente separável, o processo proposto consiste em incrementar a dimensão do conjunto de dados a fim de tornar o problema  $\mathcal{P}$  em um problema de classificação linearmente separável. Neste processo também é suposto que o conjunto de dados original é um observável no mundo quântico, *i.e.*, o conjunto de dados original (números reais) é criado a partir de uma medida do valor esperado de um estado de um sistema quântico (números complexos). Um Algoritmo Genético (AG) é aplicado para buscar uma transformação inversa ( $T$ ) da medida do valor esperado, voltando do mundo clássico para o mundo quântico, *i.e.*, tornando números reais em números complexos, sujeito à restrição de conservação de magnitude. E para constatar quanto à discriminação dos dados, métodos de classificação como *k-Nearest Neighbor* (kNN), *K-means* e Análise Discriminante Linear (LDA) são aplicados ao problema de classificação em duas condições: conjunto de dados originais e conjunto de dados transformados com a metodologia proposta. Verificada a separabilidade dos dados no mundo quântico, a técnica de Análise de Componentes Principais (PCA) é aplicada para selecionar um conjunto mínimo de características. Uma Rede Neural Artificial (RNA) é aplicada para aprender a transformação  $T$  e conduzir quaisquer novas observações do mundo clássico para o mundo quântico (conjunto gerado a partir do PCA). Experimentos realizados comprovam a robustez do método proposto. Tendo sido testado em dez conjuntos de dados com diferentes características como: natureza, cardinalidade, número de classes, número de instâncias por classe e dimensionalidade. Dois foram gerados artificialmente (Círculos concêntricos e Espirais concêntricos), sete são *benchmark* (Diabetes, Câncer de mama, Iris, Coluna vertebral, Sementes, Vinho e *Yeast*) e um conjunto de dados ainda não disponível na literatura, o conjunto de dados de Mata Atlântica pernambucana. Por fim, a comparação dos resultados de classificação é apresentada, indicando um aprimoramento no poder de discriminação dos dados quando esta metodologia é aplicada. Com uma taxa de reconhecimento de aproximadamente cem por cento para todos os conjuntos de dados, tem resultados superior às demais técnicas investigadas em vários trabalhos encontrados na literatura.

**Palavras-chave:** Classificação. Teorema de Cover. Computação Quântica. Poder de discriminação. Sistemas Inteligentes.

## Abstract

In this work is proposed a methodology for the solution of the Classification problems. Inspired in the quantum mechanics and using artificial Intelligence techniques, the method consists in enhancement the power of data discrimination based on the Cover theorem. Let  $\mathcal{P}$  a non linearly separable classification problem, the proposed process consists in increasing the data dimension in order to convert the problem  $\mathcal{P}$  into a linearly separable classification problem. In this process also is supposed that the original data are observables in the quantum world, *i.e.*, the original data (real number) are created from an expected value measurements of a quantum system state (complex numbers). A Genetic Algorithm (GA) is applied for search an inverse mapping ( $T$ ) of the expected value measurements, going back of the classical world into the quantum world, *i.e.*, transforming the real number into complex number, subject to the constraintment of magnitude conservation. To check of the data discrimination, methods of classification like k-Nearest Neighbor (kNN), K-means and Linear Discriminant Analysis (LDA) are applied to benchmark classification problems in two conditions: raw data set and transformed data set with the proposed methodology. After verified the separability of data in the quantum world, the Principal Component Analysis technique is applied to select a minimal set of features. An Artificial Neural Network is applied to learn the transformation ( $T$ ) and to lead any new observations of the classical world to the quantum world (generated by PCA). Experiments conducted prove the robustness of the proposed method. Were tested ten data set with features different, like: size data set, number of classes, number of elements per class and dimensionality. Two data set were generated artificially: Concentric circles and Concentric spirals; seven are benchmark: Diabetes (Pima Indians Diabetes), Breast Cancer, Iris, Vertebral Column, Seeds, Wine and Yeast; And a data set not yet presented in the literature, the data of Atlantic Forest of Pernambuco. Finally, the comparison of the classification results are presented, indicating a enhancement in the data power discrimination when the proposed methodology is applied. With a rate of recognition of nearly hundred percent for all data set, has higher results of that other techniques investigated in several studies in the literature.

**Key words:** Classification. Cover Theorem. Quantum Computing. Power Discriminant. Intelligent Systems.

# Lista de Figuras

1	Procedimento padrão de um Algoritmo Genético genérico. . . . .	p. 7
2	Rede MLP de três camadas totalmente conectadas. . . . .	p. 12
3	Gráficos das funções de ativação mais comuns: (a) Função Sigmóide Logística, (b) Função Tangente Hiperbólica, (c) Função Seno e (d) Função Linear. . . . .	p. 13
4	Classificação feita pelo kNN para um padrão desconhecido $x_d$ . . . . .	p. 15
5	Funcionamento do método <i>k-means</i> para três classes ( $k = 3$ ). . . . .	p. 17
6	A Espera de Bloch. $ \Psi\rangle = \alpha 0\rangle + \beta 1\rangle$ . . . . .	p. 21
7	Ilustração da metodologia proposta. . . . .	p. 28
8	Projeções no $\mathbb{R}^3$ do conjunto de dados de Mata Atlântica pernambucana, obtidas a partir da combinação dos quatro atributos que o descrevem (d1 = clorofila foliar, d2 = área foliar específica, d3 = teor de matéria seca da folha, d4 = teor de matéria úmida). . . . .	p. 32
9	Representação no espaço real, (a) 600 pontos distribuídos em dois Círculos concêntricos e (b) 600 pontos distribuídos em dois Espirais concêntricos. . . . .	p. 33
10	Ilustração do particionamento dos conjuntos de dados e aplicação da metodologia proposta. . . . .	p. 35
11	Assertividade do método kNN para valores de $k$ ( $1 \leq k \leq \sqrt{N}$ ) para os dados originais de: (a) Círculos concêntricos, (b) Espirais concêntricos, (c) Diabetes, (d) Câncer de mama, (e) Iris, (f) Mata Atlântica, (g) Coluna vertebral, (h) Sementes, (i) Vinho e (j) <i>Yeast</i> . . . . .	p. 37
12	Projeção sobre o $\mathbb{R}^3$ do espaço complexo obtido com a aplicação da metodologia, (a) 600 pontos distribuídos em dois Círculos concêntricos e (b) 600 pontos distribuídos em dois Espirais concêntricos. . . . .	p. 38

13	Taxa de reconhecimento para os métodos kNN, <i>K-means</i> e LDA para os dados brutos e dados transformados (AG) de: (a) Círculos concêntricos, (b) Espirais concêntricos, (c) Diabetes, (d) Câncer de mama, (e) Iris, (f) Mata Atlântica, (g) Coluna vertebral, (h) Sementes, (i) Vinho e (j) <i>Yeast</i> . . . . .	p. 51
14	Projeção do espaço complexo do conjunto de dados de Mata Atlântica pernambucana no $\mathbb{R}^3$ , espaço obtido com a aplicação da metodologia. . . . .	p. 52
15	Projeções no $\mathbb{R}^3$ para os dados originais, transformados pelo AG e transformados pelo método PCA, respectivamente. Para os conjuntos de dados: (a) Diabetes, (b) Câncer de mama, (c) Iris, (d) Coluna vertebral, (e) Sementes, (f) Vinho e (g) <i>Yeast</i> . . . . .	p. 54
16	Os Círculos concêntricos, (a) 600 pontos distribuídos em dois Círculos concêntricos no espaço real e (b) uma projeção no $\mathbb{R}^3$ do espaço obtido com a aplicação da técnica PCA. . . . .	p. 55
17	Os Espirais concêntricos, (a) 600 pontos distribuídos em dois Espirais concêntricos no espaço real e (b) uma projeção no $\mathbb{R}^3$ do espaço obtido com a aplicação da técnica PCA. . . . .	p. 56
18	Taxa de reconhecimento para os métodos kNN, <i>K-means</i> e LDA para os dados brutos e dados mapeados (RNA) de: (a) Círculos concêntricos, (b) Espirais concêntricos, (c) Diabetes, (d) Câncer de mama, (e) Iris, (f) Mata Atlântica, (g) Coluna vertebral, (h) Sementes, (i) Vinho e (j) <i>Yeast</i> . . . . .	p. 61
19	Taxa de reconhecimento para o método LDA aplicado aos dados brutos e dados transformados (AG) da classe Conservativa dos dados de Mata Atlântica pernambucana considerando diferentes números de espécies. . . . .	p. 62
20	Projeções do espaço complexo no $\mathbb{R}^3$ para o conjunto de dados da classe Conservativa com diferentes números de espécies. . . . .	p. 63

# Lista de Tabelas

1	Descrição sintética do conjunto de dados de espécies da Mata Atlântica pernambucana. . . . .	p. 31
2	Descrição sintética dos conjuntos de dados usados. . . . .	p. 33
3	Taxa de reconhecimento para o método kNN para os dados brutos e dados transformados (AG) de Círculos concêntricos, Espirais concêntricos, Diabetes, Câncer de mama, Iris, Mata Atlântica, Coluna vertebral, Sementes, Vinho e <i>Yeast</i> . . . . .	p. 40
4	Taxa de reconhecimento para os métodos <i>K-means</i> e LDA para os dados brutos e dados transformados (AG) de Círculos concêntricos, Espirais concêntricos, Diabetes, Câncer de mama, Iris, Mata Atlântica, Coluna vertebral, Sementes, Vinho e <i>Yeast</i> . . . . .	p. 49
5	Energia acumulada para dados $d$ -dimensional e taxa de reconhecimento para os métodos kNN, <i>K-means</i> e LDA aplicados aos dados transformados pelo PCA. Dados de Círculos concêntricos, Espirais concêntricos, Diabetes, Câncer de mama, Iris, Mata Atlântica, Coluna vertebral, Sementes, Vinho e <i>Yeast</i> . . . . .	p. 52
6	Configurações de modelos de redes neurais MLP para Círculos concêntricos, Espirais concêntricos, Diabetes, Câncer de mama, Iris, Mata Atlântica, Coluna vertebral, Sementes, Vinho e <i>Yeast</i> . . . . .	p. 55
7	Taxa de reconhecimento para o método kNN para os dados originais preservados e dados mapeados pela RNA de Círculos concêntricos, Espirais concêntricos, Diabetes, Câncer de mama, Iris, Mata Atlântica, Coluna vertebral, Sementes, Vinho e <i>Yeast</i> . . . . .	p. 58
8	Taxa de reconhecimento para os métodos <i>K-means</i> e LDA para os dados originais preservados e dados mapeados pela RNA de Círculos concêntricos, Espirais concêntricos, Diabetes, Câncer de mama, Iris, Mata Atlântica, Coluna vertebral, Sementes, Vinho e <i>Yeast</i> . . . . .	p. 59

9 Taxa de reconhecimento para o método LDA aplicado aos dados brutos e dados transformados (AG) da classe Conservativa dos dados de Mata Atlântica pernambucana considerando diferentes números de espécies. . . . . p. 62

# Sumário

<b>1</b>	<b>Introdução</b>	p. 1
<b>2</b>	<b>Métodos Clássicos Abordados e Definições</b>	p. 5
2.1	Algoritmos Genéticos . . . . .	p. 5
2.1.1	Indivíduo e População . . . . .	p. 5
2.1.2	Função de custo . . . . .	p. 6
2.1.3	Seleção de Indivíduos . . . . .	p. 6
2.1.4	Operadores Genéticos: Cruzamento e Mutação . . . . .	p. 7
2.2	Análise de Componentes Principais - PCA . . . . .	p. 8
2.3	Redes Neurais Artificiais - RNAs . . . . .	p. 11
2.3.1	Rede <i>Multilayer Perceptron</i> - MLP . . . . .	p. 11
2.4	Métodos de Classificação . . . . .	p. 14
2.4.1	<i>k-Nearest Neighbor</i> - kNN . . . . .	p. 14
2.4.2	<i>K-means</i> . . . . .	p. 15
2.4.3	Análise Discriminante Linear - LDA . . . . .	p. 17
<b>3</b>	<b>Inspiração Quântica</b>	p. 20
3.1	Estado Quântico . . . . .	p. 20
3.2	Observável Quântico . . . . .	p. 21
<b>4</b>	<b>Metodologia Proposta</b>	p. 23
4.1	Metodologia . . . . .	p. 23
4.2	A transformada inversa - Algoritmo Genético . . . . .	p. 25

4.3	A transformada - PCA . . . . .	p. 26
4.4	Mapeamento de novos pontos . . . . .	p. 27
4.5	Validação da metodologia . . . . .	p. 29
4.6	Dados . . . . .	p. 30
4.6.1	Dados das espécies de Mata Atlântica pernambucana . . . . .	p. 30
4.6.2	Conjuntos de dados usuais na literatura . . . . .	p. 31
<b>5</b>	<b>Resultados Experimentais</b>	<b>p. 34</b>
5.1	Experimentos realizados . . . . .	p. 34
5.2	Taxa de reconhecimento da transformada inversa - AG . . . . .	p. 35
5.3	Taxa de reconhecimento da transformada - PCA . . . . .	p. 48
5.4	Mapeando um novo ponto no espaço de representação complexo . . . . .	p. 49
5.5	Estudo adicional - Comparando performance por número de classes . . . . .	p. 57
<b>6</b>	<b>Conclusões</b>	<b>p. 64</b>
6.1	Conclusões . . . . .	p. 64
6.2	Trabalhos Futuros . . . . .	p. 66
	<b>Referências</b>	<b>p. 68</b>

# 1 Introdução

Denomina-se clusterização ao processo de agrupamento de objetos físicos ou abstratos em classes de objetos similares. O grau de dificuldade para classificar um conjunto de dados em algum número de classes depende da variabilidade nos valores característicos para certos objetos na mesma categoria, relativa à diferença entre os valores característicos para objetos em diferentes categorias. Basicamente, dois fatores podem ter efeito sobre a variabilidade dos valores característicos para um certo objeto: a complexidade dos dados e o ruído [1–4].

Se o conjunto de dados apresenta um baixo nível de ruído e complexidade com relações linearmente independentes para as características, então a tarefa de classificação é um problema linearmente separável e existe um hiperplano ótimo que conduz à uma classificação com taxa de reconhecimento de 100%. Por outro lado, se o problema de classificação é um problema não linearmente separável, essa taxa de reconhecimento é frequentemente impossível de ser alcançada. Em geral busca-se determinar a probabilidade de um dado objeto pertencer a cada uma das possíveis classes. Portanto, o problema linearmente separável é relativamente mais simples para resolver do que um problema não linearmente separável [5, 6].

Entretanto, se o conjunto de dados é não linearmente separável e  $d$ -dimensional (cada objeto é um vetor de padrões  $d$ -dimensional), o Teorema de Cover [7] postula que expandindo a dimensionalidade do espaço de representação (espaço onde o conjunto de dados é representado) a probabilidade assintótica de classificações ambíguas irá decrescer. Assim, se esta dimensionalidade cresce o suficiente, o problema não linearmente separável, com alta probabilidade, pode tornar-se um problema linearmente separável.

Em geral, as técnicas de agrupamento não tem um tutor ou professor, caracterizando um processo de aprendizagem não supervisionado. De modo que os dados são agrupados em *clusters* de acordo com alguma medida de similaridade. Assim, um sistema de clusterização pode definir classes (ou *clusters*) para um conjunto de dados usando um vetor característico. A ideia principal é especificar um objeto de um conjunto de dados para uma destas classes definidas.

Métodos de aprendizagem não supervisionado como o *K-means* [8] e supervisionado como *k-Nearest Neighbor* - kNN [9] e a Análise Discriminante Linear (LDA) [1, 10] são usados frequentemente para a tarefa de classificação. Algumas variações destes métodos tem sido propostas. Prabhu e Anbazhagan [11] propuseram um método para encontrar os centróides iniciais para tornar o *K-means* mais eficaz e eficiente. Zhang e Fang [12] analisaram as vantagens e desvantagens do *K-means* tradicional e elaboraram um método de fornecer um *K-means* baseado no ponto inicial, além de determinar o valor de *k*. Huang *et al* [13] propuseram um subespaço discriminativo *Kmeans* (*DSKmeans*), que integra a compacticidade intra classe e a separação inter classe simultaneamente.

Introduzido por Gao e Gao [14] um AdaBoost [15] editado pelo kNN ponderado (EAdaBoost) é projetado, onde AdaBoost e kNN naturalmente se complementam. Resultando em melhorar a precisão da classificação e evitar superajuste. Huang *et al.* [16] propuseram um método que utiliza um algoritmo kNN melhorado para resolver os conflitos de regras, denominado Classificação Associativa com kNN (AC-kNN). AC-kNN gera para cada instância de teste de um conjunto de treinamento específico composto de instâncias cobertas pelas melhores *n* regras que correspondem à instância de teste. Assim, os vizinhos mais próximos do conjunto de treinamento específico não são apenas similares, são também associativos com a instância de teste. Em [17] Panahi *et al.* propuseram um método híbrido formado pelos métodos Análise de componentes principais (PCA) e LDA. O objetivo é encontrar uma classificação bem sucedida, rápida e de baixo custo computacional.

Metodologias de clusterização/classificação frequentemente são apresentadas a partir da Inteligência Artificial (IA), um ramo da Ciência da Computação cujo objetivo é desenvolver sistemas que executam funções desempenhadas pelo ser humano usando conhecimento e raciocínio. Estas metodologias são de grande importância e um ramo usual em algumas áreas como mineração de dados, estatística, engenharia, ciências da computação entre outras ciências que trabalham com análises de dados [1, 18, 19].

A IA é uma área da ciência que busca, através de técnicas inspiradas na Natureza, o desenvolvimento de sistemas inteligentes que imitam aspectos do comportamento humano, como o aprendizado. Por exemplo, as Redes Neurais Artificiais (RNAs) [20–22] são modelos computacionais não lineares, inspirados na estrutura e operação do cérebro humano; Os Algoritmos Genéticos (AGs) [23] são algoritmos matemáticos inspirados nos mecanismos de evolução natural e recombinação genética.

Um recente paradigma da computação tem atraído atenção crescente não só da área científica, mas também da área tecnológica/industrial. A Computação Quântica (CQ), surgida

no início da década de 80, é um domínio de pesquisa que utiliza elementos de três áreas de conhecimento: Matemática, Física e Computação. Pesquisadores tem sido atraídos pelas vantagens teóricas advindas da utilização de fenômenos quânticos para a representação e o processamento de informação e sua potencialidade no uso do paralelismo quântico como ferramenta para resolver problemas matemáticos mais eficientemente.

Um sistema quântico é governado pela mecânica quântica [24]. A mecânica quântica é um ramo da Física que lida com fenômenos em uma escala microscópica, onde a magnitude das variáveis são da ordem da Constante de Planck ( $h = 6.626069 \times 10^{-34} J \cdot s$ ). Neste sentido, a mecânica quântica é uma teoria mais fundamental do que a mecânica clássica, pois um sistema macroscópico (sistema clássico) pode ser entendido como a composição de sistemas microscópicos.

Em um sistema quântico [24, 25], um vetor de estado é um vetor complexo, comumente descrito por um *ket*  $|\Psi\rangle$ , o espaço de representação quântica é um espaço de Hilbert, onde este pode ser visto como uma expansão do espaço Euclidiano. Uma característica do sistema quântico pode ser medida aplicando um operador específico ao vetor de estado. Se o valor da característica é um número real, então esta característica é chamada de *observável* e pode ser observada no mundo clássico (mundo real) [26–29].

Nos últimos anos, vários esforços tem sido aplicados no intuito de utilizar este novo paradigma da computação na elaboração de estruturas da Inteligência Artificial, destacando-se as Redes Neurais Quânticas [30–35]. No entanto, a inspiração quântica no ramo da IA para problemas de classificação ainda é pouco explorada. Poucos resultados em torno do seu uso são encontrados na literatura, [36–40].

Em contraste aos métodos tradicionais de classificação, como Análise Discriminante Linear (LDA) [10] e suas variações [41, 42], as redes neurais como a Kohonen SOM [43, 44], ou outros métodos [45–47] onde a fim de controlar a taxa de reconhecimento diminui-se a dimensionalidade do espaço característico, a metodologia proposta neste trabalho realiza um pré processamento dos dados incrementando a dimensionalidade do espaço de representação, similarmente ao *Support Vector Machine* (SVM) [48] e as funções de *Kernel* [49–51], para dados não linearmente separáveis. É proposto um pré-processamento de dados baseado no Teorema de Cover e inspirado no sistema quântico. Considerando um conjunto de dados, é suposto que os pontos deste são medidas de alguma característica de um sistema quântico, em particular, um observável.

O objetivo principal desta Tese está contido na proposta de um mapeamento de um conjunto de dados para um espaço de representação de dimensionalidade superior, possivelmente,

tornando-o linearmente separável e possibilitando facilitar a classificação. Com a aplicação de técnicas inteligentes pretende-se construir esta transformação, que pode ser tratada como um pré-processamento de dados, uma vez que qualquer método de classificação pode ser aplicado posteriormente, como o *K-means* [8], *k-Nearest Neighbor* (kNN) [9] e Análise Discriminante Linear (LDA) [10].

Este trabalho é composto por seis capítulos, organizados da seguinte forma:

Capítulo 1 - Introdução: no capítulo inicial é realizada uma pequena introdução ao problema de classificação, bem como é sugerida a criação de um método que possibilita uma resolução viável deste problema.

Capítulo 2 - Métodos Clássicos Abordados e Definições: é feita uma breve revisão de literatura sobre os métodos clássicos abordados no trabalho. Os métodos inteligentes que virão a contribuir no desenvolvimento da metodologia proposta (AGs e RNAs), bem como métodos de classificação que tornam plausível a afirmação de que qualquer método de classificação poderá ser aplicado com um alto desempenho na classificação (kNN, *K-means* e LDA).

Capítulo 3 - Inspiração Quântica: é apresentada uma breve revisão de literatura sobre alguns conceitos quânticos abordados. Na metodologia proposta supõe-se que os pontos de um conjunto de dados a ser classificado são medidas de alguma característica de um sistema quântico, em particular, um observável.

Capítulo 4 - Metodologia Proposta: é apresentado o método proposto para o incremento da dimensão de um conjunto de dados, possibilitando a separabilidade de seus elementos em classes. Bem como as técnicas usadas para obtenção do incremento da dimensionalidade e realização do mapeamento de novos pontos no espaço de representação criado com a transformação sugerida. É apresentado também o processo usado para validar a metodologia e a descrição dos conjuntos de dados usados em sua aplicação.

Capítulo 5 - Resultados Experimentais: são apresentados os resultados da aplicação da metodologia proposta para discriminar dados. Uma vez que métodos de classificação são aplicados a conjuntos de dados usuais (originais) e posteriormente a conjuntos de dados transformados, os resultados são comparados para observância do desempenho nas duas condições: originais e transformados.

Capítulo 6 - Conclusões: são apresentadas as conclusões dos estudos realizados e dos resultados experimentais para a resolução de problemas de classificação. Como também as perspectivas de possíveis desdobramentos para trabalhos futuros que poderão contribuir na solução de difíceis problemas de classificação.

## 2 Métodos Clássicos Abordados e Definições

Uma breve revisão de literatura sobre os métodos clássicos abordados no trabalho é apresentada. Os métodos inteligentes que virão a contribuir no desenvolvimento da metodologia proposta (AGs e RNAs), bem como métodos de classificação que tornam plausível a afirmação de que qualquer método de classificação poderá ser aplicado fornecendo uma alta taxa de reconhecimento na classificação (kNN, *K-means* e LDA).

### 2.1 Algoritmos Genéticos

São modelos computacionais inspirados na teoria da evolução de Charles Darwin [52]. O Algoritmo Genético (AG) refere-se ao modelo introduzido e estudado por John Holland [23], de modo que a maior parte da teoria existente sobre os AGs aplica-se total ou primariamente ao modelo original. Utilizam conceitos provenientes do princípio de seleção natural para abordar uma série ampla de problemas, em especial de busca e otimização. Robustos, genéricos e facilmente adaptáveis, consistem de uma técnica amplamente estudada e utilizada em diversas áreas do conhecimento.

#### 2.1.1 Indivíduo e População

Na busca da melhor solução para o problema a ser resolvido, um AG parte de um certo conjunto de estados iniciais que é formado por possíveis soluções para o problema. Cada uma destas possíveis soluções é vista como sendo um *indivíduo* e o grupo de indivíduos, como uma *população*. Através de mudanças no conjunto de soluções, o algoritmo tenta encontrar as melhores, em analogia à evolução das espécies proposta por Darwin, indivíduos mais aptos.

O conceito de população aplica-se a um grupo de indivíduos - cada um representado por um cromossomo - que são prováveis soluções para o mesmo problema, sendo diferentes entre si. De modo geral, os cromossomos podem ser codificados utilizando uma cadeia de bits ou

diretamente por números inteiros. No que diz respeito à codificação do cromossomo, deve-se ter em mente que o mesmo deverá possuir material genético que poderá ser trocado entre os cromossomos na reprodução, os genes. Um gene nada mais é do que uma divisão do cromossomo, é nesse sentido que entra a cadeia de bits, na qual cada bit é um gene e o conjunto é o cromossomo.

A aplicação de um AG é a busca em um espaço constituído de hipóteses ou candidatas à solução. A melhor hipótese é determinada através de uma medida numérica pré-determinada, comumente chamada função de aptidão, custo ou *fitness*.

### 2.1.2 Função de custo

A função de custo de um problema de otimização é uma função matemática construída a partir dos parâmetros que o envolvam. Esta função gera um índice de ordenamento por aptidão para todos os indivíduos da população, representando o grau de adaptação do indivíduo ao problema, isto é, a qualidade de uma solução. Esta função é de extrema importância por determinar a capacidade de um indivíduo sobreviver e se reproduzir, sendo a avaliação de seus resultados que guia quanto à evolução para resultados melhores.

O cálculo da aptidão dá-se da seguinte forma: toma-se como entrada um indivíduo e retorna-se um número, ou uma lista de números, que representam o desempenho do indivíduo em relação a como o mesmo auxilia na resolução do problema, aquele que possuir o maior valor de *fitness* ou o menor custo, será considerado a melhor solução.

Existe um grande número de diferentes implementações para os AGs, mas de forma geral são variações do pseudo código ilustrado na Figura 1, um algoritmo representando um AG genérico.

### 2.1.3 Seleção de Indivíduos

A função da seleção em um AG é escolher os elementos da população que participarão do processo de reprodução, isto é, selecionar indivíduos, geralmente dois, denominados pais para gerar indivíduos na nova população. Esta escolha deve ser feita de tal forma que os membros da população mais adaptados ao meio ambiente, tenham maior chance de reprodução, ou seja, aqueles que apresentam um valor de *fitness* mais elevado. Existem vários métodos para selecionar os indivíduos sobre os quais serão aplicados os operadores genéticos: seleção baseada em meta, método da amostragem estocástica (método da roleta), seleção por torneio, dentre outras, porém o método de seleção por roleta é bastante usual [53, 54].

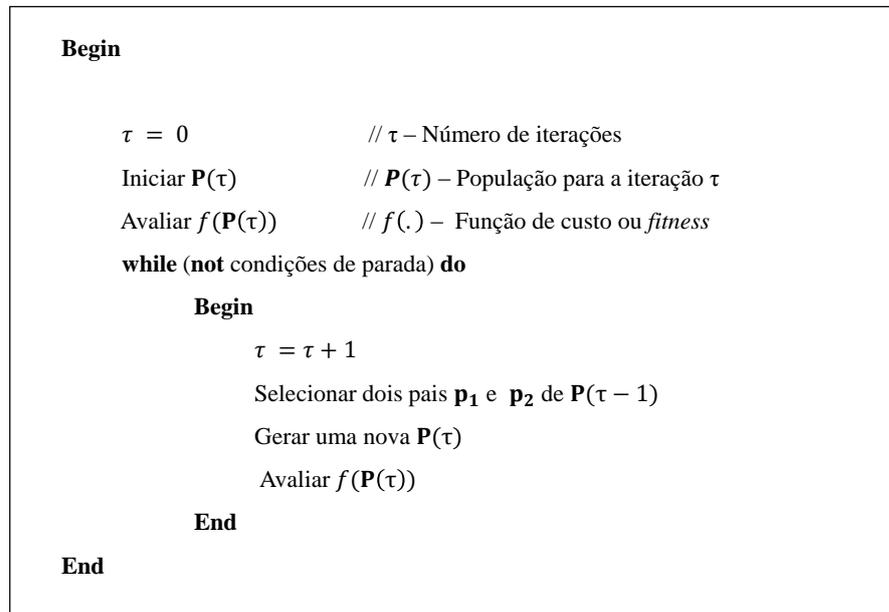


Figura 1: Procedimento padrão de um Algoritmo Genético genérico.

No método da amostragem estocástica (método da roleta) o indivíduo mais apto terá maior chance de sobreviver. Desta forma, para uma população com  $N$  indivíduos, a probabilidade de um indivíduo  $i$  ser selecionado é dado por

$$P(ind_i) = \frac{f(ind_i)}{\sum_{j=1}^N f(ind_j)}$$

em que  $ind_i$  representa o indivíduo  $i$  e  $f(\cdot)$  é a função de *fitness*.

#### 2.1.4 Operadores Genéticos: Cruzamento e Mutação

O princípio básico dos operadores genéticos é transformar a população através de sucessivas gerações, realizando a busca até chegar a um resultado satisfatório. Os operadores genéticos são necessários para que a população se diversifique e mantenha características de adaptação adquiridas pelas gerações anteriores. O algoritmo genético descrito por David Goldberg [55] utiliza os operadores de cruzamento e mutação. Para que a nova população seja gerada, uma seleção é feita mantendo, ou não, alguns dos membros da população atual e gerando novos através de cruzamento (*crossover*) e mutação.

O *crossover* é a operação mais representativa de um algoritmo genético. Ela tenta imitar a reprodução sexuada criando novos indivíduos a partir de dois indivíduos "pais" que tem

suas características misturadas. Trechos das características de um indivíduo são trocados pelo trecho equivalente do outro. O resultado desta operação é um indivíduo que potencialmente combine as melhores características dos indivíduos usados como base.

Enquanto o operador de mutação é responsável pela inserção de pequenas mudanças aleatórias nos cromossomos dos "filhos". Opera simplesmente modificando aleatoriamente alguma característica do indivíduo sobre o qual é aplicado. Esta modificação é importante por criar novos valores de características que não existiam ou apareciam em pequena quantidade na população em análise. O operador de mutação é necessário para a introdução e manutenção da diversidade genética da população. Desta forma, a mutação assegura que a probabilidade de se chegar a qualquer ponto do espaço de busca possivelmente não será zero.

## 2.2 Análise de Componentes Principais - PCA

Análise de componentes principais (do inglês *Principal Component Analysis* - PCA) é uma técnica de estatística para simplificar um conjunto de dados. O objetivo do método é reduzir a dimensionalidade dos dados multivariados preservando o máximo de informação possível [56, 57]. PCA é uma transformação linear que converte os dados para um novo sistema de coordenadas tal que o novo conjunto de variáveis - as componentes principais - são funções das variáveis originais, não correlacionadas.

Os passos para calcular as componentes principais são:

- Obter a matriz de dados com  $m$  amostras de vetores de dimensão  $n$ ;
- Calcular a matriz de covariância destes dados;
- Calcular os autovetores e autovalores da matriz de covariância.

Considere a situação em que observa-se  $n$  atributos de  $m$  exemplares de um conjunto de dados. Os atributos observados são representados pelas variáveis  $X_1, X_2, X_3, \dots, X_n$ . A matriz de dados é de ordem  $m \times n$  e normalmente denominada de matriz  $X$ .

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

A estrutura de dependência linear entre as variáveis da matriz de dados é representada pela matriz de covariância  $S$  ou pela matriz de correlação  $R$ . Na prática, o entendimento dessa estrutura através das variáveis  $X_1, X_2, X_3, \dots, X_n$ , pode ser uma tarefa complicada. Dessa forma, a PCA tem como objetivo transformar essa estrutura, representada pelas variáveis  $X_1, X_2, X_3, \dots, X_n$ , em uma outra estrutura representada pelas variáveis  $Y_1, Y_2, Y_3, \dots, Y_n$ , em que nesta última as variáveis são não correlacionadas e com variâncias ordenadas, para que seja possível comparar os exemplares usando apenas as variáveis  $Y_i$ 's que apresentam maior variância.

A partir da matriz de dados  $X$  de ordem  $m \times n$  podemos fazer uma estimativa da matriz de covariância  $S$ .

$$S = \begin{bmatrix} \text{Var}(x_1) & \text{Cov}(x_1x_2) & \text{Cov}(x_1x_3) & \cdots & \text{Cov}(x_1x_n) \\ \text{Cov}(x_2x_1) & \text{Var}(x_2) & \text{Cov}(x_2x_3) & \cdots & \text{Cov}(x_2x_n) \\ \text{Cov}(x_3x_1) & \text{Cov}(x_3x_2) & \text{Var}(x_3) & \cdots & \text{Cov}(x_3x_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(x_nx_1) & \text{Cov}(x_nx_2) & \text{Cov}(x_nx_3) & \cdots & \text{Var}(x_n) \end{bmatrix}$$

Pode-se observar que a matriz  $S$  é simétrica e de ordem  $n \times n$ . E deve-se atentar para quando os atributos são observados em unidades de medidas diferentes entre si, neste caso, é conveniente padronizar as variáveis  $X_i$  ( $i = 1, 2, 3, \dots, n$ ). Geralmente, a padronização é feita com média zero e variância 1, ou com variância 1 e média qualquer. De modo que a matriz  $Z$  das variáveis padronizadas  $Z_j$  é igual a matriz de correlação da matriz de dados  $X$ .

Para determinar as componentes principais normalmente toma-se como ponto de partida a matriz de correlação  $R$ . É importante observar que o resultado encontrado para a análise a partir da matriz  $S$  pode ser diferente do resultado encontrado a partir da matriz  $R$ .

As componentes principais são determinadas resolvendo-se a equação característica da matriz  $S$  ou  $R$ :

$$\det[R - \lambda I] = 0 \quad \text{ou} \quad |R - \lambda I| = 0.$$

$$R = \begin{bmatrix} 1 & r(x_1x_2) & r(x_1x_3) & \cdots & r(x_1x_n) \\ r(x_2x_1) & 1 & r(x_2x_3) & \cdots & r(x_2x_n) \\ r(x_3x_1) & r(x_3x_2) & 1 & \cdots & r(x_3x_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r(x_nx_1) & r(x_nx_2) & r(x_nx_3) & \cdots & 1 \end{bmatrix}$$

Se a matriz  $R$  for de posto completo igual a  $n$ , isto é, não apresentar nenhuma coluna que seja combinação linear de outra, a equação  $|R - \lambda I| = 0$  terá  $n$  raízes chamadas de autovalores ou raízes características da matriz  $R$ .

Sejam  $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n$  as raízes da equação característica da matriz  $R$  ou  $S$ , então:

$$\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_n$$

Para cada autovalor  $\lambda_i$  existe um autovetor

$$u_i = \begin{bmatrix} u_{i1} \\ u_{i2} \\ u_{i3} \\ \vdots \\ u_{in} \end{bmatrix}$$

Sendo  $u_i$  o autovetor correspondente ao autovalor  $\lambda_i$ , então a  $i$ -ésimo componente principal é dado por:

$$Y_i = u_{i1}X_1 + u_{i2}X_2 + u_{i3}X_3 + \dots + u_{in}X_n$$

As componentes principais apresentam as seguintes propriedades:

1) A variância do componente principal  $Y_i$  é igual ao valor do autovalor  $\lambda_i$ :

$$Var(Y_i) = \lambda_i;$$

2) A primeira componente é o que apresenta maior variância e assim por diante:

$$Var(Y_1) > Var(Y_2) > Var(Y_3) > \dots > Var(Y_n);$$

3) O total de variância das variáveis originais é igual ao somatório dos autovalores que é igual ao total de variância dos componentes principais:

$$\sum Var(X_i) = \sum \lambda_i = \sum Var(Y_i);$$

4) As componentes principais não são correlacionados entre si:

$$Cov(Y_i, Y_j) = 0.$$

A contribuição  $C_i$  de cada componente principal  $Y_i$  é expressa em porcentagem. É calculada dividindo-se a variância de  $Y_i$  pela variância total. A importância de uma componente

principal é avaliada por meio de sua contribuição, isto é, pela proporção de variância total explicada pela componente. A soma dos primeiros  $k$  autovalores representa a proporção de informação retida na redução de  $n$  para  $k$  dimensões. Com essa informação pode-se decidir quantas componentes usar, isto é, quantas componentes serão utilizadas para diferenciar os  $m$  exemplares observados.

## 2.3 Redes Neurais Artificiais - RNAs

Redes Neurais Artificiais [20, 21] são sistemas paralelos distribuídos compostos por unidades de processamento simples (nodos) que calculam determinadas funções matemáticas, normalmente não lineares. Tais unidades são dispostas em uma ou mais camadas e interligadas por um grande número de conexões, geralmente unidirecionais. Na maioria dos modelos estas conexões estão associadas a pesos, os quais armazenam o conhecimento representado no modelo e servem para ponderar a entrada recebida por cada neurônio da rede.

O funcionamento destas redes é inspirado em uma estrutura física concebida pela natureza: o cérebro humano, assemelhando-se em dois aspectos:

- O conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem;
- Forças de conexão entre neurônios, conhecidos como pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido.

O procedimento usual para realizar o processo de aprendizagem é chamado fase de aprendizagem. Sua função é modificar de forma ordenada os pesos sinápticos da rede, de modo que sejam extraídas automaticamente as características necessárias para representar a informação fornecida. A capacidade de aprender através de exemplos e de generalizar a informação aprendida é, sem dúvida, o atrativo principal da solução de problemas através de RNAs.

Muitos tipos de modelos de RNAs têm sido propostos, dentre os quais, pode-se destacar: As Redes *Multilayer Perceptron* - MLP [21, 22]; Redes Recorrentes [21]; Redes de Kohonen [43]; Redes de Hopfield [58]; e Redes ART [59]. No entanto, as redes MLPs são as mais populares devido a sua praticidade, flexibilidade e eficiência.

### 2.3.1 Rede *Multilayer Perceptron* - MLP

Redes de uma só camada, como o perceptron [60] resolvem apenas problemas linearmente separáveis. A solução de problemas não linearmente separáveis passa pelo uso de redes com

uma ou mais camadas intermediárias ou escondidas. Uma vez que, uma rede com uma camada intermediária pode implementar qualquer função contínua, com a utilização de duas camadas intermediárias permite-se a aproximação de qualquer função matemática. As redes MLPs têm sido aplicadas com sucesso em várias áreas, desempenhando tarefas como classificação/reconhecimento de padrões, controle e processamento de sinais.

Uma RNA do tipo MLP [21, 22] é constituída por um conjunto de nodos fonte, que formam a camada de entrada, uma ou mais camadas escondidas e uma camada de saída. Exceto a camada de entrada, todas as outras camadas são constituídas de nodos. O processamento realizado por cada nodo é definido pela combinação dos processamentos realizados pelos nodos da camada anterior que estão conectados a ele. É possível dizer que as unidades intermediárias de uma rede MLP funcionam como detectores de características. Elas geram uma codificação interna dos padrões de entrada, que é então utilizada para a definição da saída da rede. Dado um número suficientemente grande de unidades intermediárias, é possível formar representações internas para qualquer conjunto de padrões de entrada. Se cada nodo de uma determinada camada for conectado a todos os nodos da próxima camada, tem-se uma rede MLP totalmente conectada, como é ilustrado na Figura 2.

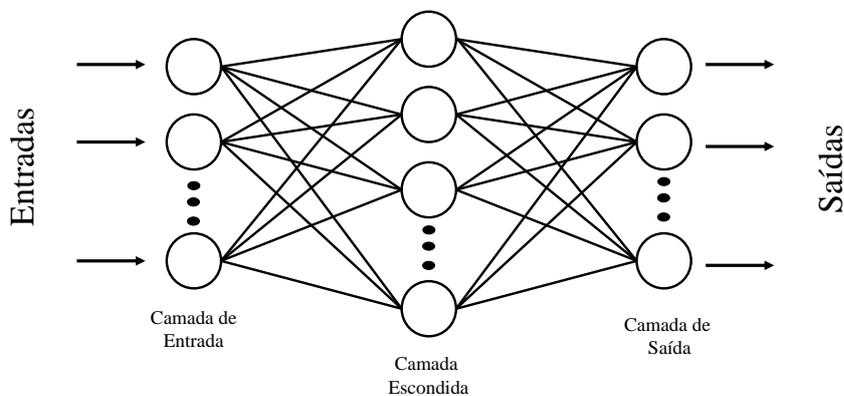


Figura 2: Rede MLP de três camadas totalmente conectadas.

Para treinar as redes com mais de uma camada, foi proposto um método que se baseia em gradiente descendente [22]. A fim de que este método possa ser utilizado, a função de ativação precisa ser contínua, diferenciável e preferencialmente não decrescente.

Um dos principais aspectos relacionados ao projeto de redes MLP diz respeito à função de ativação utilizada, dado que esta deve informar os erros cometidos pela rede para as camadas anteriores com a maior precisão possível. Desse modo, buscando superar os problemas citados,

diversas funções de ativação têm sido propostas para redes multicamadas. As mais usuais são:

- Função Sigmóide Logística

$$f(x) = \frac{1}{1 + \exp^{-\lambda x}}$$

em que  $\lambda$  é um parâmetro que determina a sensibilidade da resposta da função, comumente ajustado para 1.

- Função Tangente Hiperbólica

$$f(x) = \frac{\exp^x - \exp^{-x}}{\exp^x + \exp^{-x}}$$

- Função Seno ou Cosseno

$$f(x) = \text{sen}(x) \text{ ou } f(x) = \text{cos}(x)$$

- Função Linear

$$f(x) = \alpha x + \beta$$

Na Figura 3 são exibidos os gráficos dessas funções.

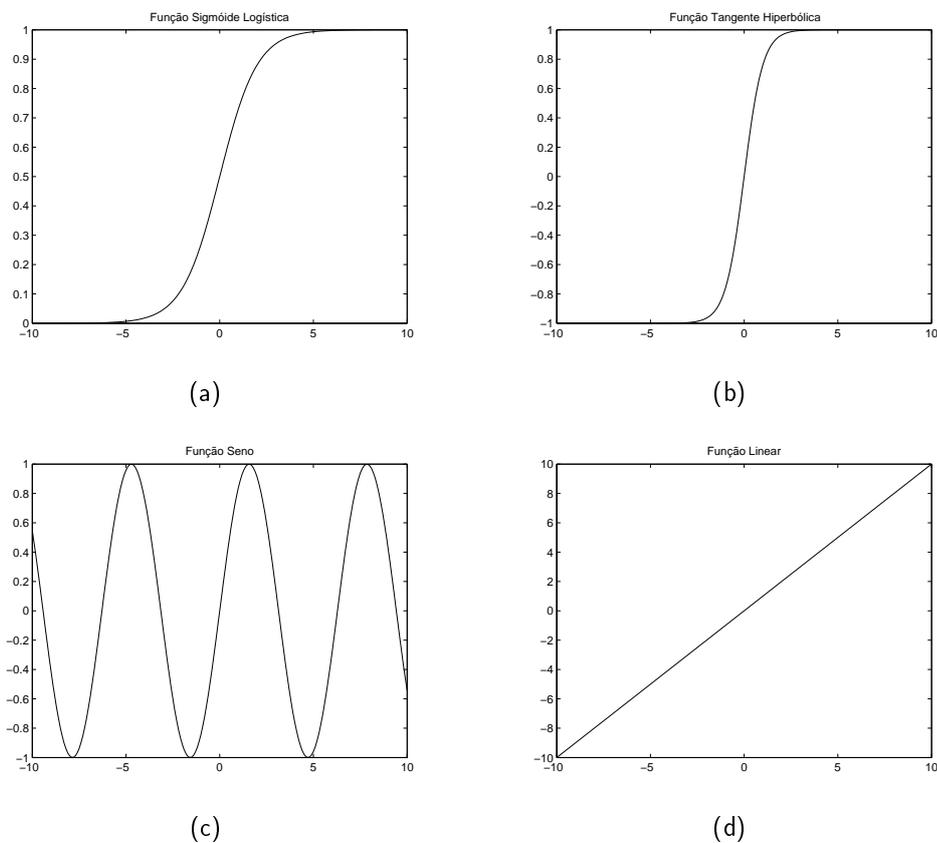


Figura 3: Gráficos das funções de ativação mais comuns: (a) Função Sigmóide Logística, (b) Função Tangente Hiperbólica, (c) Função Seno e (d) Função Linear.

Os algoritmos de aprendizagem das RNAs geralmente são do tipo supervisionado. O

mais conhecido para treinamento destas redes é o algoritmo *backpropagation* [22]. A maioria dos métodos de aprendizado para RNAs do tipo MLP utiliza variações deste algoritmo. O *backpropagation* utiliza pares (entrada, saída desejada) para, por meio de um mecanismo de correção de erros, ajustar os pesos da rede. O treinamento ocorre em duas fases, fase *forward* e fase *backward*, em que cada fase percorre a rede em um sentido. A fase *forward* é utilizada para definir a saída da rede para um dado padrão de entrada. A fase *backward* utiliza a saída desejada e a saída fornecida pela rede para atualizar os pesos de suas conexões.

Para a concepção de uma RNA, a definição da arquitetura é fundamental. No projeto da arquitetura de uma MLP os seguintes parâmetros devem ser definidos:

- A quantidade de entradas;
- O número de camadas escondidas da rede e o número de nodos em cada camada;
- A quantidade de saídas.

Além destes parâmetros, o tipo de conexão entre os nodos e a topologia da rede também devem ser definidas. A seleção destes parâmetros depende do problema a ser resolvido. Na maioria dos casos, a arquitetura é definida com base em experimentos preliminares. É interessante observar que arquiteturas e topologias diferentes podem resolver o mesmo problema.

## 2.4 Métodos de Classificação

### 2.4.1 *k-Nearest Neighbor* - kNN

O *k-Nearest Neighbor* [1, 9] é um método de classificação de padrões onde o aprendizado é baseado na analogia. O conjunto de treinamento é formado por vetores  $n$ -dimensionais e cada elemento deste conjunto representa um ponto no espaço  $n$ -dimensional. Consiste em atribuir uma classe a um elemento desconhecido usando a classe da maioria de seus vizinhos mais próximos, segundo uma determinada distância no espaço de atributos.

A métrica mais utilizada no cálculo da distância entre dois pontos no kNN é a distância Euclidiana. É calculada da seguinte forma:

Sejam  $X = (x_1, x_2, \dots, x_n)$  e  $Y = (y_1, y_2, \dots, y_n)$  dois pontos do  $\mathbb{R}^n$ .

A distância Euclidiana entre  $X$  e  $Y$  é dada por:

$$d(X, Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

O método kNN apresenta-se de modo muito simples:

Dado um conjunto  $X$  de tuplas de treinamento. Cada elemento de  $X$  é uma tupla  $(x_1, x_2, \dots, x_n, c)$ , onde  $c$  é a classe à qual pertence a tupla  $(x_1, x_2, \dots, x_n)$ , e esta tupla pode ser vista como um ponto em um espaço  $n$ -dimensional. Seja  $Y = (y_1, y_2, \dots, y_n)$  uma nova tupla, ainda não classificada. A fim de classificá-la, calculam-se as distâncias de  $Y$  a todas as tuplas de treinamento e consideram-se as  $k$ -tuplas de treinamento mais próximas de  $Y$  (as que apresentam menor distância). Dentre estas  $k$ -tuplas, verifica-se qual a classe de padrões aparece com mais frequência. A tupla  $Y$  será classificada dentro desta classe mais frequente.

O parâmetro  $k$  indica o número de vizinhos que serão usados pelo algoritmo durante a fase de teste. Este parâmetro faz com que o algoritmo consiga uma classificação mais refinada, porém o seu valor ótimo varia de um problema para outro, sendo comum considerar  $k < \sqrt{N}$ , em que  $N$  é o número de tuplas do conjunto de treinamento.

Um exemplo simples de como funciona a classificação feita pelo algoritmo kNN é ilustrado na Figura 4, em que é dado um padrão desconhecido  $x_d$  a ser classificado. O kNN o classificará como sendo pertencente a uma das classes definidas no problema. Dependendo do número de vizinhos  $k$ ,  $x_d$  será classificado da seguinte forma: se  $k = 1$ ,  $x_d$  será classificado como elemento da classe 2; para  $k = 3$ ,  $x_d$  será novamente classificado como elemento da classe 2; no entanto, se  $k = 5$ ,  $x_d$  será classificado na classe 1.

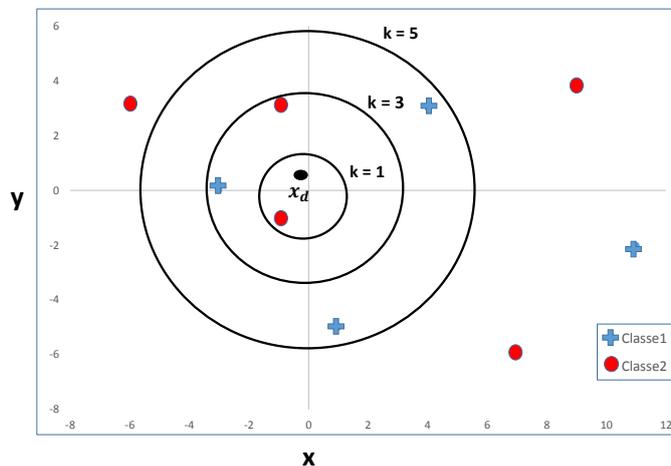


Figura 4: Classificação feita pelo kNN para um padrão desconhecido  $x_d$ .

## 2.4.2 *K-means*

*K-means* [8] é um processo que particiona um conjunto de dados  $n$ -dimensional em  $k$  subconjuntos (classes), fornecendo partições que são razoavelmente eficientes no sentido de variância mínima.

Para aplicação do método *K-means* é oferecido o conjunto de dados  $X$  a ser particionado e um número  $k$ , representando o número de classes que se deseja formar entre os elementos do banco de dados.

O processo de classificação *K-means* pode ser descrito como segue:

Sejam  $C_j$ ,  $j = 1, 2, \dots, k$  uma partição do conjunto de dados e  $x_i \in X$ , um exemplo  $n$ -dimensional no conjunto de dados.

1. Defina aleatoriamente  $k$  centróides;
2. Para cada exemplo  $x_i \in X$  e cada grupo  $C_j$ ,  $j = 1, 2, \dots, k$ ,  
Calcular a distância entre  $x_i$  e o centróide de  $C_j$ ;
3. Para cada exemplo  $x_i$ ,  
Associar  $x_i$  ao grupo com centróide mais próximo;
4. Para cada grupo  $C_j$ ,  
Recalcular o centróide;
5. Repetir os passos anteriores até que algum critério de parada seja satisfeito.

Comumente utiliza-se como critério de parada a execução de um determinado número de iterações ou a estabilização dos centróides. Estes critérios podem ser utilizados individualmente ou conjuntamente.

O método *K-means* busca minimizar a distância entre cada ponto e o centróide do grupo ao qual pertence.

$$E = \sum_{j=1}^k \sum_{x_i \in C_j} d(x_i, m_j)$$

em que  $m_j$  é o vetor centróide do grupo  $C_j$  e  $d(x_i, m_j)$  é a distância Euclidiana entre  $x_i$  e  $m_j$ .

Para isso, o erro quadrático deve ser minimizado. Este erro é definido pela equação:

$$E = \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - m_j\|^2.$$

De modo informal, pode-se dizer que o procedimento *K-means* consiste simplesmente em começar com  $k$  grupos, cada um dos quais formado por um único ponto aleatório e, posteriormente adicionar cada novo ponto observado ao grupo cujo valor médio for mais próximo. Assim, em cada iteração, as *K-means* são de fato as médias dos grupos que representam. Na

Figura 5 é ilustrada a aplicação do método *K-means*. Para um  $k = 3$ , foram realizadas cinco iterações, dado que da quarta para a quinta iteração houve uma estabilização do valor dos centróides.

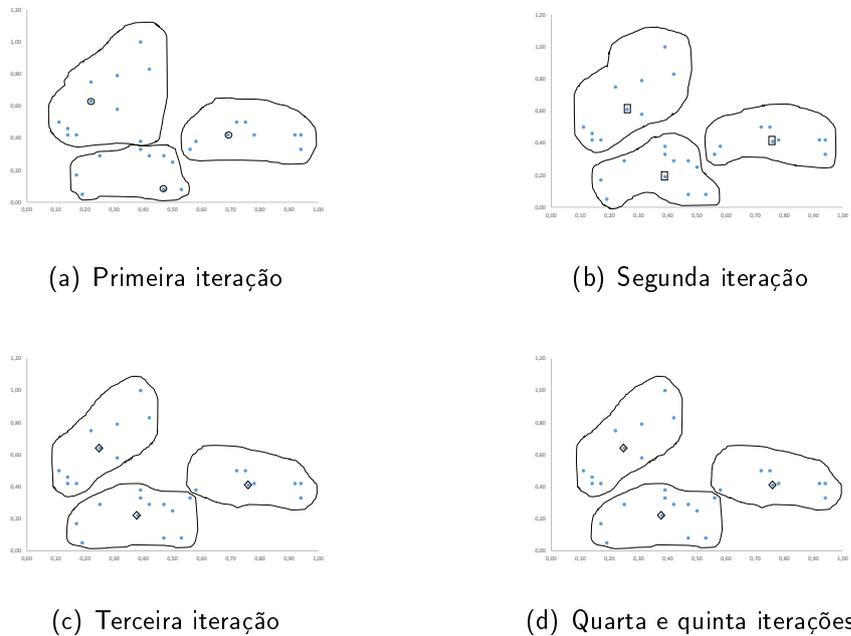


Figura 5: Funcionamento do método *k-means* para três classes ( $k = 3$ ).

### 2.4.3 Análise Discriminante Linear - LDA

A Análise Discriminante Linear (do inglês *Linear Discriminant Analysis* - LDA), também conhecida como Discriminante Linear de Fisher [10] é uma técnica comumente usada para classificação e redução de dimensionalidade de dados. Busca-se encontrar uma transformação linear através da maximização da variância entre classes e minimização da variância intra classe em qualquer conjunto particular de dados, com isso garantindo separabilidade máxima.

LDA assemelha-se ao PCA como um processo de classificação e redução de dimensionalidade de dados, no entanto, LDA não muda a localização dos dados, como acontece com o PCA, apenas tenta fornecer mais separabilidade das classes, desenhando uma região de decisão entre elas. O método tenta encontrar a melhor direção dos eixos, de modo que quando projetados os dados em um plano, as classes possam ser separadas.

O ponto de partida para uma Análise Discriminante é uma matriz  $X$  de dados  $n$ -dimensional e o conhecimento do número de classes  $k$  em que se distribuem os  $N$  objetos observados (cada objeto pertence a apenas uma classe). Busca-se determinar funções das  $n$  variáveis observadas que melhor permitam a distinção (ou discriminação) desses subgrupos (ou classes).

Conforme descrito em [17], para um conjunto de  $N$  objetos rotulados  $x_1, x_2, \dots, x_N$  em  $\mathbb{R}^n$  que pertencem a  $k$  diferentes classes, a função objetivo do LDA ( $w^*$ ) é obtida como segue:

$$w^* = \arg \max_w \frac{w^T S_b w}{w^T S_w w}. \quad (2.1)$$

Em que:

$$S_b = \sum_{j=1}^k N_j (\mu_j - \mu)(\mu_j - \mu)^T$$

e

$$S_w = \sum_{j=1}^k \sum_{i=1}^{N_j} (x_{i,j} - \mu_j)(x_{i,j} - \mu_j)^T$$

- $N_j$  é o número de objetos na  $j$ -ésima classe;
- $\mu$  é a média de todo o conjunto de dados;
- $\mu_j$  é a média da  $j$ -ésima classe;
- $x_{i,j}$  é o  $i$ -ésimo objeto da  $j$ -ésima classe;
- $S_b$  e  $S_w$  são as matrizes de dispersão intra-classe e entre-classes, respectivamente.

A matriz de dispersão total é definida por

$$S_t = \sum_{i=1}^N (x_i - \mu)(x_i - \mu)^T.$$

Assim, a função objetivo do LDA, definida da Equação 2.1, será equivalente a:

$$w^* = \arg \max_w \frac{w^T S_b w}{w^T S_t w}.$$

Os  $w$ 's são os autovetores correspondentes aos autovalores diferentes de zero.

$$\frac{S_b}{S_t} w = \lambda w.$$

Sem perda de generalidade, pode-se assumir  $\mu = \mathbf{0}$  ( $\mathbf{0}$  é o vetor nulo). Então

$$S_b = \sum_{j=1}^k N_j \mu_j \mu_j^T = \sum_{j=1}^k X_j P_j X_j^T = X P X^T$$

em que  $P_j$  é uma matriz  $N_j \times N_j$  com todos os elementos iguais a  $1/N_j$ ,  $X_j = [x_{1,j}, \dots, x_{N_j,j}]$  é a matriz de dados da  $j$ -ésima classe,  $X = [X_1, X_2, \dots, X_k]$  e  $P$  é a matriz  $N \times N$  como segue:

$$P = \begin{bmatrix} P_1 & 0 & \dots & 0 \\ 0 & P_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P_k \end{bmatrix}$$

Portanto, a função objetivo pode ser representada como:

$$w^* = \arg \max_w \frac{w^T X P X^T w}{w^T X X^T w}.$$

## 3 Inspiração Quântica

Um sistema quântico é governado pela mecânica quântica [24]. A mecânica quântica é um ramo da Física que lida com fenômenos em uma escala microscópica, onde a magnitude das variáveis são da ordem da Constante de Planck ( $h = 6.626069 \times 10^{-34} J \cdot s$ ). Neste sentido, a mecânica quântica é uma teoria mais fundamental do que a mecânica clássica, pois um sistema macroscópico pode ser entendido como a composição de sistemas microscópicos.

Matematicamente, os estados possíveis de um sistema da mecânica quântica podem ser representados por um vetor unitário, chamado vetor de estados e representado na notação de Dirac por  $|\Psi\rangle$ . Formalmente, estes vetores de estado residem em um espaço de Hilbert complexo [61], comumente chamado de espaço de estados do sistema quântico. Este espaço de estado é bem definido como um número complexo de norma 1.

### 3.1 Estado Quântico

Em mecânica quântica, um estado quântico refere-se ao estado de um sistema quântico, dado por um vetor no espaço de Hilbert, chamado vetor de estado. De modo geral, um estado quântico pode ser puro ou misto. Matematicamente, um estado quântico puro é representado por um vetor de estado no espaço de Hilbert sobre os números complexos, como uma generalização do usual espaço tridimensional [62]. E pode ser graficamente representado na esfera de Bloch [24, 25], como mostrado na Figura 6.

Para o computador quântico [25] a base canônica do espaço é  $|0\rangle$  e  $|1\rangle$ , onde

$$|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle,$$

com

$$|\alpha|^2 + |\beta|^2 = 1.$$

O conceito matemático do espaço de Hilbert, generaliza a noção de espaço euclidiano. Estendendo-se do plano Euclidiano bidimensional e/ou do espaço tridimensional para espaços

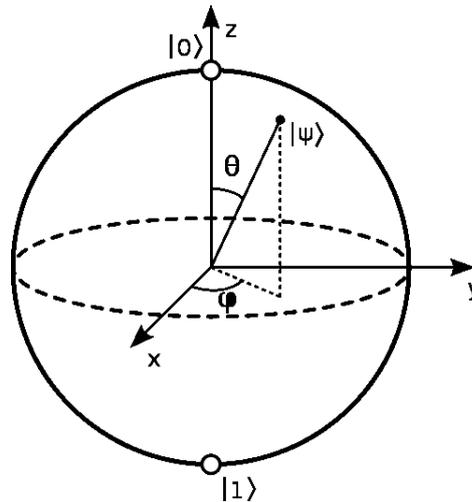


Figura 6: A Esfera de Bloch.  $|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle$ .

com qualquer número finito ou infinito de dimensões. Intuitivamente, o espaço de Hilbert nada mais é do que um espaço vetorial complexo gozando da estrutura de um produto interno, ou seja, um espaço vetorial no qual os vetores são expressos por números complexos (com um módulo e uma fase que podem ser medidos). Uma maneira de intuir um espaço vetorial complexo de  $N$  dimensões é imaginar um espaço euclidiano de dimensão  $N$  no qual cada vetor tem  $N - 1$  fases associadas, onde cada fase pode variar entre  $0$  e  $2\pi$ .

A cada instante, um estado de um sistema quântico pode ser representado por um único vetor no espaço de Hilbert correspondente. E pelo princípio de superposição, qualquer combinação linear de dois estados possíveis para um sistema quântico, também é um estado possível [63]. Dessa forma, o espaço de Hilbert fornece um meio pelo qual podemos considerar funções como pontos pertencentes a um espaço de dimensão infinita.

## 3.2 Observável Quântico

Uma medida é simplesmente uma operação física bem definida que, quando realizada sobre um sistema físico, produz um único número real sem qualquer erro ou ambiguidade no sentido de que não existe incerteza experimental associada com o número obtido. Uma medida de um observável quântico afeta o sistema, e por isso pode ser esperado aparecer explicitamente na teoria. Em outras palavras, não é possível medir dois operadores observáveis diferentes de um sistema simultaneamente, a medida de um pode mudar o valor do outro.

As operações de medida geralmente afetam o sistema. Portanto, é quase impossível prever uma única medida com certeza absoluta. Naturalmente, esta dificuldade pode, de alguma maneira ser resolvida pelo procedimento com um grande número de medidas em

condições idênticas e então introduzindo a média de todos os resultados obtidos. Esta média, chamada valor esperado de um operador observável é definida como segue.

O valor esperado  $\langle A \rangle$  de um operador observável  $A$  em um estado  $\Psi(x)$  de um sistema físico é definido por:

$$\langle A \rangle = \frac{\langle \Psi | A | \Psi \rangle}{\langle \Psi | \Psi \rangle}.$$

Se o estado  $\Psi$  é normalizado, então o valor esperado é  $\langle A \rangle = \langle \Psi | A | \Psi \rangle$ .

Na notação de Dirac [24], é possível definir um operador  $A$  onde a equação é válida,

$$A|\Psi\rangle = \lambda|\Psi\rangle. \quad (3.1)$$

O  $|\Psi\rangle$  é um autovetor e  $\lambda$  é um autovalor. Em geral,  $\lambda \in \mathbb{C}$ , mas se  $\lambda \in \mathbb{R}$  implica que o operador  $A$  é Hermitiano e o valor de  $\lambda$  pode ser medido no mundo real. No último caso, o operador  $A$  é chamado *observável* e o seu valor pode ser obtido por uma medida do valor esperado dado por,

$$\langle \Psi | A | \Psi \rangle = \langle \Psi | \lambda | \Psi \rangle = \lambda \langle \Psi | \Psi \rangle = \lambda$$

onde toda informação sobre as fases (Figura 6:  $\theta$  e  $\varphi$ ) é perdida.

Um observável quântico é matematicamente representado por um operador Hermitiano linear em um espaço de Hilbert separável de dimensão infinita e existe um tal operador para cada observável quântico, tais como a posição, o momento, a energia, e assim segue. Tais operadores em mecânica quântica são chamados operadores observáveis, e usualmente satisfazem as regras da teoria dos operadores no espaço de Hilbert. Assim, sendo os observáveis quânticos, operadores no espaço de Hilbert. Entretanto, um observável em mecânica clássica é simplesmente uma função real de variáveis conjugadas no espaço Euclidiano.

A diferença mais profunda entre as regras de dois operadores e as regras para manipulação de observáveis clássicos é: dois observáveis  $A$  e  $B$  sempre comutam,  $AB = BA$ , mas dois operadores observáveis  $A$  e  $B$  em geral não comutam. O comutador destes operadores é definido por  $[A, B] = AB - BA$ . Assim, em geral, o comutador de dois operadores observáveis é não nulo. Fisicamente, o valor não nulo deve ser associado com a magnitude dos distúrbios inevitáveis entre as duas medidas de  $A$  e  $B$  em duas diferentes ordens.

## 4 Metodologia Proposta

Método proposto para o incremento da dimensão de um conjunto de dados, possibilitando a separabilidade de seus elementos em classes. Técnicas usadas para obtenção do incremento da dimensionalidade e realização do mapeamento de novos pontos no espaço de representação criado com a transformação. Como também o processo usado para validar a metodologia e a descrição dos conjuntos de dados usados em sua aplicação.

A heurística aqui assume que o conjunto de dados é composto pelos valores de um observável no mundo quântico, a menos de uma constante multiplicativa. Assim, o procedimento irá procurar por um vetor de estado  $|\Psi\rangle$  onde um operador projetivo observável aplicado a este vetor de estado gera os dados analisados.

### 4.1 Metodologia

Seja  $X = \{x_1^{(d)}, x_2^{(d)}, \dots, x_N^{(d)}\}$  um conjunto de dados a ser classificado em  $C$  classes, em que cada  $x_j^{(d)} \in \mathbb{R}^d$ . Se esta tarefa de classificação referir-se a um problema de classificação de padrões não linearmente separável, sua solução não será trivial. De acordo com o teorema de Cover [7], problemas assim podem ser resolvidos com uma transformação de seu espaço de representação para um espaço de maior dimensionalidade. O teorema, em termos qualitativos, postula que: um problema complexo de classificação de padrões disposto não linearmente tem maior probabilidade de ser linearmente separável em um espaço de alta dimensão do que em um espaço de baixa dimensionalidade.

Inspirada nesse teorema, a proposta desse processo é incrementar a dimensionalidade do conjunto de dados. Para tanto, propõe-se uma aplicação da seguinte forma:  $x_j^{(d)} \rightarrow z_j^{(d)}$ , em que  $z_j^{(d)} \in \mathbb{C}^d$ . Assim, para cada  $x_j^{(d)}$  existe um par ordenado  $(a_j^{(d)}, b_j^{(d)})$ , satisfazendo a restrição

$$(x_j^{(d)})^2 = |z_j^{(d)}|^2 = (a_j^{(d)})^2 + (b_j^{(d)})^2. \quad (4.1)$$

Além disso,  $z_j^d$  é associado ao estado quântico  $|\Psi\rangle$  (a menos de uma constante multiplicativa - a fase), e existe um operador observável  $T$  que quando aplicado ao estado quântico fornece o autovalor  $|z_j^{(d)}|^2$ ,

$$\langle\Psi|T|\Psi\rangle = |z_j^{(d)}|^2|\Psi\rangle.$$

Para obter o par  $(a_j^{(d)}, b_j^{(d)})$ , foi usado um algoritmo genético modificado proposto em Ferreira *et al* [53] e descrito na próxima subseção. Com a aplicação da heurística proposta, um padrão  $x_j^{(d)}$   $d$ -dimensional no espaço real torna-se  $d$ -dimensional no espaço complexo ( $x_j^{(d)} \rightarrow z_j^{(d)}$ ).

Uma vez que, após medir o valor do observável toda informação sobre a fase do estado quântico é perdida, os valores de  $|z_j^{(d)}|^2$  serão conhecidos, porém o mesmo não acontece com as fases que os determinam. Assim, o algoritmo genético busca um par ordenado  $(a, b)$  que determina tais parâmetros restrito à condição de maximização do poder de discriminação dos dados. Esta situação é alcançada se,

- A variância em cada classe é minimizada;
- As distâncias entre os centróides das classes são mutuamente maximizadas.

Portanto, uma breve descrição da metodologia proposta pode ser dada por:

1. Para cada padrão  $x_j^{(d)}$  do conjunto de dados a ser classificado, o algoritmo genético encontra um par ordenado  $(a_j^{(d)}, b_j^{(d)})$  que satisfaz à restrição  $(x_j^{(d)})^2 = |z_j^{(d)}|^2 = (a_j^{(d)})^2 + (b_j^{(d)})^2$ , ou seja, o quadrado do padrão é tratado como sendo o quadrado do módulo de um número complexo e uma busca do par ordenado que o determina é realizada;
2. Na etapa seguinte, a partir da função de *fitness*, o algoritmo genético ajusta todos os pares  $(a_j^{(d)}, b_j^{(d)})$  gerando as fases que irão discriminar os dados, assim, um novo conjunto de dados é criado com o dobro da dimensionalidade do conjunto original;
3. Uma vez gerado o novo conjunto de dados com a dimensionalidade incrementada, algum algoritmo de classificação é aplicado para essa tarefa.

Seguindo essa metodologia é possível gerar uma representação no espaço dos números complexos, ou seja, dobrar o número de dimensões do conjunto de dados com o propósito de ter menor complexidade na tarefa de classificação comparada a mesma tarefa no espaço real, como postulado no teorema de Cover.

## 4.2 A transformada inversa - Algoritmo Genético

A população é composta de indivíduos, ou cromossomos, cada um representando um possível conjunto de pares  $(a_j^{(d)}, b_j^{(d)})$  correspondendo ao padrão  $x_j^{(d)}$  do conjunto de dados original. Desta forma, um indivíduo é um conjunto de  $N$  3-tuplas como  $(a_j^{(d)}, b_j^{(d)}, k_j)$ , onde  $k_j$  é um rótulo indicando a classe do padrão  $x_j^{(d)}$  e  $N$  é o número de padrões no conjunto de dados original.

O algoritmo genético modificado usado aqui foi utilizado por Ferreira *et al* [53]. Em cada geração, dois cromossomos na população são selecionados para submeter-se a operações genéticas. O processo de seleção é feito pelo método da roleta [64].

Os operadores genéticos incluem operadores de cruzamento e mutação. O operador cruzamento troca informação a partir de dois pais ( $\mathbf{p}_1$  e  $\mathbf{p}_2$ ). Aqui, quatro operadores de cruzamento foram usados,

$$C_1 = \frac{\mathbf{p}_1 + \mathbf{p}_2}{2} \quad (4.2)$$

$$C_2 = \mathbf{P}_{max}(1 - w) + \max(\mathbf{p}_1, \mathbf{p}_2)w \quad (4.3)$$

$$C_3 = \mathbf{P}_{min}(1 - w) + \min(\mathbf{p}_1, \mathbf{p}_2)w \quad (4.4)$$

$$C_4 = \frac{(\mathbf{P}_{max} + \mathbf{P}_{min})(1 - w) + (\mathbf{p}_1 + \mathbf{p}_2)w}{2} \quad (4.5)$$

onde  $\mathbf{P}_{max}$  e  $\mathbf{P}_{min}$  são os vetores com os valores máximos e mínimos possíveis para cada gene de um cromossomo, respectivamente. A função matemática  $\max$  retorna o valor máximo para cada gene de  $\mathbf{p}_1$  e  $\mathbf{p}_2$ , e a função matemática  $\min$ , o valor mínimo.  $w \in \mathbb{R}$  é um peso, onde nos experimentos foi usado o valor  $w = 0,8$ , definido a partir de tentativa e erro.

Após a potencial prole ser gerada pelo operador de cruzamento, o melhor descendente é escolhido. Se este melhor descendente é melhor do que o pior cromossomo da velha população, então este descendente substituirá o pior cromossomo.

Os quatro novos cromossomos gerados pelo processo de cruzamento ( $C_1, C_2, C_3, C_4$ ) são clonados e seus clones submetem-se a operação de mutação, onde as características herdadas de seus pais podem ser mudadas. Para cada descendente clonado, três novos cromossomos

são gerados pelo operador mutação,

$$MC_{l,\alpha} = [c_l^1 c_l^2 \cdots c_l^{no\_Vars}] + [\delta_1 mc_l^1 \delta_2 mc_l^2 \cdots \delta_{no\_Vars} mc_l^{no\_Vars}] \quad (4.6)$$

onde  $\alpha = 1, 2, 3$  é o índice de mutação,  $l = 1, 2, 3, 4$  é o índice de descendente,  $\delta_u$  ( $u = 1, 2, \dots, no\_Vars$  e  $no\_Vars$  é o número de genes no cromossomo) pode tomar somente os valores 0 ou 1, e  $mc_l^u$  ( $u = 1, 2, \dots, no\_Vars$ ) são números gerados aleatoriamente que satisfazem a restrição  $parameter_{min}^u \leq c_l^u + mc_l^u \leq parameter_{max}^u$ .

A primeira operação de mutação ( $\alpha = 1$ ) é tal que apenas um  $\delta_u$  é 1 ( $u$  sendo gerado aleatoriamente dentro do intervalo válido) e todos os outros são igual a zero na Equação 4.6. A segunda operação de mutação ( $\alpha = 2$ ) é obtida com todo  $\delta_u$  igual a 1 na Equação 4.6.

Um número real é gerado aleatoriamente e comparado a um número definido pelo usuário  $p_{Mut} \in [0, 1]$  (probabilidade de mutação aceita, aqui  $p_{Mut} = 0, 1$ , definida a partir de tentativa e erro). Se o número real é menor do que  $p_{Mut}$  então o cromossomo mutado substitui o cromossomo com menor *fitness* na população. Entretanto, se o número real é maior do que  $p_{Mut}$ , então o cromossomo mutado substitui o cromossomo com o menor *fitness* da população se e somente se seu *fitness* é maior do que o *fitness* do pior cromossomo na população.

Todos os indivíduos (*ind*) são avaliados de acordo com a função de *fitness* dada por,

$$f(ind) = \frac{\sum_{t=1}^k \sum_{j>t}^k \left( Z_C^{(t)} - Z_C^{(j)} \right)^2}{1 + \sum_{g=1}^k \sum_{v>g}^k \sqrt{\frac{1}{n_v-1} \sum_{w=1}^{n_v} \left( Z_C^{(g)} - z_w^{(v)} \right)^2}} \quad (4.7)$$

em que  $Z_C^{(l)} = \frac{1}{n_l} \sum_{j=1}^{n_l} a_j^{(l)} + ib_j^{(l)}$  é o centróide da classe  $l$ ,  $n_l$  é o número de elementos da classe  $l$ ,  $i = \sqrt{-1}$  é a constante complexa e  $z_w^{(l)} = a_w^{(l)} + ib_w^{(l)}$  é um ponto no espaço complexo de acordo com o  $w$ -ésimo ponto real da classe  $l$  no conjunto de dados original. O objetivo da função definida na Equação 4.7 é maximizar a distância entre clusters (ou classes) e minimizar a variância de cada cluster no espaço complexo.

### 4.3 A transformada - PCA

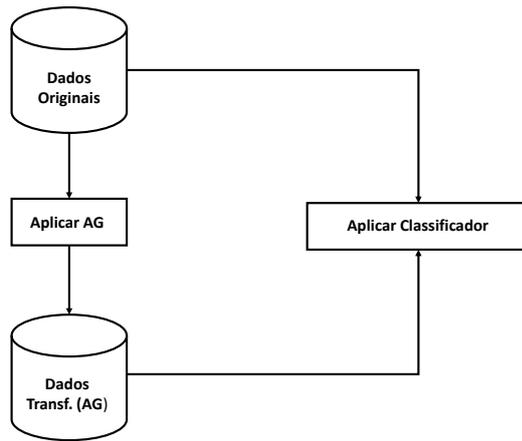
Uma vez produzido pelo Algoritmo Genético (AG) um espaço de representação  $2d$ -dimensional que fornece boa performance na discriminação de dados, o método PCA é aplicado no intuito de gerar um conjunto de dados de dimensão reduzida, mas que carregue toda a informação necessária para manter um bom desempenho nessa tarefa.

Após aplicar o PCA ao conjunto de dados produzido pelo AG e considerando as  $d$  primeiras dimensões, um espaço de representação  $d$ -dimensional é construído e é esperado que as classes se tornem separáveis com alta probabilidade, como ocorre quando a dimensionalidade é maior. Embora a ideia principal da metodologia proposta seja expandir a dimensão dos conjuntos de dados, o número de componentes principais usada será  $d$ -dimensional no intuito de não explodir a dimensionalidade do espaço de representação.

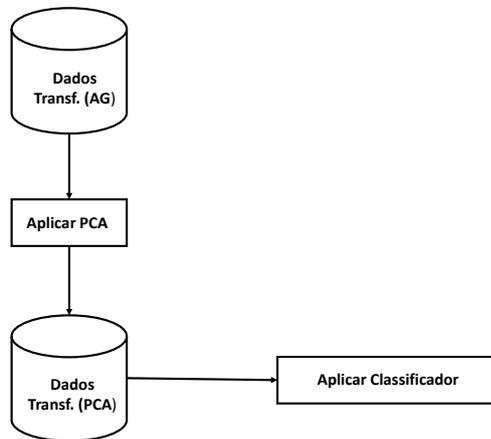
## 4.4 Mapeamento de novos pontos

Dado um novo ponto para teste, que não tenha sido apresentado ao processo de aprendizagem realizado pela metodologia proposta, como podemos classificá-lo? Um modelo de rede neural artificial *multilayer perceptron* (MLP) [22] é aplicado para aprender o mapeamento entre o conjunto de dados original e o sistema quântico no novo espaço de representação complexo, a entrada da RNA, portanto são os dados originais e a saída são os dados transformados pela PCA.

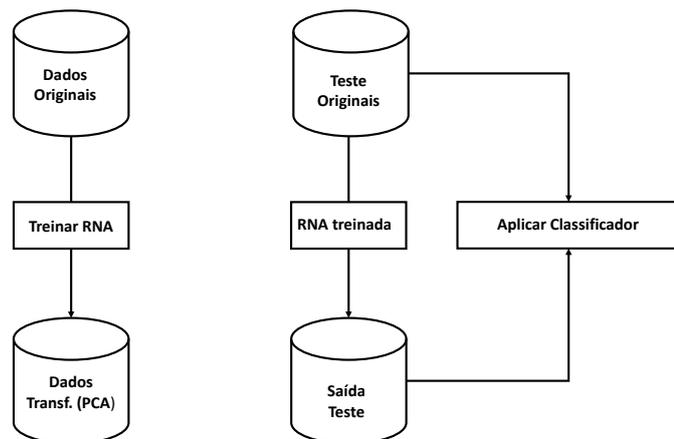
A metodologia proposta é ilustrada na Figura 7.



(a) Etapa I - O Algoritmo Genético.



(b) Etapa II - Método PCA.



(c) Etapa III - Mapeamento de novos pontos com uma RNA.

Figura 7: Ilustração da metodologia proposta.

Na Etapa I (Figura 7(a)) é aplicado um algoritmo genético que realiza a transformação dos dados, aumentando sua dimensionalidade. A seguir, os métodos de classificação são aplicados tanto aos dados originais, quanto aos dados transformados, no intuito de verificar se de fato há um melhor desempenho na tarefa de classificação quando a dimensionalidade é incrementada.

Na Etapa II (Figura 7(b)) o método PCA é aplicado a fim de reduzir os dados a partir da combinação das variáveis geradas na etapa anterior, preservando o máximo de características. Os métodos de classificação são aplicados no intuito de observar se a classificação é coerente com aquela obtida quando os dados são transformados pelo AG, ou seja, se o máximo de características é mantido.

A Etapa III (Figura 7(c)) consiste no treinamento de uma RNA, com o objetivo de mapear um novo dado no espaço de representação criado na etapa II. De modo que para o treinamento dessa RNA os dados originais serão o conjunto de dados de entrada e aqueles transformados na etapa II serão o conjunto de dados de saída. A princípio, 20% do conjunto de dados são preservados para o conjunto de teste. Tanto os dados preservados, quanto os dados de saída do conjunto de teste gerados pela RNA, são classificados para constatação da ocorrência de um mapeamento coerente.

## 4.5 Validação da metodologia

O conceito central das técnicas de validação cruzada é o particionamento do conjunto de dados em subconjuntos mutuamente exclusivos, onde posteriormente alguns destes subconjuntos são utilizados para treinamento do modelo, enquanto o restante dos subconjuntos (dados de validação ou de teste) são empregados na validação.

O *k-fold cross-validation* [65] consiste em dividir o conjunto de dados original em  $k$  subconjuntos aproximadamente do mesmo tamanho e mutuamente exclusivos. Cada subconjunto é usado uma vez como conjunto de teste, enquanto os  $k - 1$  restantes são utilizados para a tarefa de treinamento, dessa maneira todos os dados são usados para treinar e testar a metodologia.

Este processo é realizado  $k$  vezes alternando de forma circular o subconjunto de teste. Ao final das  $k$  iterações calcula-se a acurácia da metodologia, de modo que a estimativa da precisão é o número total de classificações corretas, dividido pelo número de exemplos no conjunto de dados, ou seja, uma média dos acertos de cada conjunto de teste. Além disso, a variabilidade entre estes valores também é calculada para obtenção de uma medida mais confiável sobre a potencialidade da metodologia proposta.

## 4.6 Dados

Para testar a metodologia proposta são utilizados dez conjuntos de dados, dispostos em duas, três e quatro classes, com diferentes dimensionalidades.

Para todos os conjuntos usados, os dados foram normalizados no intervalo  $[0,1]$  pela seguinte equação

$$x_n^{(d)} = \frac{x_j^{(d)} - x_{min}^{(d)}}{x_{max}^{(d)} - x_{min}^{(d)}} \quad (4.8)$$

em que  $x_j^{(d)}$ ,  $x_{min}^{(d)}$ ,  $x_{max}^{(d)}$  são respectivamente, o valor original do padrão dos dados, mínimo e máximo do padrão  $x_j^{(d)}$ .

### 4.6.1 Dados das espécies de Mata Atlântica pernambucana

Silva *et al* [66], se propuseram classificar as espécies da Mata Atlântica pernambucana em grupos funcionais, que são formados com base nas características funcionais (morfológicas, fisiológicas e/ou funcionais) das espécies.

Entende-se por espécies com baixo investimento na construção dos tecidos, aquelas com alta área foliar específica e concentração de nutrientes, baixo teor de matéria seca e concentração de clorofila. Estas espécies são consideradas aquisitivas. Enquanto espécies com baixa área foliar específica e concentração de nutrientes, alto teor de matéria seca e concentração de clorofila são consideradas de alto investimento na construção dos tecidos, e denominadas espécies conservativas [67, 68]. Entre estes dois grupos existe um contínuo de espécies que apresentam tanto características de alto como de baixo investimento (espécies intermediárias).

Os diferentes níveis de investimento influenciam para que as espécies apresentem aptidão diferenciada dependendo do ambiente em que se encontrem, direcionando a estruturação da vegetação. Ao longo da sucessão, por exemplo, nos estágios iniciais, os filtros ambientais abióticos (luz, por exemplo) selecionam espécies aquisitivas fazendo com que sejam dominantes nessas áreas, enquanto que com o avanço da sucessão os filtros bióticos (competição, por exemplo) atuam favorecendo a dominância de espécies conservativas [69, 70].

Em Silva *et al* [66] foram estudados seis fragmentos de florestas secundárias e três de florestas maduras, no domínio de Floresta Ombrófila Densa de Terras Baixas, localizados na Usina São José (USJ), no município de Igarassu, Pernambuco, Nordeste do Brasil. Foram medidas características funcionais (altura máxima da planta, área foliar específica, teor de

matéria seca da folha (mg), concentrações de nitrogênio (N), fósforo (P) e clorofila foliar, densidade da madeira e forma da semente) de 64 espécies. Estas foram agrupadas em três classes (aquisitivas, intermediárias e conservativas) através de uma análise de componentes principais - PCA, considerando todas as características funcionais amostradas.

Neste trabalho foram consideradas apenas características foliar, as seguintes: clorofila foliar, área foliar específica ( $cm^2$ ), teor de matéria seca da folha (mg) e teor de matéria úmida (g). Para cada espécie foram coletadas folhas de cinco indivíduos (10 folhas por indivíduo), resultando em 50 exemplares por espécie. Uma descrição sintética destes dados é apresentada na Tabela 1.

Tabela 1: Descrição sintética do conjunto de dados de espécies da Mata Atlântica pernambucana.

Dados	Cardinalidade	Dimensionalidade	Classes
Mata Atlântica	3200	4	3
Classes	Cardinalidade	Espécies	
Aquisitivas	700	14	
Conservativas	950	19	
Intermediárias	1550	31	

Este conjunto de dados é de grande interesse por apresentar uma característica que dificulta na tarefa de classificação, a sobreposição, como pode ser observado na Figura 8.

#### 4.6.2 Conjuntos de dados usuais na literatura

A descrição dos demais conjuntos de dados usados para testar a metodologia proposta é apresentada na Tabela 2, onde são mostrados a cardinalidade (número de instâncias), a dimensionalidade  $d$  (número de atributos), o número de classes e o número de instâncias em cada classe.

Os conjuntos de dados Diabetes (*Pima Indians Diabetes*), Câncer de mama, Iris, Coluna vertebral, Sementes, Vinho e *Yeast* foram obtidos de UCI *Machine Learning Repository* [71].

Enquanto os dados de Círculos concêntricos e Espirais concêntricos foram gerados artificialmente. Para ambos os conjuntos de dados existem 750 observações, onde cada observação é um vetor 2-dimensional, dividido em duas classes. Para os dados de Círculos concêntricos na primeira classe existem 375 observações distribuídas aleatoriamente em um raio  $0 \leq r \leq 0,5$ , o Círculo interno. Na segunda classe, as 375 observações foram gerados aleatoriamente em um raio  $1,0 \leq r \leq 1,5$ , o Círculo externo. As duas classes podem ser visualizadas na Figura 9(a). E para o conjunto de dados de Espirais concêntricos foi adotado a espiral Arquimediana

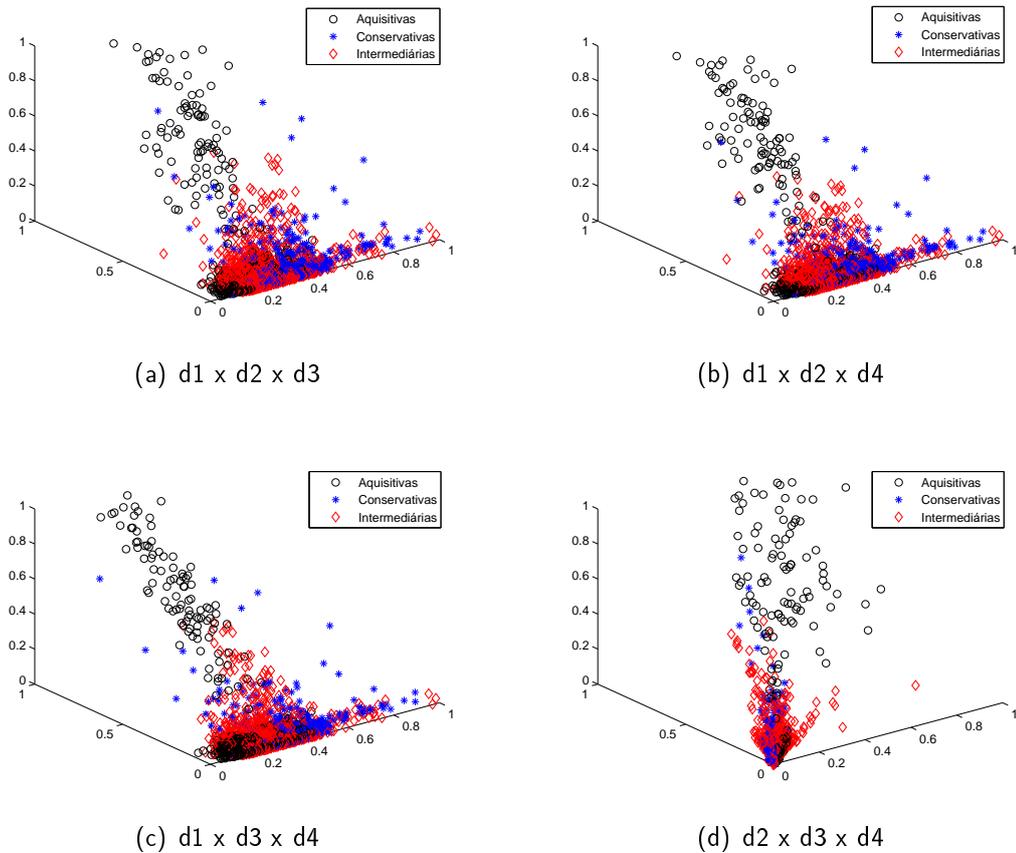


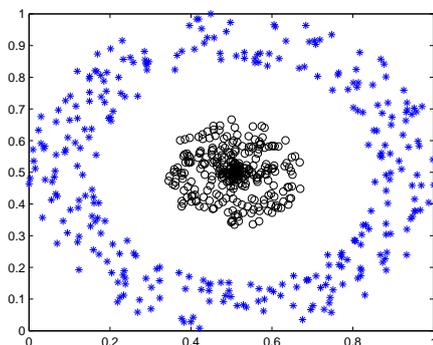
Figura 8: Projeções no  $\mathbb{R}^3$  do conjunto de dados de Mata Atlântica pernambucana, obtidas a partir da combinação dos quatro atributos que o descrevem (d1 = clorofila foliar, d2 = área foliar específica, d3 = teor de matéria seca da folha, d4 = teor de matéria úmida).

em que as coordenadas polares  $(r, \theta)$  podem ser descritas pela equação  $r = a + b\theta$ , com  $a, b \in \mathbb{R}$ . Onde  $b = 0,5$  para a Espiral interna e  $b = -0,5$  para o Espiral externa e  $a = 0$  para ambos os casos. Na Figura 9(b) as duas classes podem ser visualizadas. Para procedimentos de classificação baseados na distância entre um ponto qualquer e o centróide da classe, esses conjuntos de dados podem gerar uma classificação ambígua.

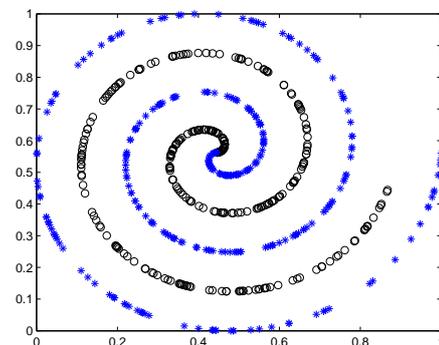
O conjunto *Yeast* (site de localização de proteínas), originalmente está dividido em dez classes, a fim de testar a metodologia para conjuntos com quatro classes, os dados foram organizados da seguinte forma: citoesqueleto (463 exemplos), nuclear (429), mitocondrial (244) e demais classes (proteína de membrana, extracelular, vacuolar, peroxisomal e lúmen do retículo endoplasmático) geraram uma classe com 348 exemplos. Oito atributos descrevem cada exemplo nesse conjunto de dados.

Tabela 2: Descrição sintética dos conjuntos de dados usados.

Dados	Instâncias	Atributos	Nº Classes	Classes	Instâncias
Círculos concêntricos	750	2	2	Círculo interno	375
				Círculo externo	375
Espirais concêntricas	750	2	2	Espiral interna	375
				Espiral externa	375
Diabetes	768	8	2	Positivo	268
				Não positivo	500
Câncer de mama	458	9	2	Benigno	458
				Maligno	241
Iris	150	4	3	Setosa	50
				Versicolor	50
				Virgínica	50
Coluna vertebral	310	6	3	Hérnia	60
				Espondilolistese	150
				Normal	100
Sementes	210	7	3	Kama	70
				Rosa	70
				Canadian	70
Vinho	188	13	3	Cultivar 1	59
				Cultivar 2	71
				Cultivar 3	58
Yeast	1484	8	4	Citoesqueleto	463
				Nuclear	429
				Mitocondrial	244
				Outras classes	348



(a) Círculos concêntricos



(b) Espirais concêntricas

Figura 9: Representação no espaço real, (a) 600 pontos distribuídos em dois Círculos concêntricos e (b) 600 pontos distribuídos em dois Espirais concêntricas.

## 5 Resultados Experimentais

Métodos de classificação são aplicados a conjuntos de dados usuais (originais) e posteriormente a conjuntos de dados transformados pela metodologia proposta. Os resultados são comparados para observância do desempenho nas duas condições: originais e transformados.

### 5.1 Experimentos realizados

Para cada um dos dez conjuntos de dados, Círculos concêntricos, Espirais concêntricas, Diabetes, Câncer de mama, Iris, Mata Atlântica, Coluna vertebral, Sementes, Vinho e *Yeast*, foi aplicada a metodologia proposta. A princípio, cada conjunto foi dividido em cinco subconjuntos de mesmo tamanho e mutuamente exclusivos, para o processo de validação. Para este processo foi utilizado o *k-fold cross validation*, com  $k = 5$ , de modo que 80% dos dados são usados para treino e 20% para teste (validação) em cada uma das iterações. Foram realizadas duas simulações em cada conjunto de dados, resultando em dez simulações.

Três diferentes algoritmos de classificação foram empregados aos dados reais (dados brutos) e dados transformados com o AG (espaço complexo). Os algoritmos de classificação escolhidos foram *K-Nearest Neighbor* (kNN) [9], *K-means* [8] e Análise Discriminante Linear (LDA) [10], que são frequentemente usados para classificação e agrupamento de dados.

Os resultados de ambas condições, real e transformado com AG, são comparados para verificar se há uma melhor taxa de reconhecimento na classificação quando os dados têm sido transformados. Ou seja, se com a transformação proposta diminui-se a sobreposição dos dados, obtendo-se um aprimoramento em sua discriminação.

Para mapear um novo ponto no espaço de representação gerado pela transformação proposta, é treinada uma RNA *multilayer perceptron*. Com o intuito de diminuir o custo deste processo, é realizada uma redução no conjunto de dados gerado pelo AG. Com tal finalidade, o método PCA é aplicado aos dados transformados, onde a ideia é usar a mesma dimensionalidade dos dados originais, mas no espaço complexo.

Aos dados gerados a partir da técnica PCA, são aplicados os métodos de classificação e observado se o desempenho é similar ao dos dados com dimensão incrementada. O objetivo é avaliar se de fato foram mantidas as características necessárias para a discriminação dos dados. E a seguir, este novo conjunto de dados será tratado como o conjunto de dados de saída no treinamento da RNA.

Finalmente, com a RNA treinada, os dados preservados para teste são integrados, gerando um novo conjunto de dados. A este novo conjunto de dados e aos dados originais preservados são aplicados os métodos de classificação, e de acordo com o desempenho é possível dizer se o mapeamento realizado é eficiente para fornecer uma boa classificação a qualquer novo dado que venha a ser integrado ao processo proposto.

A forma como os conjuntos de dados são particionados e a aplicação da metodologia proposta é ilustrada na Figura 10.

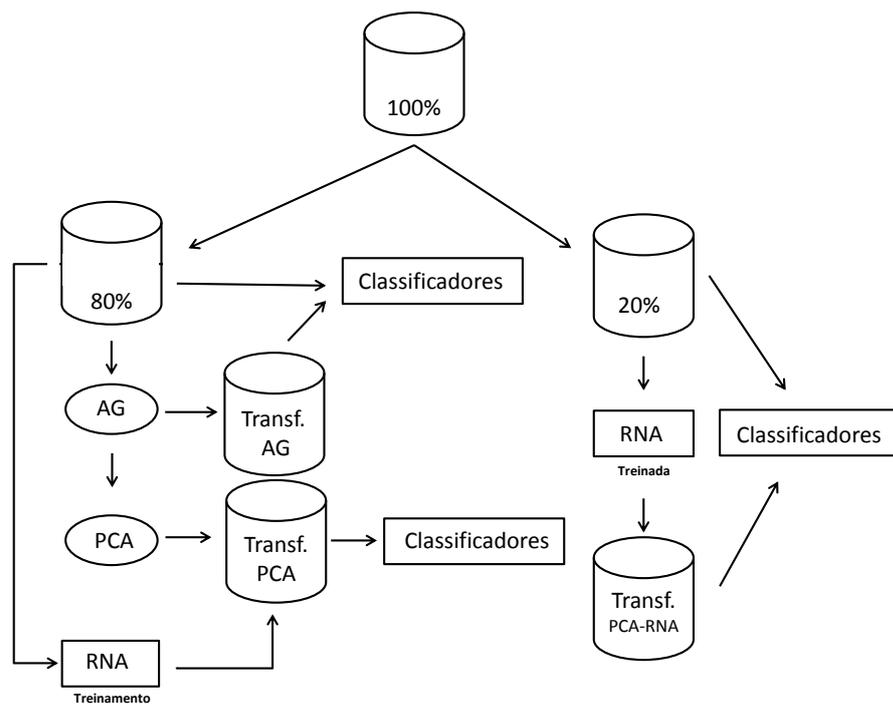


Figura 10: Ilustração do particionamento dos conjuntos de dados e aplicação da metodologia proposta.

## 5.2 Taxa de reconhecimento da transformada inversa - AG

O primeiro método empregado para a tarefa de classificação foi o algoritmo kNN [1, 9]. Variações desse método apresentando resultados relevantes são encontradas na literatura. Como

por exemplo, para o conjunto de dados de Câncer de mama, Sarkar e Leong [72] propuseram uma variação do kNN denominado *fuzzy* kNN (FkNN). Foram usados metade dos dados para treinamento e a outra metade para teste e observados valores de  $k$  entre um e quinze. Os resultados apresentaram uma taxa de reconhecimento de classificação próxima a 99%.

Outro resultado significativo para estes dados é encontrado quando o método EAdaBoost é introduzido por Gao e Gao [14]. Um AdaBoost [15] adaptado pelo kNN ponderado (EAdaBoost) é projetado, onde AdaBoost e kNN naturalmente se complementam. O conjunto de treinamento foi formado por 2/3 dos dados e o conjunto de teste pelos dados restantes. O resultado é apresentado por uma taxa de reconhecimento em relação ao erro médio de classificação de 0,03562. Um outro conjunto de dados classificado em [14] é o de Diabetes, este apresenta uma taxa de reconhecimento de 0,21484 em relação ao erro.

Para estes conjuntos de dados, outros resultados relevantes são obtidos com o método AC-kNN proposto por Huang *et al* [16]. Este método utiliza um algoritmo kNN melhorado para resolver os conflitos de regras, denominado Classificação Associativa com kNN (AC-kNN). AC-kNN gera para cada instância de teste de um conjunto de treinamento específico composto de instâncias cobertas pelas melhores  $n$  regras que correspondem à instância de teste. Assim, os vizinhos mais próximos do conjunto de treinamento específico não são apenas similares, são também associativos com a instância de teste. Para os dados de Câncer de mama, uma taxa de reconhecimento de 95,2%, para os dados de Diabetes o AC-kNN obteve 73,4% de taxa de reconhecimento e para os dados de Vinho 96,5%.

Embora muito relevantes, os resultados citados são inferiores em taxa de reconhecimento quando comparados aos resultados obtidos com o kNN após o método de discriminação de dados ser aplicado, como pode ser observado a seguir. Cada um dos dez conjuntos de dados foi particionado em dois subconjuntos: um para treinamento com 80% dos dados e um para teste com os 20% restantes. E para cada conjunto de dados foram considerados  $1 \leq k \leq \sqrt{N}$  vizinhos, em que  $N$  é o tamanho do conjunto de dados de treino. Para os dados brutos, uma vez que para cada conjunto de dados existe certa semelhança entre os resultados das dez simulações, é escolhido um resultado para ilustrar (Figura 11) a assertividade do método kNN para os valores admitidos de  $k$ .

Para os dados brutos dos Círculos concêntricos, o kNN forneceu 100% de classificação correta com o parâmetro  $k \leq 21$ , o mesmo ocorrendo para os dados transformados. Assim, embora não exista diferença na taxa de reconhecimento, existe uma considerável redução de custo na resolução do problema. Uma vez que o problema original é não linearmente separável e depois da transformação proposta o problema torna-se, com alta probabilidade, linearmente

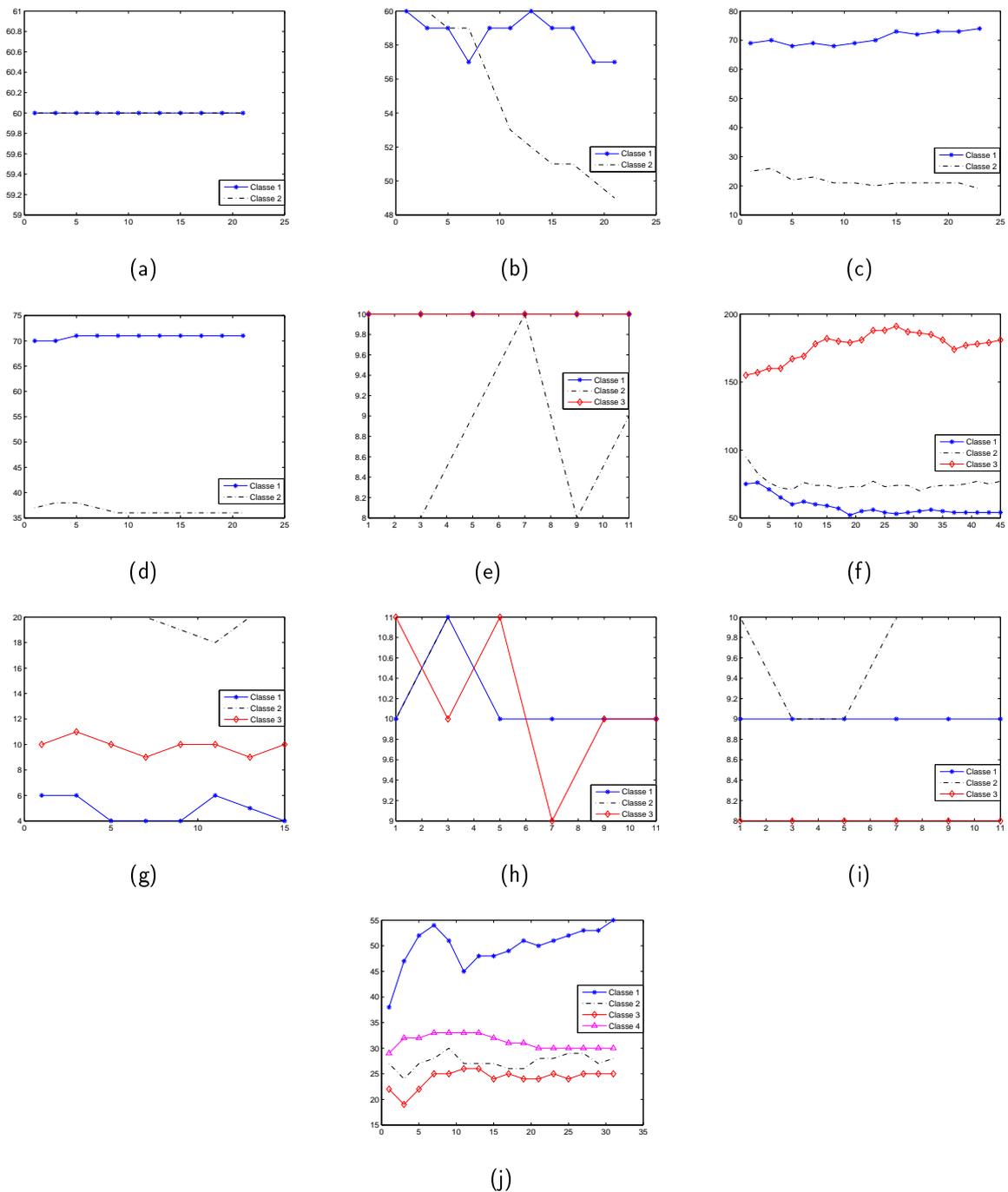


Figura 11: Assertividade do método kNN para valores de  $k$  ( $1 \leq k \leq \sqrt{N}$ ) para os dados originais de: (a) Círculos concêntricos, (b) Espirais concêntricas, (c) Diabetes, (d) Câncer de mama, (e) Iris, (f) Mata Atlântica, (g) Coluna vertebral, (h) Sementes, (i) Vinho e (j) Yeast.

separável, como é sugerido pela Figura 12(a).

O algoritmo kNN difere em desempenho de classificação para diferentes valores do vizinho  $k$  no conjunto de dados brutos de Espirais concêntricas. Para  $1 \leq k \leq 5$  apresenta taxa de acertos aproximadamente constantes, para os valores seguintes  $k \leq 21$  a classe 2 apresenta um decaimento em sua taxa de acertos, enquanto a classe 1 mantém-se aproximadamente

constante. Enquanto para os dados transformados (Figura 12(b)) o algoritmo kNN também é hábil em gerar uma taxa de classificação atingindo 100% de assertividade para ambas as classes e todos os valores testados de  $k$ . Na Tabela 3 é apresentado a média de acertos para  $1 \leq k \leq 21$  com respectivo desvio padrão.

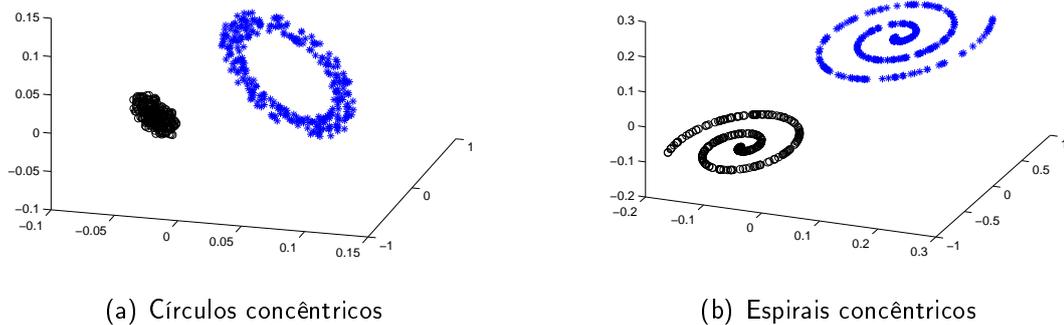


Figura 12: Projeção sobre o  $\mathbb{R}^3$  do espaço complexo obtido com a aplicação da metodologia, (a) 600 pontos distribuídos em dois Círculos concêntricos e (b) 600 pontos distribuídos em dois Espirais concêntricas.

Nos dados brutos de Diabetes são obtidos em média 87% de acertos para a classe 1 e 51% para a classe 2. Pode ser observado (Figura 11 (c)) que não existe um comportamento definido de crescimento ou decrescimento nas taxas de acertos à medida que o valor de  $k$ ,  $1 \leq k \leq 23$  aumenta. Embora exista essa inconstância, os valores giram em torno da média. No entanto, quando aplicada a transformação aos dados, uma taxa de classificação correta atingindo 100% é obtida para ambas as classes.

Os dados de Câncer de mama exibem em média 99% de acertos para a classe 1 e 93% para a classe 2. Nos dados brutos a taxa de acerto da classe 2 decresce lentamente quando o valor de  $1 \leq k \leq 21$  é incrementado. Enquanto os dados transformados fornecem classificação com 100% de acerto para ambas as classes.

Para os dados brutos de Iris o kNN fornece a mesma taxa de reconhecimento na classificação para  $1 \leq k \leq 11$  para as classes 1 e 2, com variações na classificação correta para a classe 3. Conseqüentemente para os dados brutos de Iris, o algoritmo kNN também não é hábil em gerar uma taxa de reconhecimento atingindo 100%, uma vez que as classes dos dados de Iris são sobrepostas. Mas para os dados transformados o kNN alcançou uma taxa de reconhecimento de 100% para o conjunto de teste.

Nos dados originais de Mata Atlântica com  $1 \leq k \leq 45$ , o kNN forneceu maior taxa de acertos para a classe 3, onde à medida que o valor de  $k$  é incrementado, aumenta o número

de vizinhos classificados como pertencentes a esta classe. Enquanto para as demais classes o percentual de acertos é 52% e 53%, o que de certo modo é esperado, devido à sobreposição dos dados. No entanto para os dados transformados, o percentual de acertos atinge 100% para todas as classes.

Similar aos dados de Diabetes, que não tem comportamento definido de crescimento ou decréscimo enquanto o parâmetro  $k$  é incrementado, as classes 1 e 3 dos dados brutos de Coluna vertebral apresentam taxas de acerto de 56% e 61%, respectivamente, enquanto a classe 2 apresenta comportamento aproximadamente constante em torno do número médio de acertos e uma taxa de 88% de classificação correta. E mais uma vez, para os dados transformados a classificação correta de todas as classes atinge 100%.

O algoritmo kNN foi aplicado aos dados brutos de Sementes, exibindo aproximadamente a mesma taxa de reconhecimento na classificação para as classes 2 e 3 para todo valor de  $1 \leq k \leq 11$  observado e variações para a classe 1. Enquanto para os dados de Sementes transformados o kNN forneceu uma taxa de reconhecimento de 100% para todas as classes.

E para os dados brutos de Vinho o kNN, com  $1 \leq k \leq 11$ , forneceu uma taxa de 100% para as classes 1 e 3 para todo  $k$ , enquanto a classe 2 apresentou variações no número de acertos à medida que  $k$  foi incrementado, sem um comportamento definido de crescimento ou decréscimo, como pode ser observado na Figura 11(i) e na Tabela 3 a partir dos valores do desvio padrão. No entanto para os dados transformados uma taxa de 100% de classificação correta ocorre para todas as classes.

Por fim, para os dados brutos de *Yeast* à medida que  $k$  ( $1 \leq k \leq 31$ ) foi incrementado, foram observados os seguintes comportamentos nas taxas de classificação correta: a classe 1 cresceu pouco e inconstantemente, a classe 2 mostrou-se inconstante em torno da média, a classe 3 apresentou um leve crescimento e a classe 4 exibiu um leve crescimento seguido por um leve decaimento com uma aproximação de um valor constante. As quatro classes apresentam taxa de classificação correta constante em 100% quando os dados tem sido transformados.

Os resultados da classificação com o algoritmo kNN: média e desvio para as cinco iterações (dez simulações) realizadas para todos os conjuntos de dados estudados, são mostrados na Tabela 3. Observando que para cada conjunto de dados não foi considerado apenas um valor de  $k$ , mas a média dos resultados para  $1 \leq k \leq \sqrt{N}$ . Esse resultado também é ilustrado na Figura 13.

O segundo método usado para classificar os dados foi o algoritmo *K-means*. Muitas variações dele são encontradas na literatura com resultados bem relevantes [11–13, 73] alguns, semelhantes a metodologia proposta neste trabalho, aumenta a dimensionalidade dos

Tabela 3: Taxa de reconhecimento para o método kNN para os dados brutos e dados transformados (AG) de Círculos concêntricos, Espirais concêntricos, Diabetes, Câncer de mama, Iris, Mata Atlântica, Coluna vertebral, Sementes, Vinho e Yeast.

Dados	k máximo	Classe	Original (%)	Transf. (AG) (%)
Círculos concêntricos	21	Classe 1	100,00 (0,00)	100,00 (0,00)
		Classe 2	100,00 (0,00)	100,00 (0,00)
Espirais concêntricos	21	Classe 1	99,00 (1,33)	100,00 (0,00)
		Classe 2	90,80 (1,03)	100,00 (0,00)
Diabetes	23	Classe 1	87,20 (2,25)	100,00 (0,00)
		Classe 2	51,00 (1,33)	100,00 (0,00)
Câncer de mama	21	Classe 1	99,20 (1,69)	100,00 (0,00)
		Classe 2	93,40 (3,37)	100,00 (0,00)
Iris	11	Classe 1	100,00 (0,00)	100,00 (0,00)
		Classe 2	88,00 (8,34)	100,00 (0,00)
		Classe 3	95,20 (6,20)	100,00 (0,00)
Mata Atlântica	45	Classe 1	52,40 (2,07)	100,00 (0,00)
		Classe 2	53,40 (2,37)	100,00 (0,00)
		Classe 3	71,00 (1,76)	100,00 (0,00)
Coluna vertebral	15	Classe 1	56,00 (10,75)	100,00 (0,00)
		Classe 2	87,60 (4,25)	100,00 (0,00)
		Classe 3	61,40 (8,81)	100,00 (0,00)
Sementes	11	Classe 1	91,00 (0,00)	100,00 (0,00)
		Classe 2	98,20 (3,79)	100,00 (0,00)
		Classe 3	94,60 (4,65)	100,00 (0,00)
Vinho	11	Classe 1	100,00 (0,00)	100,00 (0,00)
		Classe 2	87,40 (4,65)	100,00 (0,00)
		Classe 3	100,00 (0,00)	100,00 (0,00)
Yeast	31	Classe 1	66,40 (3,81)	100,00 (0,00)
		Classe 2	39,40 (4,74)	100,00 (0,00)
		Classe 3	61,20 (2,86)	100,00 (0,00)
		Classe 4	57,00 (7,15)	100,00 (0,00)

dados [74–77]. Entretanto, com a utilização do processo proposto para discriminar os dados antes de aplicar o método de classificação, obtém-se melhores resultados do que aqueles encontrados na literatura.

Diversas variações do método *K-means* fornecem resultados de classificação com boa

taxa de reconhecimento como, por exemplo, Huang *et al* [13] propuseram um subespaço discriminativo *Kmeans* (*DSKmeans*), que integra a compacticidade intra classe e a separação inter classe simultaneamente, obtendo uma taxa de reconhecimento de 90,73% para os dados de Iris. Com uma variação do *K-means* usando função de kernel, Sarma *et al* [77] obtiveram uma taxa de 89,76% também para dados de Iris. Enquanto Zhang e Fang [12] analisaram as vantagens e desvantagens do *K-means* tradicional e elaboraram um método de fornecer um *K-means* baseado no ponto inicial, além de determinar o valor de *k*. Foram obtidos com isso uma taxa de reconhecimento de 88,51% para dados de Iris e 74,23% para dados de Vinho. Também para dados de Vinho, Prabhu e Anbazhagan [11] propuseram um método para encontrar os centróides iniciais para tornar o *K-means* mais eficaz e eficiente, com isso obtiveram uma taxa de reconhecimento de 92,13%.

Primeiramente o método *K-means* foi aplicado aos dados originais. Para os dados brutos de Círculos concêntricos e Espirais concêntricos, a classificação com o *K-means* forneceu ambiguidades, como era esperado, uma vez que ambos conjuntos tem duas classes com centróides localizados no mesmo local no espaço de representação real. Além disso, conjuntos de dados que apresentam sobreposição ou não separabilidade de classes são hábeis na geração de resultados incoerentes, por tratar de um método de classificação baseado em distâncias. Isso pode ser observado, por exemplo, nos conjuntos de dados de Iris, Sementes, *Yeast* que geram resultados incorretos na classificação. Assim como também evidencia-se fortemente nos dados de Mata Atlântica, resultando em baixa taxa de reconhecimento na classificação.

No intuito de tornar o conjunto de dados separáveis, a metodologia proposta foi aplicada a cada um dos conjuntos de dados observados. A seguir, o algoritmo *K-means* foi aplicado aos dados transformados de Círculos concêntricos, Espirais concêntricos, Diabetes, Câncer de mama, Iris, Mata Atlântica, Coluna vertebral, Sementes, Vinho e *Yeast*, obtendo uma taxa de reconhecimento de 100%, exceto para os dados de Câncer de mama, em que a classe 2 apresentou uma classificação média de 98% de acertos, sendo ainda melhor resultado do que os dados brutos. Assim, foi possível observar um ganho na habilidade de discriminar dados e uma possível separação das classes dos conjuntos de dados.

Uma vez que, para cada conjunto de dados os resultados apresentam certa semelhança para todas as simulações realizadas, um resultado de uma das dez simulações foi escolhido para construção das matrizes de confusão. Numa matriz de confusão os detalhes relacionados a números de acertos e erros são visualizados com mais facilidade. No entanto, a Tabela 4 resume os resultados experimentais obtidos pelo algoritmo *K-means*, média e desvio padrão para as dez simulações realizadas, para todos os conjuntos de dados nas condições: dados brutos e dados transformados. Além disso, esse resultado também é apresentado na Figura 13.

Matrizes de confusão para o método *K-means* - Dados originais e dados transformados.

Círculos concêntricos

		Dados Originais		Dados transformados	
		Preditada		Preditada	
		C1	C2	C1	C2
Real	C1	79% (236)	21% (64)	100% (300)	0%
	C2	55% (166)	45% (134)	0%	100% (300)

Espirais concêntricas

		Dados originais		Dados transformados	
		Preditada		Preditada	
		C1	C2	C1	C2
Real	C1	60% (180)	40% (120)	100% (300)	0%
	C2	61% (183)	39% (117)	0%	100% (300)

Diabetes

		Dados originais		Dados transformados	
		Preditada		Preditada	
		C1	C2	C1	C2
Real	C1	76% (304)	24% (96)	100% (400)	0%
	C2	52% (112)	48% (102)	0%	100% (214)

Câncer de mama

		Dados originais		Dados transformados	
		Preditada		Preditada	
		C1	C2	C1	C2
Real	C1	98% (349)	2% (6)	100% (355)	0%
	C2	8% (15)	92% (176)	2% (3)	98% (188)

## Iris

Dados originais

Classes		Predita		
		C1	C2	C3
Real	C1	100% (40)	0%	0%
	C2	0%	95% (38)	5% (2)
	C3	0%	25% (10)	75% (30)

Dados transformados

Classes		Predita		
		C1	C2	C3
Real	C1	100% (40)	0%	0%
	C2	0%	100% (40)	0%
	C3	0%	0%	100% (40)

## Mata Atlântica

Dados originais

Classes		Predita		
		C1	C2	C3
Real	C1	12% (67)	25% (138)	63% (355)
	C2	1% (7)	72% (551)	27% (202)
	C3	1% (15)	49% (610)	50% (615)

Dados transformados

Classes		Predita		
		C1	C2	C3
Real	C1	100% (560)	0%	0%
	C2	0%	100% (760)	0%
	C3	0%	0%	100% (1240)

## Coluna vertebral

Dados originais

Classes		Predita		
		C1	C2	C3
Real	C1	85% (41)	2% (1)	13% (6)
	C2	8% (9)	49% (59)	43% (52)
	C3	17% (14)	4% (3)	79% (63)

Dados transformados

Classes		Predita		
		C1	C2	C3
Real	C1	100% (48)	0%	0%
	C2	0%	100% (120)	0%
	C3	0%	0%	100% (80)

## Sementes

Classes		Predita		
		C1	C2	C3
Real	C1	84% (47)	0%	16% (9)
	C2	11% (6)	89% (50)	0%
	C3	5% (3)	0%	95% (53)

Classes		Predita		
		C1	C2	C3
Real	C1	100% (56)	0%	0%
	C2	0%	100% (56)	0%
	C3	0%	0%	100% (56)

## Vinho

Classes		Predita		
		C1	C2	C3
Real	C1	100% (47)	0%	0%
	C2	0%	89% (51)	11% (6)
	C3	0%	0%	100% (38)

Classes		Predita		
		C1	C2	C3
Real	C1	100% (47)	0%	0%
	C2	0%	100% (57)	0%
	C3	0%	0%	100% (38)

## Yeast

Classes		Predita			
		C1	C2	C3	C4
Real	C1	57% (210)	24% (90)	14% (51)	5% (19)
	C2	42% (145)	47% (161)	9% (29)	2% (8)
	C3	25% (48)	6% (12)	59% (115)	10% (20)
	C4	33% (92)	3% (7)	23% (65)	41% (114)

Classes		Predita			
		C1	C2	C3	C4
Real	C1	100% (370)	0%	0%	0%
	C2	0%	100% (343)	0%	0%
	C3	0%	0%	100% (195)	0%
	C4	0%	0%	0%	100% (278)

O último algoritmo empregado para classificação dos conjuntos de dados foi o LDA [1, 10]. Como os dez conjuntos de dados brutos são não linearmente separáveis, o LDA não é hábil em classificar corretamente todos os dados. No entanto, para os conjuntos de dados transformados, o LDA forneceu uma taxa de reconhecimento de 100% para quase todos os

conjuntos de dados. Os dados de Câncer de mama apresentam 99% de taxa de acerto para a classe 2. Mas em geral, temos melhor desempenho do que outros métodos na literatura [17, 78]. Mais uma vez, estes resultados experimentais implicam que, com alta probabilidade, a transformação proposta é capaz de diminuir a complexidade dos conjuntos de dados e aprimorar o poder em discriminá-los, de modo que um problema não linearmente separável pode tornar-se em um problema linearmente separável, com alta probabilidade.

A Tabela 4 resume os resultados do algoritmo LDA para os conjuntos de dados brutos e transformados. Bem como, esse resultado também é ilustrado na Figura 13. Além disso, para cada conjunto de dados foram geradas matrizes de confusão para os resultados obtidos de dados reais e transformados. Como em cada conjunto de dados os resultados apresentam certa semelhança para todas as simulações realizadas, um resultado de uma das dez simulações foi escolhido para construção das matrizes de confusão.

Matrizes de confusão para o método LDA - Dados originais e dados transformados.

Círculos concêntricos

		Dados Originais		Dados transformados	
		Preditada		Preditada	
Classes		C1	C2	C1	C2
Real	C1	59% (177)	41% (123)	100% (300)	0%
	C2	50% (149)	50% (151)	0%	100% (300)

Espirais concêntricas

		Dados originais		Dados transformados	
		Preditada		Preditada	
Classes		C1	C2	C1	C2
Real	C1	57% (170)	43% (130)	100% (300)	0%
	C2	51% (154)	49% (146)	0%	100% (300)

## Diabetes

Classes		Predita	
		C1	C2
Real	C1	80% (321)	20% (79)
	C2	27% (57)	73% (157)

Classes		Predita	
		C1	C2
Real	C1	100% (400)	0%
	C2	0%	100% (214)

## Câncer de mama

Classes		Predita	
		C1	C2
Real	C1	98% (349)	2% (6)
	C2	7% (14)	93% (177)

Classes		Predita	
		C1	C2
Real	C1	100% (355)	0%
	C2	1% (2)	99% (189)

## Iris

Classes		Predita		
		C1	C2	C3
Real	C1	100% (40)	0%	0%
	C2	0%	97% (39)	3% (1)
	C3	0%	3% (1)	97% (39)

Classes		Predita		
		C1	C2	C3
Real	C1	100% (40)	0%	0%
	C2	0%	100% (40)	0%
	C3	0%	0%	100% (40)

## Mata Atlântica

Classes		Predita		
		C1	C2	C3
Real	C1	55% (306)	13% (72)	32% (182)
	C2	3% (22)	64% (486)	33% (252)
	C3	20% (243)	35% (433)	45% (564)

Classes		Predita		
		C1	C2	C3
Real	C1	100% (560)	0%	0%
	C2	0%	100% (760)	0%
	C3	0%	0%	100% (1240)

## Coluna vertebral

Dados originais

Classes		Predita		
		C1	C2	C3
Real	C1	83% (40)	0%	17% (8)
	C2	2% (2)	89% (107)	9% (11)
	C3	15% (12)	3% (2)	82% (66)

Dados transformados

Classes		Predita		
		C1	C2	C3
Real	C1	100% (48)	0%	0%
	C2	0%	100% (120)	0%
	C3	0%	0%	100% (80)

## Sementes

Dados originais

Classes		Predita		
		C1	C2	C3
Real	C1	96% (54)	0%	4% (2)
	C2	0%	100% (56)	0%
	C3	5% (3)	0%	95% (53)

Dados transformados

Classes		Predita		
		C1	C2	C3
Real	C1	100% (56)	0%	0%
	C2	0%	100% (56)	0%
	C3	0%	0%	100% (56)

## Vinho

Dados originais

Classes		Predita		
		C1	C2	C3
Real	C1	100% (47)	0%	0%
	C2	0%	100% (57)	0%
	C3	0%	0%	100% (38)

Dados transformados

Classes		Predita		
		C1	C2	C3
Real	C1	100% (47)	0%	0%
	C2	0%	100% (57)	0%
	C3	0%	0%	100% (38)

## Yeast

		Dados originais				Dados transformados						
Classes		Preditas				Classes		Preditas				
		C1	C2	C3	C4			C1	C2	C3	C4	
Real	C1	64%	16%	14%	6%	Real	C1	100%	0%	0%	0%	
		(237)	(61)	(51)	(21)			(370)				
	C2	(40%)	(43%)	(9%)	(8%)		C2	0%	100%	0%	0%	
		137	148	31	27			(343)				
C3	(22%)	(3%)	(61%)	(14%)	C3	0%	0%	100%	0%			
	43	5	119	28				(195)				
C4	(10%)	(4%)	(9%)	(77%)	C4	0%	0%	0%	100%			
	27	10	25	216					(278)			

### 5.3 Taxa de reconhecimento da transformada - PCA

O método PCA foi aplicado aos dados transformados pelo AG ( $2d$ -dimensional) gerando um novo conjunto com dimensão reduzida ( $d$ -dimensional), mas ainda no espaço de representação complexo. Uma vez que há taxa de reconhecimento de 100% para os conjuntos de dados com dimensão incrementada, os métodos de classificação foram empregados para verificar se as características mantidas eram suficientes para a tarefa de classificação. Na técnica PCA, a contribuição (energia) de cada componente principal é dada a partir dos autovalores relacionados aos autovetores que geram a componente. Assim, a energia é dada pela proporção entre cada autovalor e a soma de todos os autovalores. A taxa de reconhecimento e energia acumulada podem ser vistos na Tabela 5.

Pode-se observar na Tabela 5 que, à medida que o número de classes a ser distinguidas aumenta, a energia acumulada diminui. No entanto, as taxas de reconhecimento são máximas, sugerindo que a quantidade de informação levada é suficiente para bem realizar a tarefa de classificação. Além disso, é possível observar graficamente que a separabilidade dos dados se mantém, como por exemplo, na Figura 14 para os dados de Mata Atlântica após transformação com PCA. A partir disso, um novo conjunto de dados é gerado com dimensão reduzida, conseqüentemente favorecendo a um menor custo computacional no processo classificatório.

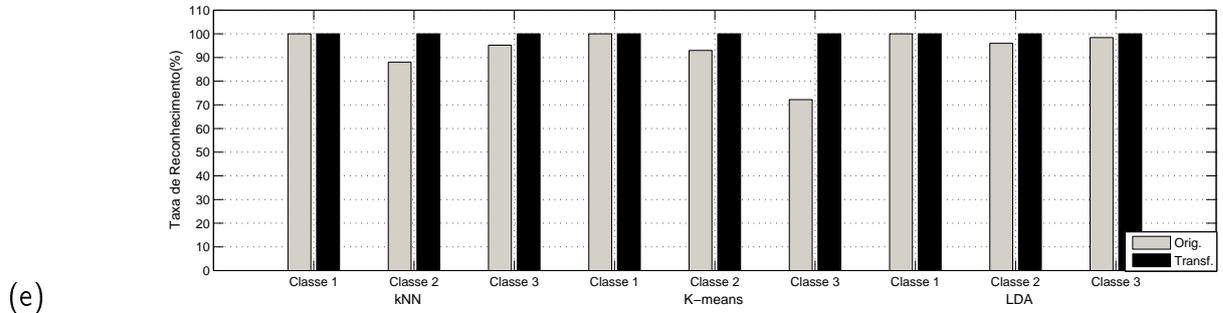
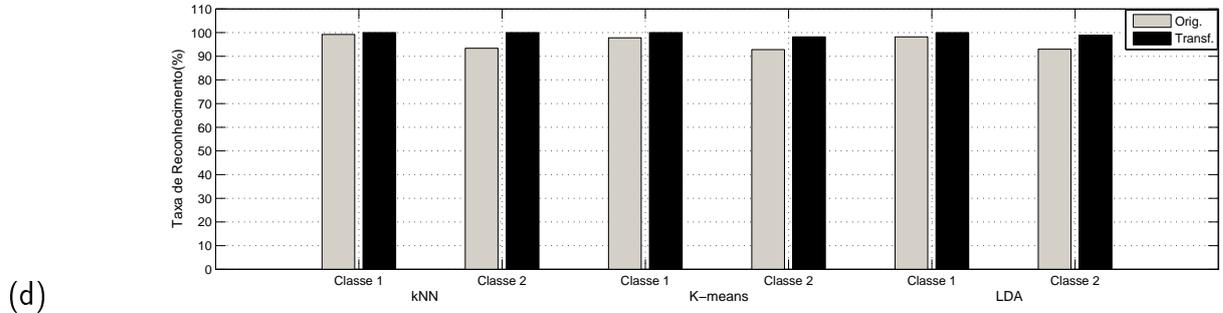
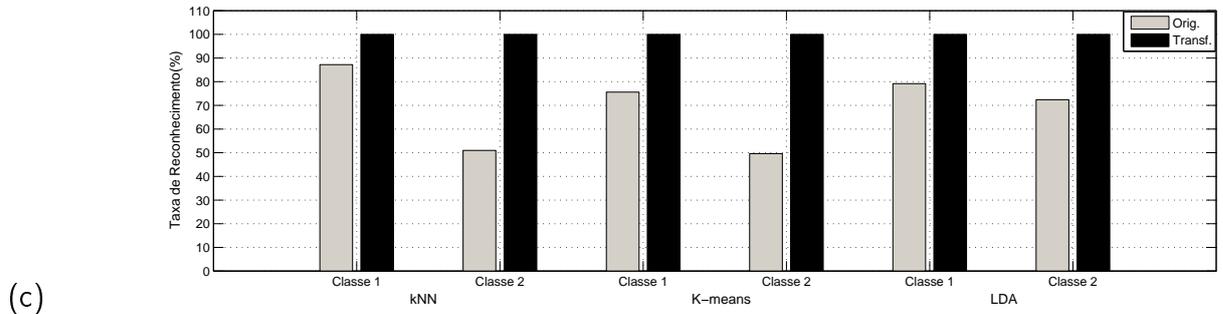
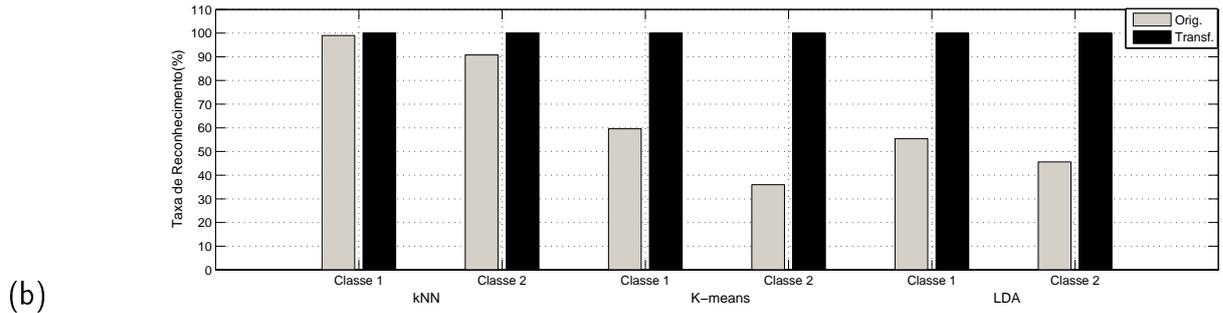
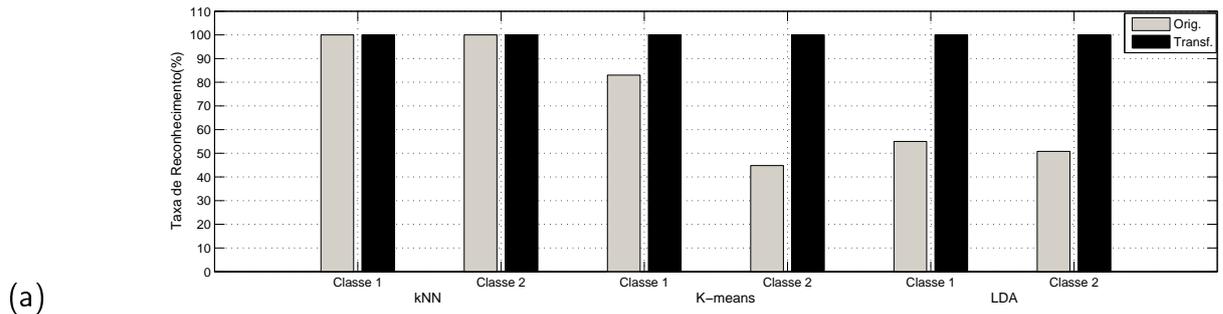
Na Figura 15, pode ser observado o comportamento dos dados após as transformações propostas para dados de Diabetes, Câncer de mama, Iris, Coluna vertebral, Sementes, Vinho e Yeast. Para gerar estas figuras foram usadas apenas as três primeiras dimensões de cada conjunto de dados.

Tabela 4: Taxa de reconhecimento para os métodos *K-means* e LDA para os dados brutos e dados transformados (AG) de Círculos concêntricos, Espirais concêntricas, Diabetes, Câncer de mama, Iris, Mata Atlântica, Coluna vertebral, Sementes, Vinho e *Yeast*.

Dados	Classe	<i>K-means</i>		LDA	
		Original (%)	Transf. (AG) (%)	Original (%)	Transf. (AG) (%)
Círculos concêntricos	Classe 1	83,00 (2,75)	100,00 (0,00)	55,00 (5,92)	100,00 (0,00)
	Classe 2	44,80 (0,79)	100,00 (0,00)	50,80 (0,79)	100,00 (0,00)
Espirais concêntricas	Classe 1	59,60 (0,52)	100,00 (0,00)	55,40 (1,07)	100,00 (0,00)
	Classe 2	36,00 (2,58)	100,00 (0,00)	45,60 (5,44)	100,00 (0,00)
Diabetes	Classe 1	75,60 (0,84)	100,00 (0,00)	79,20 (1,40)	100,00 (0,00)
	Classe 2	49,60 (2,72)	100,00 (0,00)	72,40 (1,26)	100,00 (0,00)
Câncer de mama	Classe 1	97,80 (0,79)	100,00 (0,00)	98,20 (0,42)	100,00 (0,00)
	Classe 2	92,80 (1,23)	98,20 (0,42)	93,00 (1,76)	98,90 (0,32)
Iris	Classe 1	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
	Classe 2	93,00 (1,93)	100,00 (0,00)	96,00 (1,55)	100,00 (0,00)
	Classe 3	72,20 (3,91)	100,00 (0,00)	98,00 (0,84)	100,00 (0,00)
Mata Atlântica	Classe 1	12,80 (0,42)	100,00 (0,00)	54,60 (1,43)	100,00 (0,00)
	Classe 2	72,20 (2,15)	100,00 (0,00)	63,00 (1,15)	100,00 (0,00)
	Classe 3	51,80 (1,55)	100,00 (0,00)	47,40 (1,96)	100,00 (0,00)
Coluna vertebral	Classe 1	85,20 (2,70)	100,00 (0,00)	81,00 (1,33)	100,00 (0,00)
	Classe 2	54,80 (5,18)	100,00 (0,00)	88,20 (2,86)	100,00 (0,00)
	Classe 3	84,20 (4,49)	100,00 (0,00)	76,20 (4,54)	100,00 (0,00)
Sementes	Classe 1	83,20 (1,03)	100,00 (0,00)	95,20 (0,42)	100,00 (0,00)
	Classe 2	88,00 (1,15)	100,00 (0,00)	99,60 (0,84)	100,00 (0,00)
	Classe 3	92,00 (5,29)	100,00 (0,00)	95,80 (1,23)	100,00 (0,00)
Vinho	Classe 1	99,60 (0,84)	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
	Classe 2	88,00 (1,15)	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
	Classe 3	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
<i>Yeast</i>	Classe 1	54,20 (2,04)	100,00 (0,00)	63,40 (2,55)	100,00 (0,00)
	Classe 2	47,40 (1,58)	100,00 (0,00)	45,40 (1,43)	100,00 (0,00)
	Classe 3	58,20 (1,69)	100,00 (0,00)	61,80 (1,23)	100,00 (0,00)
	Classe 4	40,60 (1,43)	100,00 (0,00)	79,00 (3,20)	100,00 (0,00)

## 5.4 Mapeando um novo ponto no espaço de representação complexo

Dado um novo ponto não envolvido no processo de discriminação dos dados, este ponto pode ser mapeado no espaço de representação obtido com a transformação que incrementa a dimensionalidade dos conjuntos de dados. Para tanto, arquiteturas de RNAs foram investigadas por meio do método tentativa e erro. Se for encontrada uma arquitetura conveniente e o mapeamento for realizado coerentemente, uma perfeita classificação é obtida para os dados



[Continua]

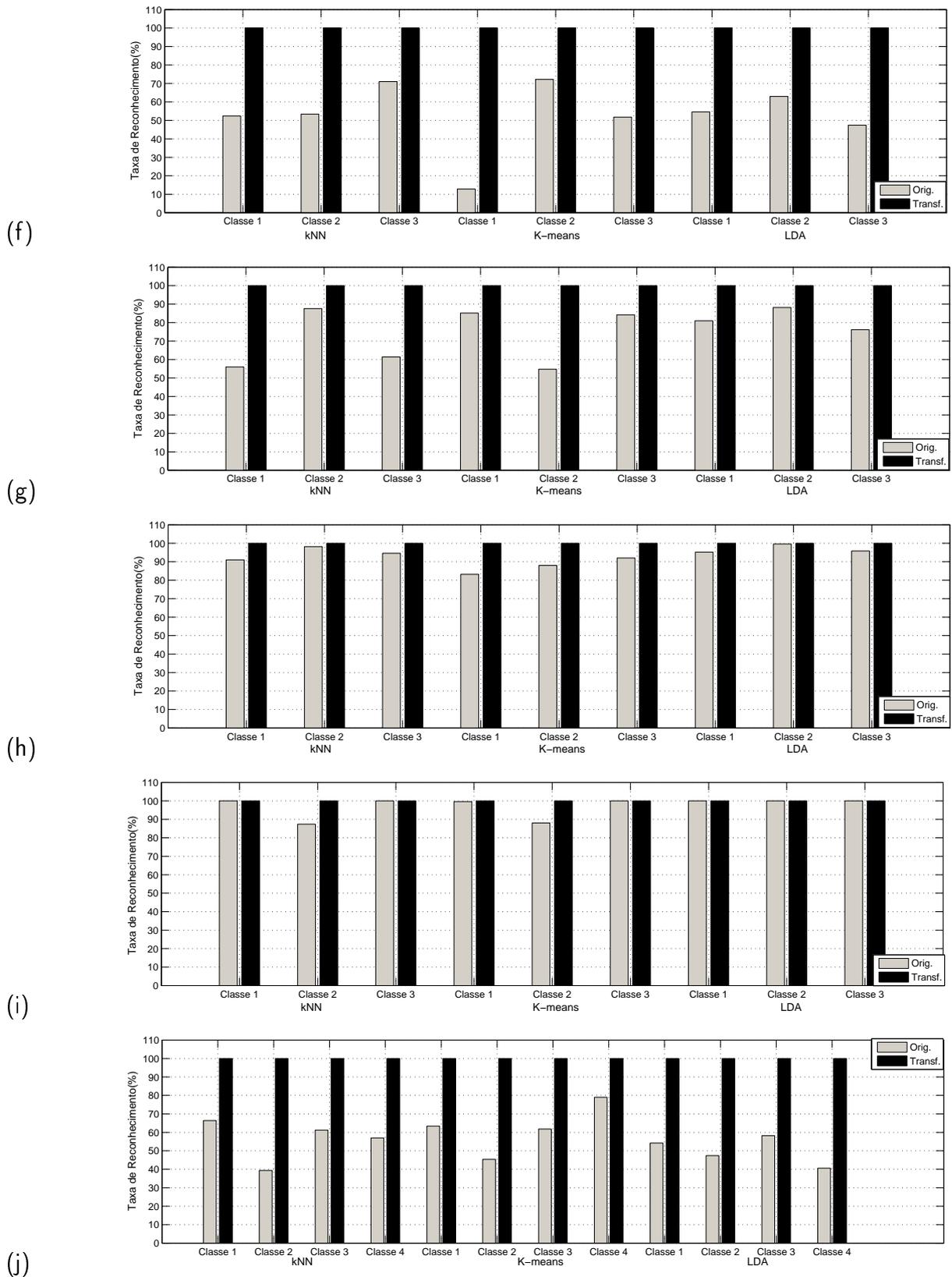
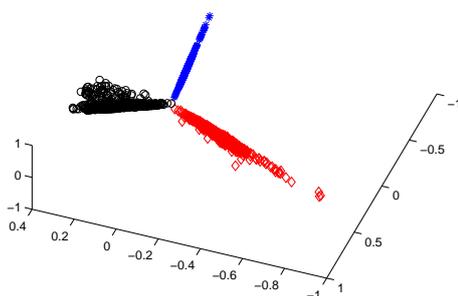


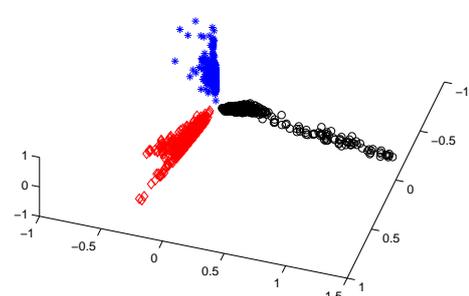
Figura 13: Taxa de reconhecimento para os métodos kNN, *K-means* e LDA para os dados brutos e dados transformados (AG) de: (a) Círculos concêntricos, (b) Espirais concêntricas, (c) Diabetes, (d) Câncer de mama, (e) Iris, (f) Mata Atlântica, (g) Coluna vertebral, (h) Sementes, (i) Vinho e (j) *Yeast*.

Tabela 5: Energia acumulada para dados  $d$ -dimensional e taxa de reconhecimento para os métodos kNN,  $K$ -means e LDA aplicados aos dados transformados pelo PCA. Dados de Círculos concêntricos, Espirais concêntricas, Diabetes, Câncer de mama, Iris, Mata Atlântica, Coluna vertebral, Sementes, Vinho e *Yeast*.

Dados	Energia (%)	Classes	kNN (%)	$K$ -means (%)	LDA (%)
Círculos concêntricos	100,00	Classe 1	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
		Classe 2	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
Espirais concêntricas	100,00	Classe 1	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
		Classe 2	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
Diabetes	100,00	Classe 1	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
		Classe 2	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
Câncer de mama	100,00	Classe 1	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
		Classe 2	100,00 (0,00)	97,40 (2,50)	98,70 (1,25)
Iris	99,54	Classe 1	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
		Classe 2	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
		Classe 3	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
Mata Atlântica	99,27	Classe 1	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
		Classe 2	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
		Classe 3	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
Coluna vertebral	98,33	Classe 1	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
		Classe 2	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
		Classe 3	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
Sementes	99,86	Classe 1	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
		Classe 2	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
		Classe 3	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
Vinho	98,44	Classe 1	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
		Classe 2	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
		Classe 3	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
<i>Yeast</i>	97,68	Classe 1	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
		Classe 2	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
		Classe 3	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
		Classe 4	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)



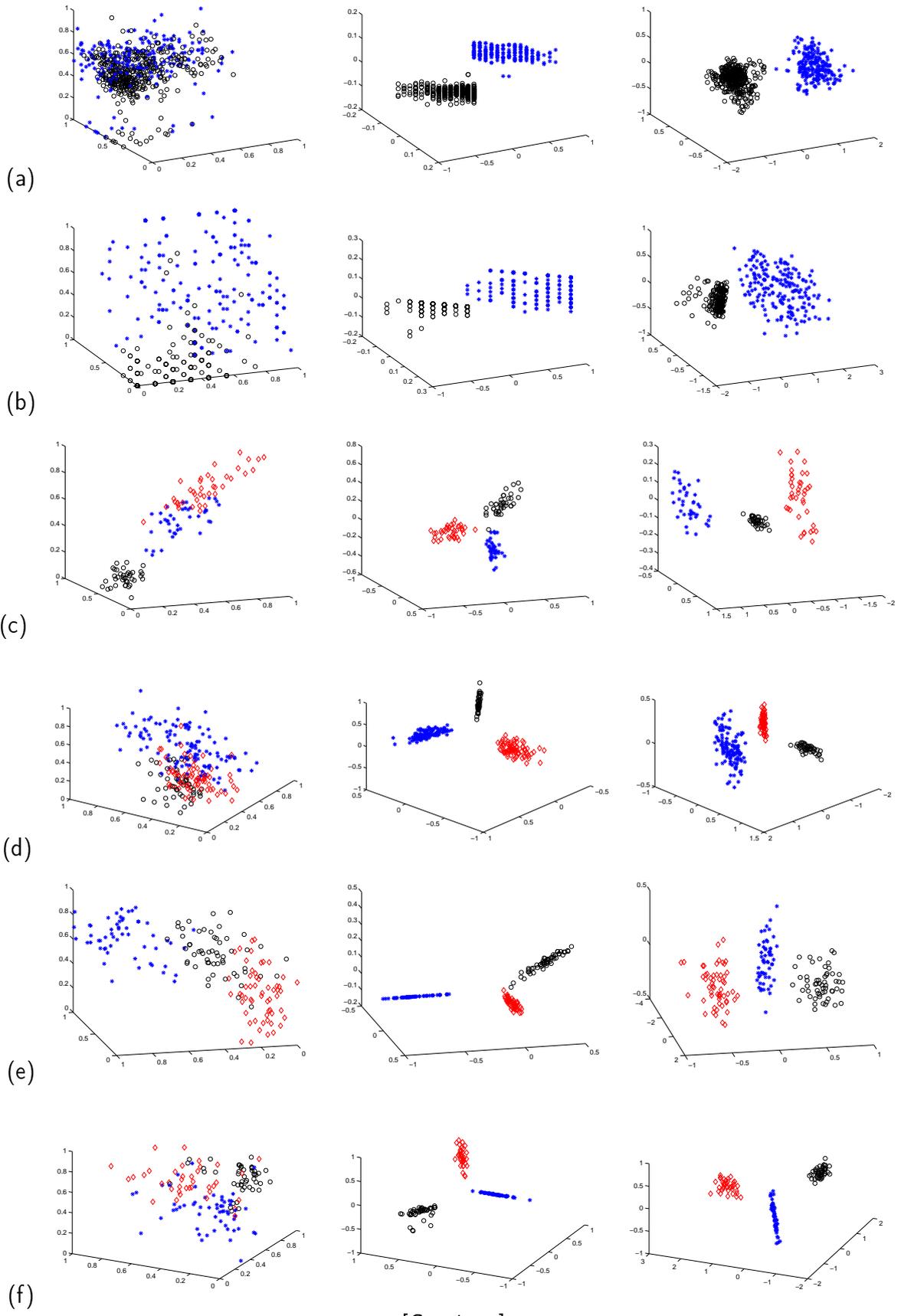
(a) Transformação com AG.



(b) Transformação com PCA.

Figura 14: Projeção do espaço complexo do conjunto de dados de Mata Atlântica pernambucana no  $\mathbb{R}^3$ , espaço obtido com a aplicação da metodologia.

de saída do conjunto de teste.



[Continua]

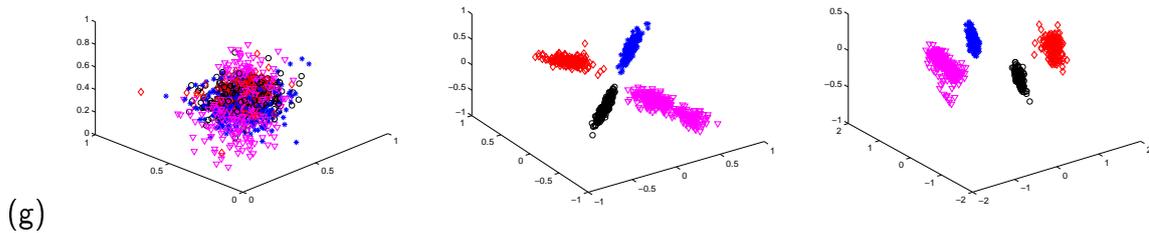


Figura 15: Projeções no  $\mathbb{R}^3$  para os dados originais, transformados pelo AG e transformados pelo método PCA, respectivamente. Para os conjuntos de dados: (a) Diabetes, (b) Câncer de mama, (c) Iris, (d) Coluna vertebral, (e) Sementes, (f) Vinho e (g) *Yeast*.

Esta RNA tem como conjunto de dados de entrada para treino 80% dos dados originais, e conjunto de saída de treino os dados produzidos pelo método PCA. E para conjunto de dados de entrada para teste foram preservados 20% dos dados originais.

Após treinamento da rede e processo de mapeamento ser realizado, os métodos kNN, *K-means* e LDA foram aplicados para tarefa de classificação dos dados originais preservados para teste e dos novos dados mapeados pela RNA. Os resultados foram comparados para garantir que o método empregado de fato discrimina os dados e que o mapeamento realizado é adequado.

Para o processo de aprendizagem da transformação de dados originais em dados transformados, ou seja, para mapear novos pontos no espaço de representação gerado com o incremento da dimensionalidade, foi usada uma rede neural *multilayer perceptron feedforward* com algoritmo *backpropagation*. Funções de treinamento que atualizam valores de pesos e vieses de acordo com o gradiente descendente com momento e uma taxa de aprendizagem adaptativa (*gda\_mlr*), de acordo com o gradiente descendente com taxa de aprendizagem adaptativa (*gda\_lr*) e de acordo com a otimização de *Levenberg-Marquardt* (LM). A medida de desempenho é o erro quadrático médio (MSE).

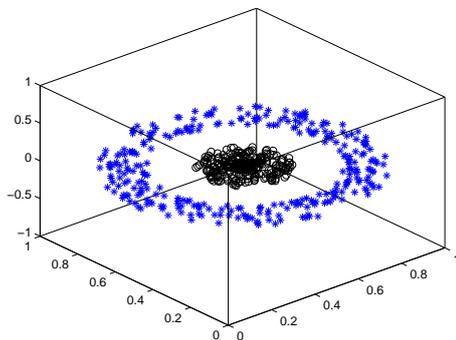
Para todos os conjuntos de dados usados a função de aprendizagem é o gradiente descendente com momento (*gdm*) e a função de transferência é a tangente sigmóide hiperbólica (*htansig*). Para o treinamento as taxas de incremento e decremento são 1,01 e 0,3, respectivamente. Nos casos em que a função de treinamento é a (*gda\_mlr*) o momento adotado é 0,9. As arquiteturas podem ser vistas na Tabela 6.

Para os Círculos concêntricos os dados mapeados pela RNA exibem desempenho superior em relação aos dados originais para os três métodos aplicados, kNN, *K-means* e LDA. Comportamento bastante similar ocorre para os Espirais concêntricos, os dados mapeados são todos classificados corretamente se for considerado o desvio padrão, como pode ser visto nos

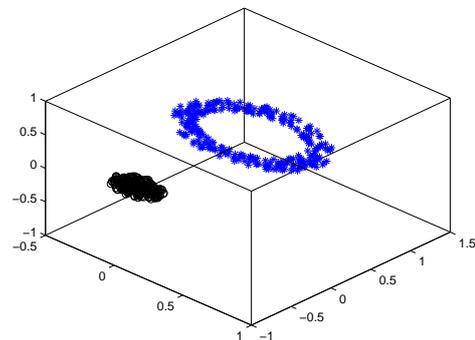
Tabela 6: Configurações de modelos de redes neurais MLP para Círculos concêntricos, Espirais concêntricas, Diabetes, Câncer de mama, Iris, Mata Atlântica, Coluna vertebral, Sementes, Vinho e Yeast.

Dados	MLP	Epoca	Treinamento	Taxa Aprend.	$\mu$
Círculos concêntricos	MLP(2,4,2)	5.000	LM		0,001
Espirais concêntricas	MLP(2,32,2)	5.000	LM		0,001
Diabetes	MLP(8,6,8)	5.000	LM		0,1
Câncer de mama	MLP(9,9,9)	20.000	gda_lr	0,1	
Iris	MLP(4,4,4)	50.000	gda_mlr	0,01	
Mata Atlântica	MLP(4,16,4)	5.000	LM		0,001
Coluna vertebral	MLP(6,6,6)	10.000	gda_lr	0,1	
Sementes	MLP(7,7,7)	20.000	gda_lr	0,01	
Vinho	MLP(13,13,13)	40.000	gda_lr	0,01	
Yeast	MLP(8,8,8)	5.000	LM		0,01

resultados exibidos nas Tabelas 7 e 8. Nas Figuras 16 e 17 pode ser observado que as classes são bem separáveis, dando evidência de que se a RNA adotada para mapear os dados for a mais adequada, para qualquer método de classificação empregado pode ser obtido um bom resultado.



(a) Projeção sobre  $\mathbb{R}^3$  do Espaço Real



(b) Projeção sobre  $\mathbb{R}^3$  dos dados transformados pelo PCA

Figura 16: Os Círculos concêntricos, (a) 600 pontos distribuídos em dois Círculos concêntricos no espaço real e (b) uma projeção no  $\mathbb{R}^3$  do espaço obtido com a aplicação da técnica PCA.

Os três métodos de classificação foram aplicados aos dados originais e mapeados de Diabetes, apenas o método LDA apresentou classificação inferior para a classe 2 dos dados mapeados. Embora os demais resultados para os dados mapeados apresentem maiores taxas de classificação, a diferença não é tão expressiva como era esperado. Uma vez que, quando transformados pelo PCA estes dados tornaram-se bem separáveis, como é ilustrado na Figura 15(a). Essa baixa taxa de classificação correta pode ser explicada como um mapeamento inadequado a partir de uma arquitetura de rede não conveniente para este conjunto de dados.

Nos dados originais e mapeados de Câncer de mama, para a classe 1 os resultados são quase

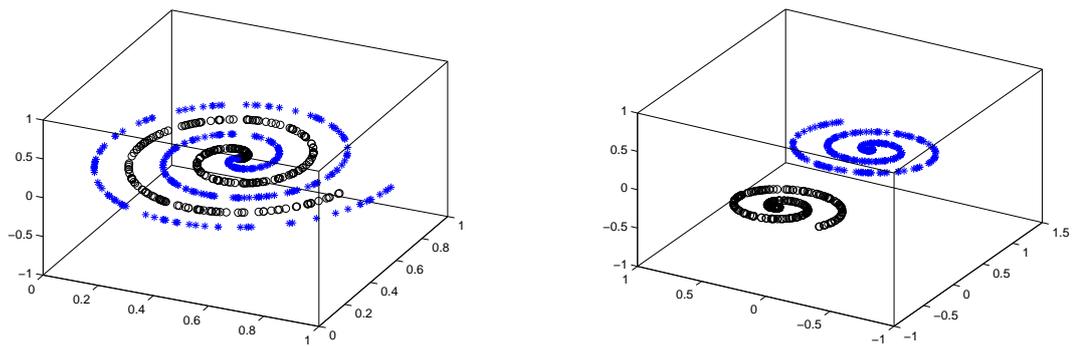
(a) Projeção sobre  $\mathbb{R}^3$  do Espaço Real(b) Projeção sobre  $\mathbb{R}^3$  dos dados transformados pelo PCA

Figura 17: Os Espirais concêntricas, (a) 600 pontos distribuídos em dois Espirais concêntricas no espaço real e (b) uma projeção no  $\mathbb{R}^3$  do espaço obtido com a aplicação da técnica PCA.

similares em ambos os conjuntos de dados (originais e mapeados) e para todos os métodos de classificação, mas para a classe 2 um desempenho menos favorável é exibido para os dados mapeados. Dessa forma, não sendo vantajoso os custos da aplicação dessa metodologia com essa arquitetura de rede sugerida.

Para os dados de Iris os três métodos apresentam resultados diferentes para as classes 2 e 3. A classe 1, para ambos conjuntos (originais e mapeados) exibe classificação máxima. O método *K-means* apresenta melhor desempenho para os dados mapeados, enquanto o LDA, para os dados originais. E o kNN tem melhor desempenho para os dados mapeados da classe 3 enquanto para a classe 2 é inferior.

Apenas o método LDA fornece melhor desempenho de classificação para as 3 classes dos dados mapeados em comparação aos dados originais de Mata Atlântica. Para os métodos kNN e *K-means* os resultados são similares, desempenho superior para os dados mapeados das classes 1 e 3. Ressalta-se que, embora melhores, estes resultados não são os esperados para aplicação dessa metodologia. Outra arquitetura de RNA deve ser buscada para o aprendizado do mapeamento proposto.

Aplicados os três métodos de classificação aos dados originais e mapeados de Coluna vertebral foram obtidos melhores resultados de classificação para a classe 2 dos dados mapeados, aproximando-se de 99% de acurácia. Enquanto para a classe 1, todos os resultados são inferiores para as três metodologias, o mesmo ocorrendo para a classe 3 dos dados mapeados quando o método *K-means* é aplicado.

Com os dados mapeados de Sementes os métodos *K-means* e LDA apresentam resultados contraditórios, enquanto com o *K-means* o desempenho é superior para todas as classes, com

o LDA é inferior, em comparação a classificação dos dados dados originais. Com o método kNN apresenta desempenho superior para a classe 1, inferior para a classe 2 e igual para a classe 3.

Para os dados mapeados de Vinho o método LDA exibe igual desempenho (100%) para todas as classes de ambos conjuntos de dados. Enquanto apenas a classe 1 exibe desempenho inferior na aplicação do método *K-means*, para os demais métodos e classes o desempenho é igual ou superior.

Na classificação dos dados mapeados de *Yeast*, o método LDA apresenta desempenho igual para a classe 3 e superior para as demais, enquanto para o método kNN é obtido desempenho inferior para a classe 1 e superior para as demais classes quando comparados ao desempenho na classificação para os dados originais. E o *K-means*, desempenho superior para as classes 1 e 4, e inferior para as demais.

Embora, em geral haja melhor taxa de reconhecimento quando os dados tem sido mapeados pela RNA, este resultado não é o esperado quando considera-se a discriminação dos dados observada a partir dos resultados exibidos na Tabela 5. Com isso, uma forma de mapeamento mais coerente deve ser proposta. E no uso das RNA's deve-se buscar uma arquitetura de rede que favoreça ao aprendizado da metodologia proposta. A taxa de reconhecimento para o método kNN é visto na Tabela 7 e para os métodos *K-means* e LDA na Tabela 8.

## 5.5 Estudo adicional - Comparando performance por número de classes

Adicionalmente um estudo foi realizado para verificação do número de classes possível a uma boa discriminação dos dados quando a dimensionalidade é incrementada. Foi considerado a classe Conservativa do conjunto de dados de Mata Atlântica pernambucana. Esta classe contém 19 espécies de 50 exemplares cada, e é descrita por quatro atributos.

Para a tarefa de classificação foram utilizados apenas os dados originais e os transformados pelo AG, com o método LDA, uma vez que tem-se por objetivo observar a organização dos dados no espaço complexo (nesse caso, quando a dimensionalidade cresce).

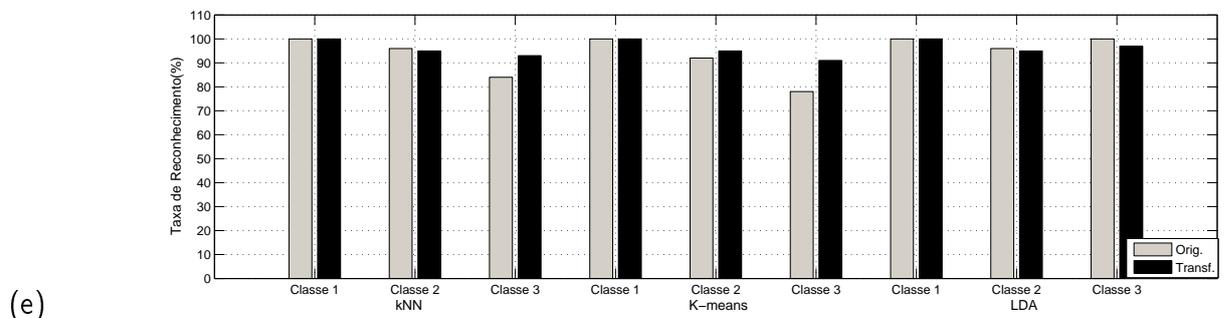
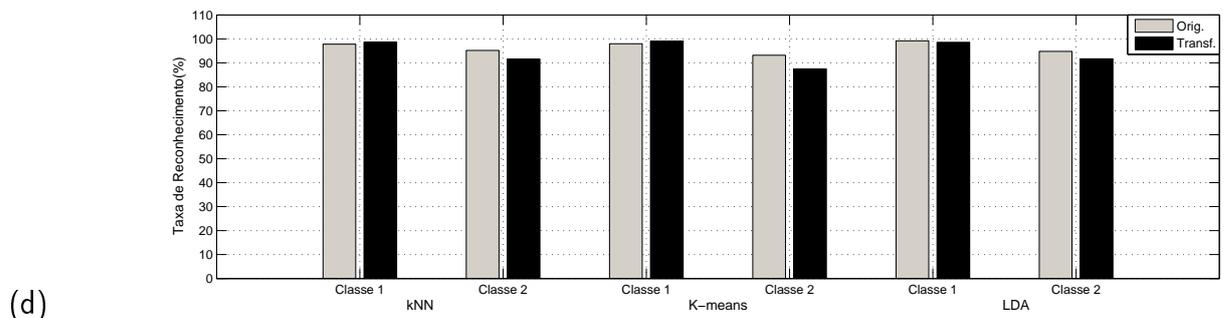
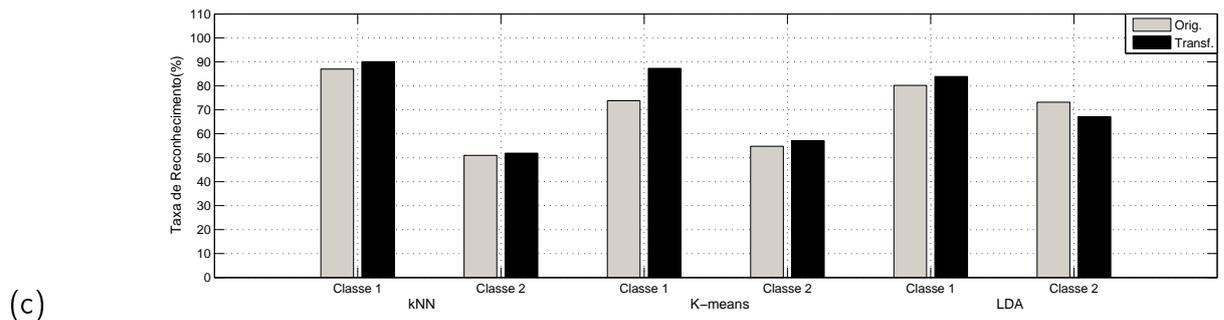
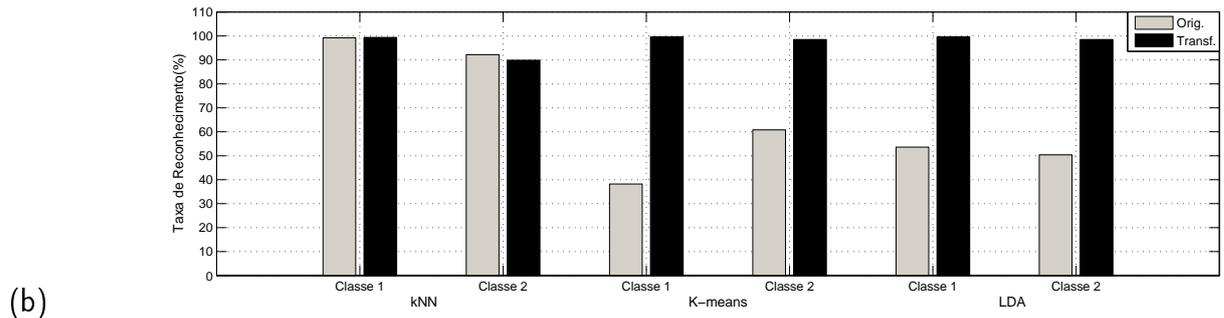
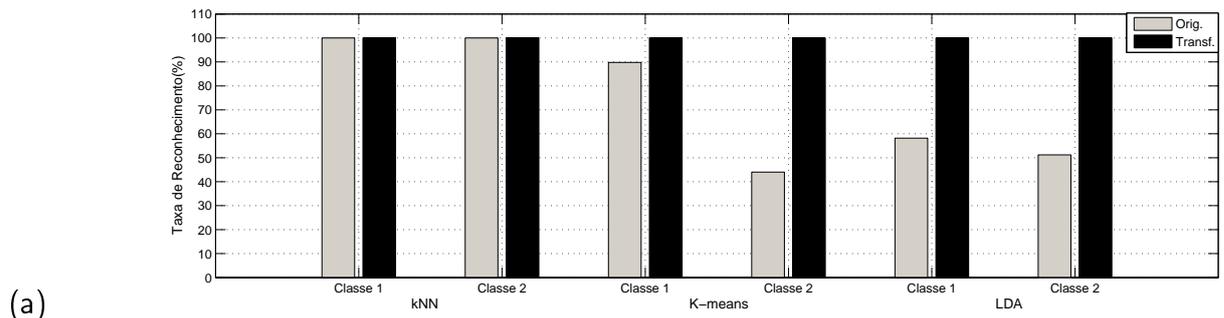
Cada conjunto de dados foi simulado dez vezes, sendo fornecido na Tabela 9 o número médio do percentual de acertos, com respectivo desvio padrão. Além disso, uma comparação da taxa de reconhecimento para os conjuntos de dados nas duas condições pode ser observada graficamente na Figura 19.

Tabela 7: Taxa de reconhecimento para o método kNN para os dados originais preservados e dados mapeados pela RNA de Círculos concêntricos, Espirais concêntricos, Diabetes, Câncer de mama, Iris, Mata Atlântica, Coluna vertebral, Sementes, Vinho e *Yeast*.

Dados	k máximo	Classe	Original (%)	Mapeado (RNA) (%)
Círculos concêntricos	25	Classe 1	100,00 (0,00)	100,00 (0,00)
		Classe 2	100,00 (0,00)	100,00 (0,00)
Espirais concêntricos	25	Classe 1	99,20 (1,23)	99,40 (0,97)
		Classe 2	92,20 (2,44)	98,90 (1,37)
Diabetes	25	Classe 1	87,00 (1,63)	90,10 (0,03)
		Classe 2	51,00 (4,00)	51,90 (4,07)
Câncer de mama	23	Classe 1	97,80 (2,35)	98,70 (1,25)
		Classe 2	95,20 (5,59)	91,60 (5,25)
Iris	11	Classe 1	100,00 (0,00)	100,00 (0,00)
		Classe 2	96,00 (5,16)	95,00 (5,27)
		Classe 3	84,00 (12,65)	93,00 (8,23)
Mata Atlântica	51	Classe 1	53,20 (5,30)	59,50 (3,83)
		Classe 2	61,00 (8,61)	57,80 (5,31)
		Classe 3	66,00 (10,15)	78,80 (2,94)
Coluna vertebral	15	Classe 1	80,00 (13,38)	72,30 (13,18)
		Classe 2	88,60 (8,66)	94,40 (4,25)
		Classe 3	71,00 (11,74)	78,50 (10,29)
Sementes	13	Classe 1	87,40 (8,60)	91,60 (2,95)
		Classe 2	95,80 (3,61)	95,10 (4,72)
		Classe 3	95,80 (5,90)	95,80 (6,76)
Vinho	13	Classe 1	100,00 (0,00)	100,00 (0,00)
		Classe 2	91,60 (5,52)	95,10 (4,72)
		Classe 3	98,00 (4,22)	100,00 (0,00)
<i>Yeast</i>	35	Classe 1	62,60 (5,27)	49,60 (7,52)
		Classe 2	47,20 (5,43)	48,50 (4,93)
		Classe 3	54,20 (4,54)	61,20 (7,21)
		Classe 4	74,60 (11,97)	82,50 (10,60)

Tabela 8: Taxa de reconhecimento para os métodos *K-means* e LDA para os dados originais preservados e dados mapeados pela RNA de Círculos concêntricos, Espirais concêntricos, Diabetes, Câncer de mama, Iris, Mata Atlântica, Coluna vertebral, Sementes, Vinho e Yeast.

Dados	Classe	<i>K-means</i>		LDA	
		Original (%)	Mapeado (%)	Original (%)	Mapeado (%)
Círculos concêntricos	Classe 1	89,80 (4,73)	100,00 (0,00)	58,20 (8,07)	100,00 (0,00)
	Classe 2	44,00 (2,40)	100,00 (0,00)	51,20 (2,70)	100,00 (0,00)
Espirais concêntricos	Classe 1	38,20 (6,05)	99,60 (0,51)	53,60 (5,48)	99,60 (0,52)
	Classe 2	60,80 (8,01)	98,50 (1,84)	50,40 (3,75)	98,50 (1,84)
Diabetes	Classe 1	73,80 (2,35)	87,30 (2,54)	80,20 (4,34)	83,90 (2,42)
	Classe 2	54,80 (5,63)	57,10 (3,60)	73,20 (4,78)	67,10 (4,04)
Câncer de mama	Classe 1	98,00 (2,40)	99,10 (1,19)	99,20 (0,79)	98,60 (1,58)
	Classe 2	93,20 (4,34)	87,50 (4,99)	94,80 (3,15)	91,70 (4,34)
Iris	Classe 1	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
	Classe 2	92,00 (7,89)	95,00 (5,27)	96,00 (5,16)	95,00 (5,27)
	Classe 3	78,00 (19,32)	91,00 (7,38)	100,00 (0,00)	97,00 (6,75)
Mata Atlântica	Classe 1	12,60 (2,07)	17,20 (17,87)	55,20 (5,05)	72,10 (3,57)
	Classe 2	71,60 (7,41)	68,10 (6,03)	62,40 (3,31)	69,00 (3,62)
	Classe 3	52,60 (4,65)	75,30 (7,51)	47,40 (6,42)	62,70 (2,75)
Coluna vertebral	Classe 1	85,00 (10,31)	84,20 (10,01)	83,60 (11,17)	82,50 (7,43)
	Classe 2	57,40 (6,20)	91,70 (7,10)	88,60 (8,66)	95,40 (3,69)
	Classe 3	81,00 (10,22)	73,50 (10,29)	77,00 (12,29)	79,00 (7,38)
Sementes	Classe 1	86,00 (4,67)	90,90 (3,38)	98,60 (2,95)	93,70 (2,21)
	Classe 2	94,40 (5,52)	95,10 (4,72)	98,60 (2,95)	97,20 (3,61)
	Classe 3	92,80 (11,84)	97,90 (3,38)	98,60 (2,95)	96,50 (4,95)
Vinho	Classe 1	100,00 (0,00)	99,20 (2,53)	100,00 (0,00)	100,00 (0,00)
	Classe 2	87,40 (7,23)	96,50 (4,95)	100,00 (0,00)	100,00 (0,00)
	Classe 3	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)	100,00 (0,00)
Yeast	Classe 1	59,80 (3,36)	67,60 (6,96)	56,33 (7,50)	58,60 (7,21)
	Classe 2	58,40 (9,08)	55,80 (3,77)	54,66 (1,37)	56,10 (7,03)
	Classe 3	61,40 (4,09)	59,20 (8,66)	61,00 (7,15)	60,80 (6,14)
	Classe 4	45,20 (8,09)	66,00 (10,24)	77,33 (12,72)	78,70 (11,66)



[Continua]

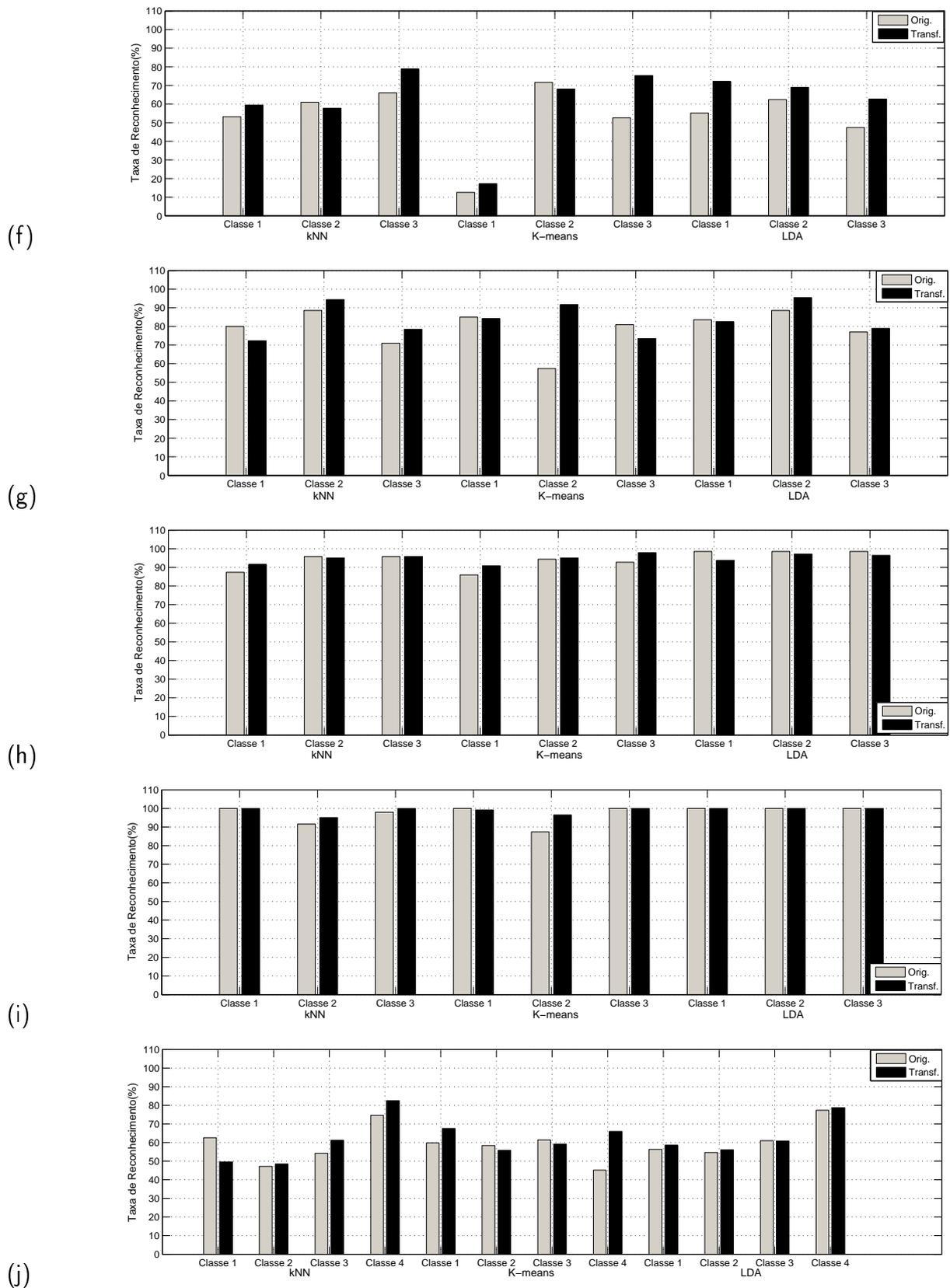


Figura 18: Taxa de reconhecimento para os métodos kNN, *K-means* e LDA para os dados brutos e dados mapeados (RNA) de: (a) Círculos concêntricos, (b) Espirais concêntricas, (c) Diabetes, (d) Câncer de mama, (e) Iris, (f) Mata Atlântica, (g) Coluna vertebral, (h) Sementes, (i) Vinho e (j) *Yeast*.

Tabela 9: Taxa de reconhecimento para o método LDA aplicado aos dados brutos e dados transformados (AG) da classe Conservativa dos dados de Mata Atlântica pernambucana considerando diferentes números de espécies.

Número de espécies	Original (%)	Transf. (AG) (%)
4 Espécies	87,00 (3,00)	100,00 (0,00)
5 Espécies	86,00 (1,00)	100,00 (0,00)
6 Espécies	86,00 (1,00)	100,00 (0,00)
7 Espécies	86,00 (1,00)	100,00 (0,00)
8 Espécies	81,00 (1,00)	100,00 (0,00)
9 Espécies	79,00 (1,00)	99,00 (0,00)
10 Espécies	72,00 (2,00)	97,00 (2,00)
11 Espécies	71,00 (1,00)	96,00 (2,00)
12 Espécies	71,00 (1,00)	97,00 (2,00)
13 Espécies	66,00 (1,00)	95,00 (2,00)
14 Espécies	66,00 (1,00)	96,00 (2,00)
15 Espécies	67,00 (1,00)	93,00 (2,00)
16 Espécies	65,00 (1,00)	94,00 (1,00)
17 Espécies	62,00 (1,00)	93,00 (1,00)
18 Espécies	61,00 (1,00)	92,00 (2,00)
19 Espécies	55,00 (2,00)	94,00 (1,00)

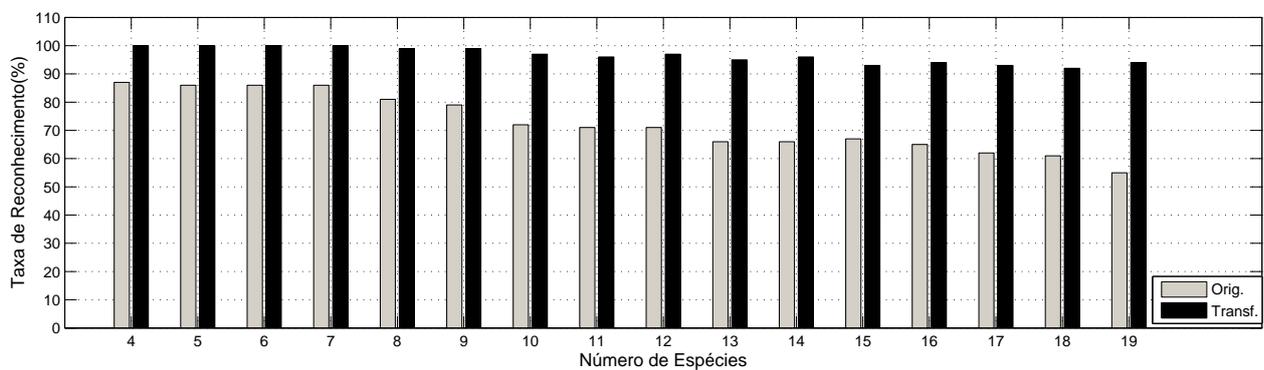
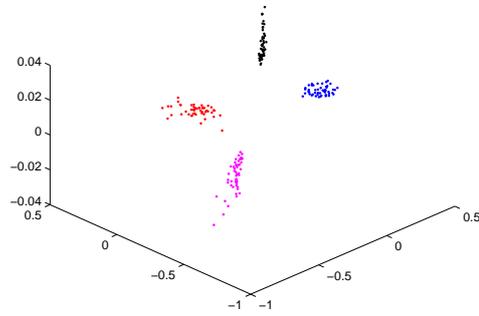
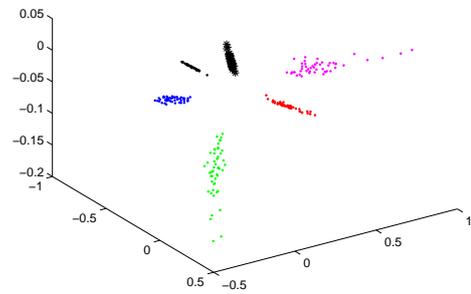


Figura 19: Taxa de reconhecimento para o método LDA aplicado aos dados brutos e dados transformados (AG) da classe Conservativa dos dados de Mata Atlântica pernambucana considerando diferentes números de espécies.

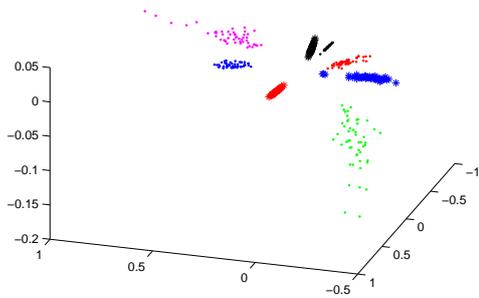
Na Figura 20 são exibidas projeções da discriminação dos dados no espaço complexo, obtido pela aplicação da metodologia proposta.



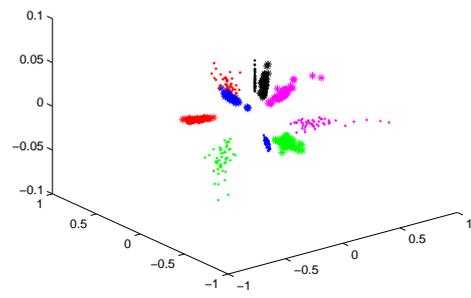
(a) 4 Espécies



(b) 6 Espécies



(c) 8 Espécies



(d) 10 Espécies

Figura 20: Projeções do espaço complexo no  $\mathbb{R}^3$  para o conjunto de dados da classe Conservativa com diferentes números de espécies.

## 6 Conclusões

Conclusões e perspectivas de possíveis desdobramentos para trabalhos futuros que poderão contribuir na solução de difíceis problemas de classificação.

### 6.1 Conclusões

Neste trabalho foi proposto um método com inspiração na Mecânica Quântica e baseado no teorema de Cover, desenvolvido para incrementar a dimensionalidade de conjuntos de dados, tornando problemas não linearmente separáveis em problemas linearmente separáveis em um outro espaço de representação. Com isso, o poder de discriminar dados é aprimorado e a complexidade na tarefa de classificação é diminuída, o que implica em um melhor desempenho na classificação para inúmeros métodos adotados. Assumiu-se a existência de um observável quântico gerando os valores observados dos conjuntos de dados e buscou-se a transformada inversa que torna dados reais em dados complexos, aumentando a dimensionalidade do conjunto de dados original. Um algoritmo genético foi usado para buscar esta transformação inversa, agindo como um operador inverso. A busca é a tarefa de definir um conjunto de pares numéricos ordenados que mapeia um elemento do espaço de representação real em um elemento no espaço de representação complexo.

Dez conjuntos de dados de problemas não linearmente separáveis foram usados para testar a metodologia proposta. Três diferentes algoritmos clássicos kNN, *K-means* e LDA, foram aplicados aos dados originais e transformados pelo novo método, os resultados foram comparados demonstrando a viabilidade do processo na discriminação dos dados. Os resultados experimentais mostram a capacidade da metodologia em fornecer um melhor desempenho na tarefa de classificação. Além disso, baseado no teorema de Cover [7], esta metodologia é capaz, com alta probabilidade, de tornar os dez problemas não linearmente separáveis, Círculos concêntricos, Espirais concêntricas, Diabetes, Câncer de mama, Iris, Mata Atlântica, Coluna vertebral, Sementes, Vinho e *Yeast*, em problemas linearmente separáveis em outro espaço de representação.

Em seguida foi proposta uma transformação que diminui a complexidade dos dados, uma vez que a dimensionalidade é reduzida de  $2d$ -dimensional para  $d$ -dimensional, mas com a capacidade de discriminação e classificação semelhante aos dados com dimensão incrementada. A técnica PCA foi assumida como sendo esta transformação e para verificar quanto à capacidade em discriminar os dados os métodos kNN, *K-means* e LDA foram utilizados novamente.

Para mapear um novo ponto no espaço de representação obtido anteriormente, uma rede neural MLP foi treinada e as unidades de saída vinculado ao conjunto de teste geraram um novo conjunto de dados. Os métodos de classificação kNN, *K-means* e LDA foram empregados e os resultados para os dados originais e mapeados pela RNA, de Círculos concêntricos, Espirais concêntricas, Diabetes, Câncer de mama, Iris, Mata Atlântica, Coluna vertebral, Sementes, Vinho e *Yeast* foram comparados para garantir a robustez da metodologia.

Nos conjuntos de dados transformados com o algoritmo genético, onde a dimensionalidade foi incrementada, a taxa de reconhecimento na classificação foi 100% para todos os métodos de classificação empregados, exceto para os dados de Câncer de mama, onde os métodos *K-means* e LDA não apresentaram igual taxa de reconhecimento, mas ainda forneceram resultados superiores àqueles obtidos para os dados originais.

Similarmente, os dados transformados pelo PCA apresentaram um excelente desempenho na tarefa de classificação, exceto para os dados de Câncer de mama, quando aplicados os métodos *K-means* e LDA. Uma vez que o comportamento na classificação se repete dos dados transformados pelo AG para os dados transformados pelo PCA, pode-se considerar que as características sobre a separabilidade são mantidas, no entanto, um novo conjunto de dados com menor complexidade é gerado.

Na classificação dos dados mapeados pela RNA, em geral obteve-se desempenho igual ou superior aos dados originais. Entretanto, o resultado obtido não é o esperado, uma vez que a discriminação dos dados tem sido verificada com o desempenho na tarefa de classificação para os dados usados na fase de treinamento da RNA. Desse modo, uma forma de mapeamento mais coerente ao aprendizado da metodologia deve ser proposta, partindo da busca de uma arquitetura de rede mais conveniente para essa tarefa.

Adicionalmente, foi verificado quantas classes esta metodologia é hábil em discriminar. Um conjunto de dados com 19 classes e 4 dimensões foi considerado, a dimensionalidade foi incrementada e a seguir aplicado o método de classificação LDA. Até 8 classes obteve uma taxa de reconhecimento de 100%, à medida que o número de classes consideradas aumentou, a taxa de reconhecimento caiu. No entanto, ainda é considerado um bom resultado uma taxa de reconhecimento de 91% para um conjunto de 19 classes. E quando comparado ao desempenho

no conjunto de dados sem a transformação proposta, este resultado mostra-se ainda mais relevante. Além dessa observação em relação ao número de classes, uma investigação adicional deve ser feita em torno da dimensionalidade do conjunto de dados original.

Assim, o método proposto para discriminar dados a partir do incremento de sua dimensionalidade apresenta-se como uma nova metodologia viável na abordagem do problema de classificação. Uma vez preprocessados os dados, discriminados em um novo espaço de representação, qualquer método usado para classificação pode oferecer um bom desempenho.

## 6.2 Trabalhos Futuros

A proposta inicial deste trabalho era obtenção de uma forma algébrica para resolução de problemas de classificação. Inspirado no teorema de Cover, a dimensionalidade de um conjunto de dados deveria ser incrementada gerando um espaço de representação aonde a tarefa de classificação seria viável devido a discriminação dos dados.

Dado a não trivialidade em obter uma solução algébrica, uma alternativa foi a busca de uma solução por meio da inteligência artificial. Um algoritmo genético foi usado com esse propósito. E resultados consideráveis foram obtidos e apresentados neste trabalho de doutorado.

A solução da seguinte equação fornece resultados numéricos para solução deste problema,

$$f(a_j^{(d)}) = \frac{\sum_{t=1}^k \sum_{j>t}^k \left( Z_{\mathbf{C}}^{(t)} - Z_{\mathbf{C}}^{(j)} \right)^2}{1 + \sum_{g=1}^k \sum_{v>g}^k \sqrt{\frac{1}{n_v-1} \sum_{w=1}^{n_v} \left( Z_{\mathbf{C}}^{(g)} - z_w^{(v)} \right)^2}}$$

em que  $Z_{\mathbf{C}}^{(l)} = \frac{1}{n_l} \sum_{j=1}^{n_l} a_j^{(l)} - \imath b_j^{(l)}$  é o centróide da classe  $l$ ,  $n_l$  é o número de elementos da classe  $l$ ,  $\imath = \sqrt{-1}$  é a constante complexa e  $z_w^{(l)} = a_w^{(l)} + \imath b_w^{(l)}$  é um ponto no espaço complexo de acordo com o  $w$ -ésimo ponto real da classe  $l$  no conjunto de dados original.

Uma forma mais bem definida já foi obtida e está em fase de teste.

- Para a definição das equações do sistema de equações não lineares, foi utilizado o método dos multiplicadores de Lagrange [79]. Usual em problemas de otimização, ele permite encontrar extremos de uma função de uma ou mais variáveis suscetíveis a uma ou mais restrições.
- A solução numérica foi dada a partir do método de Jacobi, por este ser bastante usual na solução de problemas similares.

Como trabalhos futuros, tem-se:

- Ainda no uso do algoritmo genético:
  - Investigar a que tipos de conjuntos de dados essa metodologia se adequa. Observar parâmetros como número de classes, número de dimensões, tamanho de conjunto de dados;
  - Buscar outras arquiteturas de RNAs que foreneçam o perfeito mapeamento dos dados.
  
- A partir da solução do sistema de equações que está em fase de teste:
  - Aplicá-la aos mesmos conjuntos de dados observados anteriormente, para uma possível comparação entre as metodologias;
  - Observar a que tipos de dados se aplica;
  - Testar uma outra forma de mapeamento dos dados.

# Referências

- 1 DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern Classification*. Second. New York: John Wiley & Sons, 2000.
- 2 WU, X.; ZHU, X. Mining with noise knowledge: error-aware data mining. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 2008.
- 3 SAEZ, J. A.; LUENGO, J.; HERRERA, F. Predicting noise filtering efficacy with data complexity measures for nearest neighbor classification. *Pattern Recognition*, 2013.
- 4 CANO, J.-R. Analysis of data complexity measures for classification. *Expert Systems with Applications*, 2013.
- 5 VAPNIK, V. N. *Statistical Learning Theory*. New York: John Wiley & Sons, 1998.
- 6 ELIZONDO, D. The linear separability problem: Some testing methods. *IEEE Transactions on Neural Networks*, 2006.
- 7 COVER, T. M. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers*, 1965.
- 8 MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: *5th Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press, 1967. p. 281–297.
- 9 COVER, T.; HART, P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, v. 13, n. 1, p. 21–27, 1967. ISSN 0018-9448.
- 10 FISHER, R. A. The use of multiple measurements in taxonomic problems. *Ann. Eugen.*, v. 7, p. 179–188, 1936.
- 11 PRABHU, P.; ANBAZHAGAN, N. Improving the performance of k-means clustering for high dimensional data set. *International Journal on Computer Science and Engineering (IJCSE)*, v. 3, n. 6, p. 2317–2322, 2011.
- 12 ZHANG, C.; FANG, Z. An improved k-means clustering algorithm. *Journal of Information & Computational Science*, v. 10, p. 193–199, 2013.
- 13 HUANG, X. et al. Dskmeans: A new kmeans-type approach to discriminative subspace clustering. *Knowledge-Based Systems*, v. 70, p. 293–300, 2014.
- 14 GAO, Y.; GAO, F. Edited adaboost by weighted knn. *Neurocomputing*, v. 73, p. 3079–3088, 2010.
- 15 FREUND, Y.; SCHAPIRE, R. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 1997.

- 16 HUANG, Z.; ZHOU, Z.; HE, T. Associative classification with knn. *Journal of Theoretical and Applied Information Technology*, v. 49, 2013.
- 17 PANAH, N. et al. Recognition of different datasets using pca, lda, and various classifiers. In: *Application of Information and Communication Technologies (AICT), 5th International Conference*. Baku: [s.n.], 2011.
- 18 XU, R.; LI, D. W. Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 2005.
- 19 WU, X. et al. Top 10 algorithms in data mining. *Knowledge and Information Systems*, 2008.
- 20 BRAGA, A. P.; LUDERMIR, T. B.; CARVALHO, A. C. P. L. F. *Redes Neurais Artificiais Teoria e aplicações*. First. Rio de Janeiro: LTC, 2000.
- 21 HAYKIN, S. *Redes Neurais Principios e pratica*. Second. Porto Alegre: Bookman, 2001.
- 22 RUMELHART, D. E.; HINTON, G.; WILLIAMS, R. Learning representations by back-propagating errors. *Nature*, v. 323, n. 11, p. 533–536, 1986.
- 23 HOLLAND, J. H. *Adaptation in natural and artificial systems*. Ann Arbor: University of Michigan Press, 1975.
- 24 COHEN-TANNOUDJI, C.; DIU, B.; LALOE, F. *Quantum Mechanics*. Paris: John Wiley & Sons, 1977.
- 25 NIELSEN, M. A.; CHUANG, I. L. *Quantum Computation and Quantum Information*. First. New York: Cambridge University Press, 2000.
- 26 OECKL, R. Observables in the general boundary formulation. In: *Quantum Field Theory and Gravity*. [S.l.]: Springer, 2012. p. 137–156.
- 27 HOLEVO, A. S. Information capacity of a quantum observable. *Problems of Information Transmission*, Springer, v. 48, n. 1, p. 1–10, 2012.
- 28 PAPE, M.; STREICHER, T. Computability for basic quantum mechanics based on the Hilbert lattice. In: *Eleventh International Conference on Computability and Complexity in Analysis*. Darmstadt, Germany: [s.n.], 2014. p. 50.
- 29 TRUONG, T. T. A mathematical model for measurements in quantum mechanics. *arXiv preprint arXiv:1407.5519*, 2014.
- 30 KOUDA, N. et al. An examination of qubit neural network in controlling an inverted pendulum. *Neural Processing Letters*, v. 22, p. 277–290, 2005.
- 31 AZEVEDO, C. R. B.; FERREIRA, T. A. E. The application of qubit neural networks to time series forecasting with automatic phase adjustment mechanism. In: *Anais do XXVII Congresso da Sociedade Brasileira de Computação, VI Encontro Nacional de Inteligência Artificial. Sociedade Brasileira de Computação*. Rio de Janeiro: SBC Press, 2007. p. 1112–1121.
- 32 OLIVEIRA, W. et al. Quantum logical neural networks. In: *Neural Networks. SBRN '08. 10th Brazilian Symposium*. Salvador: IEEE computer society, 2008. p. 147–152.

- 33 TAKAHASHI, K.; KUROKAWA, M.; HASHIMOTO, M. Remarks on multi-layer quantum neural network controller trained by real-coded genetic algorithm. In: *Intelligent Science and Intelligent Data Engineering*. [S.l.]: Springer, 2012. p. 50–57.
- 34 LIU, C.-Y. et al. Single-hidden-layer feed-forward quantum neural network based on Grover learning. *Neural Networks*, Elsevier, v. 45, p. 144–150, 2013.
- 35 SCHULD, M.; SINAYSKIY, I.; PETRUCCIONE, F. The quest for a quantum neural network. *Quantum Information Processing*, Springer, v. 13, n. 11, p. 2567–2586, 2014.
- 36 NASIOS, N.; BORS, A. G. Kernel-based classification using quantum mechanics. *Pattern Recognition*, v. 40, p. 875–889, 2007.
- 37 WIEBE, N.; KAPOOR, A.; SVORE, K. M. *Quantum Nearest-Neighbor Algorithms for Machine Learning*. 2014. Artigo submetido para Quantum Physics: arXiv:1401.2142v2.
- 38 SINGHA, M. P.; RAJPUT, B. Pattern classification using maximally entangled quantum states (MES). *Eur. Phys. J. Plus*, 2014.
- 39 SCHULD, M.; SINAYSKIY, I.; PETRUCCIONE, F. *Quantum computing for pattern classification*. 2014. Artigo submetido para Quantum Physics: arXiv:1412.3646v1.
- 40 SCHULD, M.; SINAYSKIY, I.; PETRUCCIONE, F. *An introduction to quantum machine learning*. 2014. Artigo submetido para Quantum Physics: arXiv:1409.3097v1.
- 41 TANG, E. et al. Linear dimensionality reduction using relevance weighted LDA. *Pattern Recognition*, v. 38, p. 485 – 493, 2005.
- 42 JARCHI, D.; BOOSTANI, R. A new weighted LDA method in comparison to some versions of LDA. *World Academy of Science, Engineering and Technology*, v. 2, n. 12, p. 759 – 764, 2008.
- 43 KOHONEN, T. Self-organized formation of topologically connect feature maps. *Biological Cybernetics*, v. 43, p. 59–69, 1982.
- 44 MOKRIS, I.; FORGAC, R. *Decreasing the Feature Space Dimension by Kohonen Self-Organizing Maps*. 2004. 2nd Slovakian - Hungarian Joint Symposium on Applied Machine Intelligence, Herlany, Slovakia.
- 45 PLASTRIA, F.; BRUYNE, S. D.; CARRIZOSA, E. Dimensionality reduction for classification. In: *4th international conference on Advanced Data Mining and Applications*. Berlin Heidelberg: Springer-Verlag, 2008. p. 411–418.
- 46 MASAELI, M.; DY, J. G.; FUNG, G. M. From transformation-based dimensionality reduction to feature selection. In: *27th International Conference on Machine Learning (ICML-10)*. Omnipress, 2010. p. 751–758. Disponível em: <<http://www.icml2010.org/papers/333.pdf>>.
- 47 KONG, S.; WANG, D. Learning exemplar-represented manifolds in latent space for classification. *Machine Learning and Knowledge Discovery in Databases*, v. 8190, p. 240–255, 2013.
- 48 CORTES, C.; VAPNIK, V. Support-vector networks. *Machine Learning*, v. 20, n. 3, p. 273–297, 1995. Disponível em: <<http://dx.doi.org/10.1007/BF00994018>>.

- 49 CAMPBELL, C. Kernel methods: A survey of current techniques. *Neurocomputing*, v. 48, p. 63–84, 2000.
- 50 ZHANG, D.; CHEN, S. Fuzzy clustering using kernel method. In: *2002 International Conference on Control and Automation*. China: [s.n.], 2002. p. 162–163.
- 51 HOFMANN, T.; SCHOLKOPF, B.; SMOLA, A. J. Kernel methods in machine learning. *Annals of Statistics*, 2008.
- 52 DARWIN, C. *The Origin of Species*. [S.l.: s.n.], 1859.
- 53 FERREIRA, T. A. E.; VASCONCELOS, G. C.; ADEODATO, P. J. L. A new intelligent system methodology for time series forecasting with artificial neural networks. *Neural Processing Letters*, Springer US, v. 28, n. 2, p. 113–129, 2008. ISSN 1370-4621. Disponível em: <<http://dx.doi.org/10.1007/s11063-008-9085-x>>.
- 54 NETO, P. S. de M. et al. Hybrid intelligent system for air quality forecasting using phase adjustment. *Engineering Applications of Artificial Intelligence*, v. 32, p. 185–191, 2014.
- 55 GOLDBERG, D. E. *Genetic algorithms in search, optimization and machine learning*. Reading, MA.; Alabama: Addison-Wesley, 1989.
- 56 PEARSON, K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, v. 2, p. 559–572, 1901.
- 57 HOTELLING, H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, v. 24, p. 417–441 and 498–520, 1933.
- 58 HOPFIELD, J. J. Neural networks and physical systems with emergent collective computational abilities. In: *National Academy of the Sciences of the U.S.A.* Washington: National Academy of the Sciences, 1982. v. 79, p. 2554–2558.
- 59 CARPENTER, G. A.; GROSSBERG, S. The art of adaptive pattern recognition by a self-organizing neural network. *Computer*, v. 21, n. 3, p. 77–88, 1988.
- 60 ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.*, v. 65, p. 386–408, 1958.
- 61 DEBNATH, L.; MIKUSIŃSKI, P. *Hilbert Spaces with Applications*. Third. China: Elsevier Academic Press, 2005.
- 62 GRIFFITHS, D. J. *Introduction to quantum mechanics*. Second. New Jersey: Prentice Hall, 2004.
- 63 JR, O. P. *Conceitos de física quântica*. First. São Paulo: Livraria da Física, 2003.
- 64 LEUNG, F. H. F. et al. Tuning of the structure and parameters of a neural network using an improved genetic algorithm. *Neural Networks, IEEE Transactions on*, v. 14, n. 1, p. 79–88, 2003. ISSN 1045-9227.
- 65 KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Fourteenth International Joint Conference on Artificial Intelligence*. Montreal: [s.n.], 1995. v. 2, p. 1137–1143.

- 66 SILVA, M. A. M. et al. *Does the Plant Economics Spectrum Change with Secondary Succession in the Forest?* 2014. Artigo submetido para a Revista *Trees - Structure and Function*.
- 67 BAZZAZ, F. A.; PICKETT, S. T. A. Physiological ecology of tropical succession: a comparative review. *Annual Review of Ecology and Systematics*, v. 11, p. 287–310, 1980.
- 68 WESTOBY, M.; WRIGHT, I. Land-plant ecology on the basis of functional traits. *Trends in Ecology and Evolution*, v. 21, p. 261–268, 2006.
- 69 CORNWEL, W. K.; ACKERLY, D. D. Community assembly and shifts in plant trait distributions across an environmental gradient in coastal California. *Ecological Monographs*, v. 79, p. 109–126, 2009.
- 70 SCHLEICHER, A.; PEPPLER-LISBACH, C.; KLEYER, M. Functional traits during succession: is plant community assembly trait-driven? *Preslia*, v. 83, p. 347–370, 2011.
- 71 FRANK, A.; ASUNCION, A. *UCI Machine Learning Repository*. abril 2013. Irvine, CA: University of California, School of Information and Computer Science. Disponível em: <<http://archive.ics.uci.edu/ml>>.
- 72 SARKAR, M.; LEONG, T.-Y. Application of k-nearest neighbor algorithm on breast cancer diagnosis problem. In: *AMIA Symposium*. Los Angeles: [s.n.], 2000. p. 759–763.
- 73 NASSER, A.; HAMAD, D. K-means clustering algorithm in projected spaces. In: *9th Int. Conf. Inf. Fusion*. Florence: [s.n.], 2006. p. 1–6.
- 74 LIU, H. X. et al. Diagnosing breast cancer based on support vector machines. *J. Chem. Inf. Comput. Sci.*, v. 43, n. 3, p. 900–907, 2003.
- 75 CAMASTRA, F. A novel kernel method for clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 27, n. 5, p. 801–804, 2005.
- 76 CAMASTRA, F. Kernel methods for clustering. In: *WIRN/NAIS Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 2005. v. 3931, p. 1–9.
- 77 SARMA, T. H.; VISWANATH, P.; REDDY, B. E. Single pass kernel k-means clustering method. *Sadhana - Academy Proceedings in Engineering Sciences*, Indian Academy of Sciences, v. 38, n. 3, p. 407–419, 2013.
- 78 PECHENIZKIY, M.; TSYMBAL, A.; PUURONEN, S. On combining principal components with Fisher's linear discriminants for supervised learning. *Foundations of Computing and Decision Sciences*, v. 31, n. 1, p. 59–73, 2006.
- 79 KLEIN, D. *Lagrange multipliers without permanent scarring*. [Http://www.cs.berkeley.edu/klein/papers/lagrangemultipliers.pdf](http://www.cs.berkeley.edu/klein/papers/lagrangemultipliers.pdf).