

TIAGO MARTINS PEREIRA

**DISCRIMINAÇÃO DE POPULAÇÕES COM DIFERENTES GRAUS DE
SIMILARIDADE POR REDES NEURAS ARTIFICIAIS**

Dissertação apresentada à
Universidade Federal de Viçosa,
como parte das exigências do
Programa de Pós-Graduação em
Estatística Aplicada e Biometria,
para obtenção do título de *Magister
Scientiae*.

VIÇOSA
MINAS GERAIS - BRASIL
2009

TIAGO MARTINS PEREIRA

**DISCRIMINAÇÃO DE POPULAÇÕES COM DIFERENTES GRAUS DE
SIMILARIDADE POR REDES NEURAS ARTIFICIAIS**

Dissertação apresentada à
Universidade Federal de Viçosa,
como parte das exigências do
Programa de Pós-Graduação em
Estatística Aplicada e Biometria,
para obtenção do título de *Magister
Scientiae*.

APROVADA: 15 de dezembro de 2009.

Prof. Adair José Regazzi
(Co-Orientador)

Prof. José Ivo Ribeiro Júnior
(Co-Orientador)

Prof. Luiz Alexandre Peternelli

Prof. Pedro Crescêncio S. Carneiro

Prof. Cosme Damião Cruz
(Orientador)

Ao
meu querido avô, Sr. Renato
Pereira, que por um acaso do
destino não está mais aqui para
presenciar essa conquista...

DEDICO

Agradecimentos

"Não mostre para Deus o tamanho do seu problema, mostre para o problema o tamanho do seu Deus" (Autor desconhecido)

Agradeço imensamente à minha mãe, Luzia, pelo carinho, amor, incentivo, apoio permanente e ao meu pai, José Raimundo, por estarem, apesar da distância, sempre presentes ao longo desta caminhada.

Ao meu irmão Raphael e sua esposa Bianca, e a toda minha família pela compreensão durante todos esses anos de Viçosa.

A minha namorada Diana, pelo carinho, companheirismo, amizade e principalmente pela compreensão nas horas mais difíceis dessa caminhada.

Ao Prof. Cosme Damião Cruz, por ter aceitado orientar e planejar este trabalho, durante todo o seu desenvolvimento, sendo sua atuação e colaboração essenciais para a conclusão desta pesquisa.

Aos co-orientadores Prof. Adair José Regazzi e Prof. José Ivo Ribeiro Júnior, e a todos os outros professores do Programa pela cordialidade e atenção que sempre me foram dispensadas.

A Universidade Federal de Viçosa, através do Curso de Pós-Graduação em Estatística Aplicada e Biometria, por colocar a disposição seus laboratórios e instalações.

À Coordenadoria de Aperfeiçoamento de Pessoal do Ensino Superior (CAPES) pelo apoio financeiro através da bolsa de estudos.

Ao Altino pela grande e valiosa ajuda sempre que precisei e a todos os meus amigos do mestrado que estiveram sempre presentes nos melhores (e nos piores) momentos dessa pesquisa.

Sumário

1. Introdução	1
1.1 Objetivos	2
2. Referencial Teórico	4
2.1 Análise de Agrupamento	4
2.1.1 Medidas de Dissimilaridade	4
2.1.2 Métodos de agrupamento	6
2.2 Análise discriminante	8
2.2.1 Discriminação entre populações	9
2.2.2 Discriminação e Classificação	10
2.2.3 Regiões de Alocação	10
2.2.4 Regras de Classificação	11
2.2.5 Análise discriminante de Fisher	12
2.2.6 Análise discriminante de Anderson	13

2.3	Redes Neurais Artificiais	16
2.3.1	Os neurônios biológicos	19
2.3.2	Os neurônios artificiais	19
2.3.3	Funções de ativação	21
2.3.4	Arquiteturas das redes neurais artificiais	24
2.3.5	Aprendizado	27
2.3.6	Regras de Aprendizado	29
2.4	Taxa de erro aparente	30
2.5	Simulação	31
3.	Materiais e Métodos	32
3.1	Obtenção e preparação dos dados	32
3.2	Análise discriminante	37
3.3	Redes Neurais Artificiais	38
4.	Resultados e Discussões	42
5.	Conclusão	53
6.	Referências Bibliográficas	53
A	Código fonte da rede neural	63

B	Descrição de algumas funções do <i>ntoll</i> do MATLAB	70
B.1	Funções de Inicialização	70
B.2	Funções de Transferência (Ativação)	71
B.3	Funções de Treinamento	72
B.4	Funções de Aprendizado de Pesos	72
B.5	Funções de Performance de Erro	73
B.6	Treinamento da Rede	73
B.7	Utilizando a Rede	74

Lista de Tabelas

3.1	Médias e desvios padrões (σ) das características agronômicas dos dez cultivares sorteados para simulação das populações	35
3.2	Definição dos cenários utilizados no estudo para aplicação das funções discriminantes e da rede neural	36
3.3	Critério para definição dos cenários utilizados no estudo com base nos diferentes graus de similaridade das populações	37
3.4	Estrutura dos arquivos apresentados à rede neural, com suas respectivas composições	39
3.5	Parâmetros utilizados na rede neural que apresentou melhor resultado em relação à taxa de erro aparente	40
4.1	Grupos de populações estabelecidos pelo método de Tocher com base na dissimilaridade expressa pela distância de Mahalanobis . . .	43
4.2	Composição dos cenários utilizados no estudo com base nos diferentes graus de similaridade das populações	43
4.3	Estimativa dos coeficientes das variáveis consideradas na função discriminante de Anderson para os três cenários estabelecidos, com probabilidade a priori $p_i = 0,333$	44

4.4	Porcentagem de classificação correta e incorreta para cada técnica de classificação referente às populações do cenário C_1	46
4.5	Porcentagem de classificação correta e incorreta para cada técnica de classificação referente às populações do cenário C_2	47
4.6	Porcentagem de classificação correta e incorreta para cada técnica de classificação referente às populações do cenário C_3	48
4.7	Estimativa dos coeficientes das variáveis consideradas na função discriminante de Fisher para os três cenários propostos com seus respectivos pontos médios ($m_{ii'}$)	49
4.8	Taxa de Erro Aparente (%) e Taxa de Rejeição (R) da rede neural (%) nos três cenários	50
4.9	Taxa de Erro Aparente (%) e Taxa de Rejeição (R) da rede neural (%) nos três cenários considerando o procedimento de validação cruzada	51

Lista de Figuras

2.1	Regiões de alocação para o caso de duas populações π_1 e π_2 linear (a esquerda) e quadrática (a direita)	11
2.2	Representação simplificada de um neurônio (Adaptado de Ferneda, 2006)	19
2.3	Modelo básico de um neurônio artificial (Adaptado de Haykin, 2001)	20
2.4	Gráfico da função limiar	22
2.5	Gráfico da função sigmóide	23
2.6	Gráfico da função sigmóide, com α tendendo ao infinito	23
2.7	Gráfico da função signum	24
2.8	Gráfico da função tangente hiperbólica	25
2.9	Exemplos de arquiteturas de RNA's	26
2.10	Aprendizado supervisionado, adaptado de Braga et al. 2007	28
2.11	Aprendizado não supervisionado	29
2.12	Fluxo de processamento do algoritmo backpropagation	30

3.1 Média dos 49 cultivares de milho em relação à variável Peso de Grãos - adaptado de Embrapa, 2008	33
4.1 Método de Agrupamento: Ligação Média Entre Grupos (UPGMA) . .	42

Resumo

PEREIRA, Tiago Martins, M.Sc., Universidade Federal de Viçosa, dezembro de 2009. **Discriminação de populações com diferentes graus de similaridade por redes neurais artificiais.** Orientador: Cosme Damião Cruz. Co-orientadores: Adair José Regazzi e José Ivo Ribeiro Júnior.

A correta classificação de indivíduos em grupos pré-estabelecidos tem se tornado de grande importância no melhoramento genético. As técnicas de estatística multivariada usualmente utilizadas nesse tipo de problema são as funções discriminantes de Fisher e as funções discriminantes de Anderson, que são usadas para alocar um indivíduo inicialmente desconhecido em uma das g populações ou grupos pré-definidos. Nas últimas décadas vêm surgindo um novo paradigma de computação, as redes neurais artificiais, que podem ser utilizadas para resolver diversos problemas da Estatística, como agrupamento de indivíduos similares, previsão de séries temporais e em especial, os problemas de classificação. O objetivo dessa pesquisa foi realizar um estudo comparativo entre as funções discriminantes de Fisher e de Anderson e as redes neurais artificiais quanto ao número de classificações erradas de indivíduos sabidamente pertencentes a diferentes populações, com distintos níveis de dissimilaridade. Essa dissimilaridade, medida pela distância de Mahalanobis, foi um conceito de fundamental importância na utilização das técnicas de discriminação, pois quantificou o quanto as populações eram divergentes. Quanto maior o valor observado para essa medida, menos similares foram as populações em análise. A obtenção dos dados foi feita através de simulação utilizando o programa computacional Genes (CRUZ, 2006). As re-

des neurais artificiais apresentaram uma taxa de indivíduos rejeitados por serem considerados ambíguos quanto às suas características discriminatórias. No entanto, mostraram-se uma técnica promissora no que diz respeito a problemas de classificação, uma vez que apresentaram um número de classificações erradas de indivíduos menor que aqueles dados pelas funções discriminantes.

Abstract

PEREIRA, Tiago Martins, M.Sc., Universidade Federal de Viçosa, December 2009. **Discrimination of populations different degrees of similarity in artificial neural networks.** Advisor: Cosme Damião Cruz. Co-advisors: Adair José Regazzi and José Ivo Ribeiro Junior.

The correct classification of individuals in pre-established groups has become of great importance in breeding. The multivariate statistical techniques commonly used in this type of problem are the discriminant functions of Fisher and Anderson, which are used to allocate an initially unknown individual in one of the g populations or pre-defined groups. In recent decades a new computing paradigm, artificial neural networks, has come along to solve various problems of Statistics, such as grouping of similar individuals, time series forecasting and also of particular interest, the problem of classification. The objective of this research was to conduct a simulation study in order to compare the discriminant functions of Fisher and Anderson and neural networks. We evaluated the number of incorrect classifications of individuals known to belong to different populations with different levels of dissimilarity measured by the Mahalanobis distance. Simulations were conducted using the software Genes (Cruz, 2006). Although Artificial Neural Networks presented a rate of incorrect classification of individuals rejected for being considered ambiguous as to its discriminatory characteristics, it proved to be a promising technique, since it presented a lower number of incorrect classifications of individuals when compared to the discriminant functions.

1. Introdução

Os estudos sobre discriminação de populações têm sido de grande importância para fins de melhoramento genético e para avaliar o impacto da atividade humana sobre a biodiversidade. São igualmente importantes no entendimento dos mecanismos micro e macro-evolutivos que atuam na diversificação das espécies, envolvendo estudos populacionais, bem como na otimização do processo de conservação da diversidade genética. Também são fundamentais no entendimento de como as populações naturais se estruturam no tempo e no espaço e quais são os efeitos das atividades antrópicas nesta estruturação e, por consequência, nas suas chances de sobrevivência e, ou, extinção. Essas informações constituem subsídio para constatar possíveis perdas genéticas geradas pelo isolamento das populações e dos indivíduos, o que se refletirá nas futuras gerações, permitindo o estabelecimento das melhores estratégias para incrementar e preservar a diversidade entre e dentro das espécies.

Análises discriminantes são de uso geral e têm sido utilizadas na maioria das situações em que se devem tomar decisões, pois geralmente são determinadas por nossa experiência passada. Assim, como exemplo mais geral, dizer, baseado em experiência, que um determinado genótipo, numa "primeira impressão", parece promissor, equivale a dizer: "a grande maioria dos genótipos com este comportamento, também parecem promissores". Assim, percebe-se que a incapacidade ou dificuldade de emitir qualquer opinião ou tomar qualquer decisão surge quando não há algum histórico, ainda que por aproximação ou compara-

ção, sobre a matéria a ser analisada. No caso específico do uso da análise discriminante de populações procura-se obter funções que permitam classificar um elemento amostral, a partir das informações de um conjunto de características mensuradas, em uma dentre várias populações conhecidas, buscando minimizar a probabilidade de má classificação (MINGOTI, 2005). Assim, devem-se obter funções que permitam alocar o elemento amostral na população a qual ele realmente pertence. Nestas análises os estudos são realizados preliminarmente a partir das informações de populações previamente conhecidas. Constatada a eficácia da discriminação, as funções podem ser utilizadas para alocar novos elementos, que se desconhecem a origem.

As Redes Neurais Artificiais (RNA's) são modelos de processamento de dados que emulam uma rede de neurônios biológicos, capazes de recuperar rapidamente uma grande quantidade de dados e reconhecer padrões baseados na experiência (HAYKIN, 2009). São sistemas que se adaptam utilizando uma abordagem de processamento distribuído, no qual os neurônios se comunicam por meio de uma rede de elos interconectados (BISHOP, 2007). O grande diferencial da computação por redes neurais é que o processamento das informações pelos neurônios artificiais pode ocorrer em paralelo, o que lhe confere uma capacidade de processar grande quantidade de informações de forma rápida. Como as RNA's têm a capacidade de aprender e guardar o conhecimento adquirido, elas têm sido utilizadas principalmente no reconhecimento de padrões, tanto para realizar classificações, quanto para fazer previsões.

1.1 Objetivos

Diante do exposto foi realizado este trabalho com a finalidade de utilizar e comparar a eficácia de diferentes técnicas em análise de discriminação de populações. Cada população, gerada por simulação, foi caracterizada por um conjunto de elementos mensurados por características de natureza contínua.

O objetivo deste trabalho foi estudar a viabilidade da discriminação de populações com diferentes graus de dissimilaridade, através da comparação das técnicas de discriminação por redes neurais artificiais com as técnicas de análise discriminante propostas por Anderson e por Fisher, a fim de minimizar a probabilidade de má classificação, ou seja, a probabilidade de classificar erroneamente um indivíduo em uma população π_i quando ele realmente pertence a uma população π_j , com $i \neq j$.

2. Referencial Teórico

2.1 Análise de Agrupamento

A análise de agrupamento é uma técnica estatística multivariada que tem como objetivo dividir os elementos de uma amostra ou população em grupos, de tal forma que os elementos pertencentes a um mesmo grupo sejam similares em relação a determinadas características. Os agrupamentos resultantes devem exibir alta homogeneidade interna (dentro dos grupos) e elevada heterogeneidade externa (entre os grupos). Para a formação dos grupos, é necessário um método de comparação simultânea sobre as características dos indivíduos.

2.1.1 Medidas de Dissimilaridade

Ao se aplicar as técnicas de agrupamento, faz-se necessário utilizar alguma medida de dissimilaridade entre os indivíduos a serem utilizados no processo de agrupamento. Dissimilaridade representa o grau de correspondência entre os indivíduos ao longo de todas as características usadas na análise. As principais medidas de dissimilaridade usadas na análise de agrupamento são dadas a seguir:

Distância Euclidiana

A distância euclidiana entre os indivíduos i e i' é dada por

$$d_{ii'} = \left[\sum_{j=1}^p (X_{ij} - X_{i'j})^2 \right]^{\frac{1}{2}} = [(\tilde{X}_i - \tilde{X}_{i'})^t (\tilde{X}_i - \tilde{X}_{i'})]^{\frac{1}{2}} \quad (2.1)$$

em que

$$\tilde{X}_i = \begin{bmatrix} X_{i1} & X_{i2} & \cdots & X_{ip} \end{bmatrix}^t$$

$$\tilde{X}_{i'} = \begin{bmatrix} X_{i'1} & X_{i'2} & \cdots & X_{i'p} \end{bmatrix}^t$$

Uma outra forma de expressar a distância euclidiana é por meio de sua média, dada por:

$$d_{ii'} = \left[\frac{1}{p} \sum_{j=1}^p (X_{ij} - X_{i'j})^2 \right]^{\frac{1}{2}} \quad (2.2)$$

sendo p o número de características mensuradas no indivíduo.

Distância de Mahalanobis

A distância de Mahalanobis (D^2) entre os indivíduos i e i' é dada pela raiz quadrada de:

$$D_{ii'}^2 = (\tilde{X}_i - \tilde{X}_{i'})^t S^{-1} (\tilde{X}_i - \tilde{X}_{i'}) \quad (2.3)$$

em que

S é a matriz de variâncias e covariâncias amostral comum a todos os indivíduos;

\widetilde{X}_i e $\widetilde{X}_{i'}$ são os vetores p -dimensionais de médias i e i' , com $i \neq i'$ e $i, i' = 1, 2, \dots, n$.

No caso de delineamentos experimentais, S é a matriz de variâncias e covariâncias residual. No texto, (D^2) será chamado simplesmente de distância de Mahalanobis.

2.1.2 Métodos de agrupamento

Existem inúmeros métodos de agrupamento, que se diferem pelo tipo de resultado fornecido e pelas diferentes formas de definir a distância entre um indivíduo e um grupo já formado.

Método de Tocher

O método de agrupamento otimizado de Tocher apresentado em Cruz & Carneiro (2006), constitui em um método de agrupamento simultâneo, onde a separação dos indivíduos em grupos se realiza de uma só vez. O método requer a obtenção da matriz de dissimilaridade, sobre a qual é identificado o par de indivíduos mais similares, que formarão o grupo inicial. A partir daí, é avaliada a possibilidade de inclusão de novos indivíduos, adotando-se o critério de que a distância média intragrupo deve ser menor que a distância média intergrupo (CRUZ et al. 2008).

Dessa forma, tem-se que a distância entre um indivíduo k e o grupo formado pelos indivíduos ij é dada por:

$$d_{(ij)k} = d_{ik} + d_{jk} \quad (2.4)$$

A inclusão, ou não, do indivíduo k no grupo é, então, feita considerando que o acréscimo médio promovido pela inclusão de um indivíduo k em um grupo previamente estabelecido seja menor que o valor máximo da medida de dissimilaridade encontrado no conjunto das menores distâncias envolvendo cada indivíduo, denotado por θ . Dessa forma,

$$\begin{aligned} \text{Se } \frac{d_{(grupo)k}}{n} \leq \theta, & \text{ inclui-se o indivíduo } k \text{ no grupo;} \\ \text{Se } \frac{d_{(grupo)k}}{n} > \theta, & \text{ o indivíduo } k \text{ não é incluído no grupo.} \end{aligned}$$

sendo n o número de indivíduos que constitui o grupo original.

Método da ligação média entre grupos

O método da ligação média não ponderada entre grupos (UPGMA), tem sido utilizado em diversos estudos no melhoramento genético. Este método trata-se de uma técnica de agrupamento que utiliza as médias aritméticas não ponderadas das medidas de dissimilaridade, evitando assim, caracterizar a dissimilaridade por valores extremos entre os indivíduos considerados.

Dessa forma, a construção do dendrograma é definida pelo indivíduo de maior similaridade. A distância entre um indivíduo k e um grupo, formado pelos indivíduos i e j , é dada por:

$$d_{(ij)k} = \frac{d_{ik} + d_{jk}}{2} \quad (2.5)$$

A distância entre dois grupos formados por indivíduos ij e klm é, então, dada por:

$$d_{(ij)(klm)} = \frac{d_{ik} + d_{il} + d_{im} + d_{jk} + d_{jl} + d_{jm}}{6} \quad (2.6)$$

ou seja, a distância entre dois grupos é dada pela média das distâncias entre elementos do conjunto.

No caso geral, a distância entre dois agrupamentos J e K é definida como:

$$d_{J,K} = \frac{1}{n_J \times n_K} \sum_{j \in J, k \in K} d_{jk} \quad (2.7)$$

em que n_J e n_K são o número de indivíduos dos grupos J e K , respectivamente.

2.2 Análise discriminante

A análise discriminante é uma técnica estatística que permite estudar diferenças entre dois ou mais grupos, em função de um conjunto de informações conhecidas para todos os elementos dos grupos. Segundo Khattree & Naik (2000) é uma técnica da estatística multivariada que estuda a separação de objetos de uma população em duas ou mais classes. A discriminação ou separação é a primeira etapa, sendo a parte exploratória da análise e consiste em se procurar características capazes de serem utilizadas para alocar objetos em diferentes grupos previamente definidos. A classificação ou alocação pode ser definida como um conjunto de regras que serão usadas para alocar novos objetos (JOHNSON & WICHERN, 1999). Contudo, a função que separa objetos pode também servir para alocar, e o inverso, regras que alocam objetos podem ser usadas para separar. Normalmente, discriminação e classificação se sobrepõem na

análise, e a distinção entre separação e alocação é confusa. Segundo Regazzi (2006) o problema da discriminação entre dois ou mais grupos, visando posterior classificação, foi inicialmente abordado por Fisher (1936). Consiste em obter funções matemáticas capazes de classificar um indivíduo \tilde{X} (uma observação \tilde{X}) em uma de várias populações π_i ($i=1, 2, \dots, g$), com base em medidas de um número p de características, buscando minimizar a probabilidade de má classificação, isto é, minimizar a probabilidade de classificar erroneamente um indivíduo em uma população π_i , quando realmente pertence a população π_j ($i \neq j$) para $i, j = 1, 2, \dots, g$.

2.2.1 Discriminação entre populações

Exemplos de emprego de funções discriminantes são encontrados em diversas áreas da ciência. No melhoramento genético cita-se, como ilustração o estudo realizado por Ferreira et al. (1995), cujo objetivo foi o de discriminar cultivares de arroz tolerantes à toxidez de alumínio, em condições de cultivo hidropônico. Para estabelecimento das funções que permitissem a discriminação entre cultivares tolerantes e sensíveis, foram utilizados genótipos cujo comportamento, em nível de cultivo de campo, já eram conhecidos. Desta forma, genótipos sabidamente tolerantes foram reunidos para formar a primeira população e os sabidamente sensíveis formaram a outra. Constatou ser possível, em condições de cultivo hidropônico, distinguir genótipos tolerantes e sensíveis, considerando os caracteres comprimento de raiz, pesos da matéria-seca da raiz, da parte aérea e total, e altura da planta, após dez dias de cultivo em solução nutritiva. Outro estudo baseado em função discriminante foi realizado por Pereira (1999), visando diferenciar entre cultivares de arroz considerados de padrão moderno e tradicional. As redes neurais artificiais foram usadas por Ramos (2003) para fins de classificação de frutos por meio de vetores de padrões, utilizando para isso uma rede perceptron. Nakano (1997) classificou maçãs utilizando redes neurais e, Simões et al. (2003) as utilizaram para classificar laranjas baseada em padrões visuais.

2.2.2 Discriminação e Classificação

Segundo Varella (2004), o problema de classificação de indivíduos consiste em se obter uma combinação linear de características observadas, denominada função discriminante, que apresente maior poder de discriminação entre populações. Tal função tem a propriedade de minimizar as probabilidades de má classificação, quando as populações são normalmente distribuídas com média μ e variância Σ conhecidas. Entretanto, tal situação não ocorre com frequência, isto é, a média e a variância das populações normalmente não são conhecidas, havendo a necessidade de estimação desses parâmetros. Pode-se assumir que as populações têm uma mesma matriz de covariâncias ou não. Conforme a pressuposição sobre a matriz de covariâncias, as funções discriminantes são denominadas de lineares ou quadráticas.

2.2.3 Regiões de Alocação

Regiões de alocação são conjuntos de valores separados por uma fronteira definida através de uma função discriminante qualquer, estimada a partir de amostras de treinamento, e são obtidas por meio de modelos estatísticos ou computacionais através de redes neurais (HAIR, 2009). Uma vez definidas tais regiões, uma observação pode ser alocada como sendo de uma população π_1 e ou de uma população π_2 . Contudo é importante observar que geralmente a fronteira entre regiões não está exatamente definida, podendo haver sobreposição, isto é, erro de classificação (VARELLA, 2004). A Figura 2.1 ilustra regiões de alocação para o caso de duas populações, para funções discriminantes lineares e quadráticas.

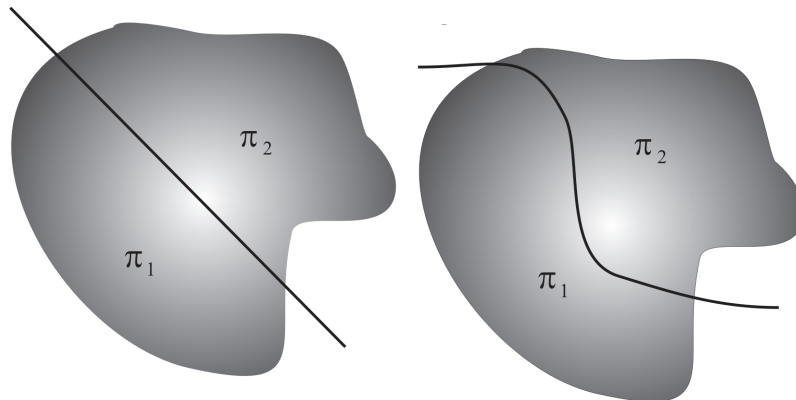


Figura 2.1: Regiões de alocação para o caso de duas populações π_1 e π_2 linear (a esquerda) e quadrática (a direita)

2.2.4 Regras de Classificação

Uma boa classificação deve resultar em pequenos erros, isto é, deve haver pouca probabilidade de má classificação. Segundo Johnson & Wichern (1999), para que isso ocorra, a regra de classificação deve considerar as probabilidades a priori e os custos de má classificação. Outro fator a ser considerado é se as variâncias das populações são iguais ou não. Quando a regra de classificação assume variâncias populacionais iguais, as funções discriminantes são ditas lineares, caso contrário, as funções discriminantes são quadráticas. Regras de classificação também podem ser construídas com base em modelos de redes neurais. Segundo Gonzalez & Woods (1992), citado por Khoury Jr. (2004), em comparação com classificadores estatísticos que determinam planos lineares ou quadráticos, o maior benefício da modelagem por redes neurais é sua capacidade de determinar planos não-lineares de separação de classes.

Os métodos de análise discriminante aplicam-se a populações munidas de uma partição definida a priori, descritas por diversas variáveis explicativas. O objetivo destes métodos é discriminar as classes da partição, através das características definidas pelas variáveis explicativas. Pretende-se, então, construir uma regra de decisão que permita, no futuro, alocar novos indivíduos, minimizando

os erros de alocação. Quando as variáveis explicativas são de natureza qualitativa, tem-se a análise discriminante discreta. Nas últimas décadas foram propostos um grande número de modelos nesta área, contudo, a resolução do principal problema em Análise Discriminante Discreta, o denominado problema da dimensionalidade, permaneceu em aberto em muitos casos, particularmente quando as amostras são de dimensão reduzida. Algumas técnicas de análise discriminante, comumente utilizadas na literatura, são apresentadas a seguir.

2.2.5 Análise discriminante de Fisher

Uma das soluções apresentadas ao problema de classificação de uma observação em g populações foi sugerida por Fisher (1936), através da utilização de uma função linear do vetor aleatório \tilde{X} , a qual se caracteriza por maximizar as razão entre as soma dos quadrados entre grupos e a soma de quadrados dentro dos grupos (MINGOTI, 2005). Segundo Mardia et al. (1979), as funções discriminantes de Fisher são de grande importância para o caso especial em que temos $g = 2$ grupos, por produzir separação máxima entre duas populações. Assim, define-se a função discriminante linear de Fisher pela expressão:

$$D_{ii'}(\tilde{X}) = \alpha^t \tilde{X} = (\mu_i - \mu_{i'})^t \Sigma^{-1} \tilde{X} \quad (2.8)$$

Se tomarmos $\alpha^t = (\mu_i - \mu_{i'})^t \Sigma^{-1}$, observa-se que α^t é um vetor linha, de dimensão $1 \times p$. Dessa forma, a função discriminante de Fisher pode ser escrita como sendo uma combinação linear das características originais do indivíduo, conforme equação 2.9.

$$D_{ii'}(\tilde{X}) = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_p X_p \quad (2.9)$$

Assim, a função discriminante de Fisher é uma combinação linear do conjunto de caracteres que possibilita alocar um indivíduo, com um vetor de observações em uma população i ou i' , com máxima probabilidade de acerto. Define-se o ponto médio entre duas populações como sendo:

$$m_{ii'} = \frac{1}{2} [D(\mu_i) + D(\mu_{i'})] \quad (2.10)$$

Logo, aloca-se um indivíduo com vetor de observações \tilde{X} em π_i se $D_{ii'}(\tilde{X}) \geq m_{ii'}$ e em π_j se $D_{ii'}(\tilde{X}) < m_{ii'}$

2.2.6 Análise discriminante de Anderson

Na análise discriminante proposta por Anderson (1958), consideram-se as informações de indivíduos sabidamente pertencentes a diferentes populações. A partir dessas informações são geradas funções que são combinações lineares das características avaliadas e têm por finalidade promover a melhor discriminação entre os indivíduos, alocando-os em suas devidas populações.

Sejam π_i ($i = 1, 2, \dots, g$) um grupo de g populações. Um ponto importante além de se assumir alguma distribuição para obter as funções discriminantes, é o estabelecimento das probabilidades a priori para as várias populações. Há casos em que a probabilidade de um indivíduo com um vetor de observações \tilde{X} , pertencer a uma determinada população já é conhecida pelo pesquisador a partir de sua experiência e do conhecimento que se tem sobre as populações pesquisadas e do indivíduo a ser classificado. Os classificadores são desenvolvidos da necessidade de se alocar um indivíduo com um vetor de observações \tilde{X} em uma dentre g populações, sendo $g \geq 2$. Para desenvolver um classificador é necessário fazer algumas pressuposições para o modelo da função discriminante, neste caso as pressuposições são as seguintes:

1. As g populações apresentam distribuição normal multivariada;
2. A soma das p_i probabilidades a priori de ocorrência das populações são iguais a $\sum_{i=1}^g p_i = 1$;
3. As populações apresentam custos de má classificação.

Sendo a probabilidade a priori de um indivíduo i pertencer a um população definida por p_i , pode-se estabelecer a função discriminante de Anderson da seguinte forma:

$$D_i(\tilde{X}) = -\frac{1}{2} [\ln(2\pi) + \ln|\Sigma_i|] - \frac{1}{2} \left[(\tilde{X} - \mu_i)^t \Sigma_i^{-1} (\tilde{X} - \mu_i) \right] + \ln(p_i) \quad (2.11)$$

Supondo a homogeneidade das matrizes de variâncias e covariâncias e retirando-se os componentes constantes da função $D_i(\tilde{X})$ que são dispensáveis na discriminação entre duas populações, tem-se:

$$D_i(\tilde{X}) = \ln(p_i) + \left(\tilde{X} - \frac{1}{2}\mu_i \right)^t \Sigma^{-1} \mu_i \quad (2.12)$$

que é a função discriminante de Anderson (1958), em que

$$\Sigma = \frac{(n_1 - 1)\Sigma_1 + (n_2 - 1)\Sigma_2 + \dots + (n_g - 1)\Sigma_g}{(n_1 - 1) + (n_2 - 1) + \dots + (n_g - 1)} \quad (2.13)$$

em que Σ_i é a matriz de variâncias e covariâncias entre as p variáveis do grupo g_i .

Esta técnica tem por finalidade classificar novos indivíduos, de comportamento desconhecido, em populações já conhecidas.

Assim, classifica-se uma observação \tilde{X} na população π_i se, e somente se

$$D_i(\tilde{X}) = \max [D_1(\tilde{X}), D_2(\tilde{X}), \dots, D_g(\tilde{X})] \quad (2.14)$$

Na prática, raramente se conhece os parâmetros μ_i e Σ . Os processos de obtenção de regras de discriminação continuam os mesmos utilizando-se as estimativas ao invés dos parâmetros.

Exemplos de aplicação de funções discriminantes podem ser encontrados em diversas áreas do conhecimento. Em diversidade genética, por exemplo, cita-se o estudo realizado por Assis et al. (2003) em que o objetivo era discriminar entre espécies de *Brachiaria* baseada em diferentes grupos de caracteres morfológicos. Neste estudo, foram analisados 301 acessos pertencentes a seis diferentes espécies de braquiária, nas quais foram avaliadas características vegetativas, reprodutivas e de pilosidade. Foram realizadas análises discriminantes de Anderson para cada um dos três grupos de caracteres morfológicos, sendo estabelecidas funções discriminantes para cada uma das seis espécies. As taxas de classificação correta para os caracteres vegetativos, reprodutivos e de pilosidade foram, respectivamente, as seguintes: *B. brizantha* (74,00; 69,33 e 30,67%), *B. decumbens* (63,04; 89,13 e 65,22%), *B. humidicola* (52,78; 69,44 e 52,78%), *B. jubata* (83,87; 77,42 e 64,52%), *B. ruziziensis* (89,29; 92,86 e 85,71%) e *B. dictyoneura* (80,00; 70,00 e 70,00%). Os caracteres vegetativos e reprodutivos mostraram ser os mais eficientes, enquanto os de pilosidade foram os menos eficientes na classificação e discriminação das espécies.

Outro estudo baseado em análise discriminante foi realizado por Moura Zanine et al. (2007), onde foi estudado o uso de funções discriminantes para comparação de cultivares dos gêneros *Cynodon* e *Digitaria* quanto à produção de matéria seca (MS) e teores de macronutrientes da parte aérea e sistema radicular, em 15 amostras de cada cultivar, coletadas no período das chuvas. A análise

discriminante mostrou que as variáveis teor de magnésio (Mg), teor de cálcio (Ca) e produção de matéria seca (MS) na parte aérea e teor de fósforo (P) e teor de nitrogênio (N) na raiz foram importantes no processo de discriminação dos cultivares estudados. Diversos outros trabalhos envolvendo o uso de funções discriminantes em diversidade genética e em outras áreas de conhecimento podem ser encontrados na literatura, o que demonstra o quão importante é esta técnica multivariada.

2.3 Redes Neurais Artificiais

O estudo de Redes Neurais tem uma história de aproximadamente cinco décadas, mas achou sólida aplicação somente nos últimos anos. O campo de aplicação está se desenvolvendo rapidamente, embora ainda haja grande carência de formalismo matemático para que se possa explicar consistentemente seu modo de operação, gerando desconfiança por parte dos especialistas quanto a sua confiabilidade. Conseqüentemente é uma área de estudo com muitos problemas abertos à pesquisa teórica.

Em 1943, McCulloch e Pitts propuseram um modelo para uma célula nervosa, chamado de neurônio formal ou neurônio artificial. Eles mostraram que uma coleção de neurônios era capaz de calcular certas funções lógicas.

Em 1949, Hebb apontou o significado das conexões entre sinapses, que são as conexões entre os neurônios, no processo de aprendizagem, e desenvolveu uma regra de aprendizagem básica. Ele propôs que as mudanças nas forças das sinapses fossem proporcionais às ativações dos mesmos.

Em 1959, Rosenblatt reunindo as idéias de Hebb, McCulloch e Pitts, descreveu o 1º modelo de Rede Neural, o Perceptron. Arranjando os neurônios em uma rede com uma topologia particular e modificando as conexões entre as sinapses, o Perceptron poderia aprender funções lógicas.

Em 1962, Widrow desenvolveu um tipo diferente de processador para Redes Neurais, denominado Adaline, o qual dispunha de uma poderosa estratégia de aprendizado. Em 1969 este direcionamento da pesquisa em Redes Neurais foi abandonado quase que completamente, por força do trabalho de Minsky e Papert que expuseram as limitações do Perceptron.

Em 1974, Werbos conseguiu o maior progresso em termos de Redes Neurais desde o Perceptron de Rosenblatt. Ele lançou as bases do algoritmo backpropagation, que permitiu que Redes Neurais com múltiplas camadas apresentassem capacidade de aprendizado.

Os primeiros resultados da retomada do desenvolvimento sobre Redes Neurais foram publicados em 1986 e 1987, onde ficou consagrada a técnica de treinamento por retropropagação.

Mangasarian et al. (1990) citam que se o objetivo é construir uma função discriminante $f : \mathbb{R}^n \rightarrow \mathbb{R}$, tal que $f(A) > 0$ e $f(B) < 0$. Quando as coberturas convexas de A e B não se interceptam, caso em que os conjuntos são linearmente separáveis, um único programa linear pode ser usado para obter uma função discriminante linear.

Mangasarian et al. (1990) têm estudado o problema de separação de padrões há aproximadamente três décadas. Vários trabalhos seus são apresentados com este objetivo: Via Programação Linear ou Programação não Linear (MANGASARIAN, 1965); via Programação Linear, com testes feitos em problemas reais pequenos (MANGASARIAN, 1968); via Programação Linear, gerando uma Superfície Linear por Partes, com aplicação no diagnóstico médico (MANGASARIAN et al. 1990), problema este que também pode ser abordado via Redes Neurais (BENNETT e MANGASARIAN, 1992); via Programação Linear, com a geração de apenas um plano (BENNETT e MANGASARIAN, 1992).

As redes neurais são compostas de muitos elementos simples, inspirados pelo sistema nervoso biológico, que operam em paralelo. A função da rede é

determinada pelas conexões entre os seus elementos. Pode-se treinar uma rede neural para executar uma função particular ajustando-se os valores das conexões entre os elementos.

As redes neurais têm sido treinadas para executar funções complexas em vários campos de aplicação, como: no diagnóstico médico (BENNETT e MANGASARIAN, 1990; STEINER et al. 1994); na predição de falência bancária (TAM et al. 1992), na aplicação à fabricação da pasta e papel industrial (FADUM, 1993), no controle do processo de produção do papel industrial (RUDD, 1991; STEINER et al. 1994), no mundo financeiro (CIPRA, 1992), no controle de processos químicos (NASCIMENTO et al. 1993), na obtenção de um modelo organizacional (ALMEIDA, 1995); para detectar fraudes com cartões (REBOUÇAS, 1993), em problemas de administração de empresas (ALMEIDA, 1995).

Tem sido comparada com outras técnicas de análise discriminante (PATUWO et al. 1993; TAM et al. 1993; BENNETT e MANGASARIAN, 1992; STEINER et al. 1994). E muito discutida por Zahedi (1991), Hammerstrom (1993), Barbosa (1989), Hilster (1989), Gorni (1993), Haykin (1994), Barbosa (1989) e inúmeros outros. Sharda (1994) aponta que no encontro da *Operations Research Society of America*, ORSA/1990 foram apresentados 10 artigos abordando Redes Neurais. Este número aumentou para 27 em 1991 e em 1992 para 45 trabalhos. Aponta também as inúmeras aplicações desta técnica na área de finanças, business, marketing, estatística, e outras.

A maior motivação para o estudo das Redes Neurais Artificiais (RNA's), comumente referenciadas apenas por Redes Neurais ou Redes Neuronais se dá na percepção de que o cérebro humano processa as informações de uma maneira completamente distinta dos computadores digitais convencionais. Segundo Haykin (2001), o cérebro é um computador altamente complexo, não-linear e paralelo. Ele tem a capacidade de organizar certos processamentos muito mais rapidamente do que o mais rápido computador digital hoje existente.

O cérebro humano contém uma enorme quantidade de neurônios, sua célula fundamental. Cada uma dessas células combinam-se de maneira contínua e paralela a fim de processar as várias informações que são obtidas por todos os nossos sentidos.

2.3.1 Os neurônios biológicos

Os neurônios são divididos em três partes, cada uma delas exercendo uma importante função para o processamento da informação: *o corpo da célula, os dendritos e o axônio*. Os dendritos têm a função de receber os impulsos nervosos ou a informação de outros neurônios e conduzi-los para o corpo da célula, onde a informação é processada e novos impulsos são gerados. O ponto de contato entre os neurônios é chamado de *sinapse*. A Figura 2.2 (FERNEDA, 2006) apresenta a representação simplificada de um neurônio biológico e suas partes.

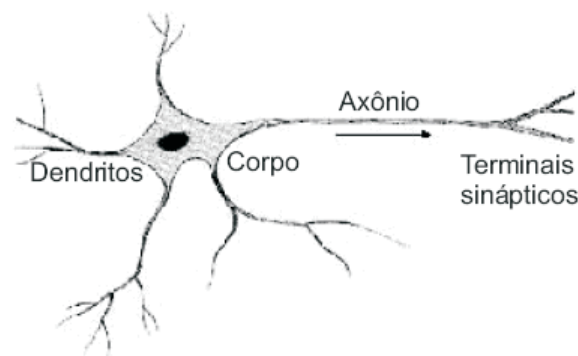


Figura 2.2: Representação simplificada de um neurônio (Adaptado de Ferneda, 2006)

2.3.2 Os neurônios artificiais

Um neurônio é uma unidade de processamento de informação que é de fundamental importância para o funcionamento de uma RNA. Basicamente, um

neurônio artificial é composto de três partes distintas: um conjunto de *sinapses* ou *elos de conexão*, cada uma delas com um peso ou força própria, um *somador* para somar os sinais de entrada e uma *função de ativação* para restringir a amplitude de saída de um neurônio. A Figura 2.3 mostra o modelo básico de um neurônio artificial adaptado por Haykin (2001).

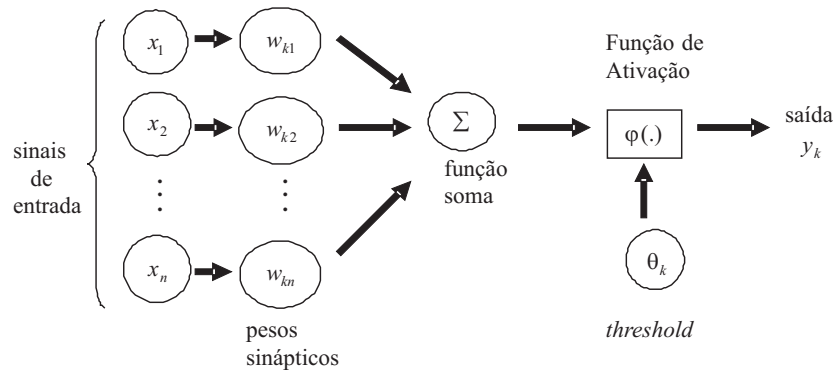


Figura 2.3: Modelo básico de um neurônio artificial (Adaptado de Haykin, 2001)

O primeiro modelo de neurônio artificial foi proposto por McCulloch e Pitts (1943) onde o neurônio era basicamente o que se entendia por um neurônio biológico. Matematicamente, era um modelo com n terminais x_1, x_2, \dots, x_n que representavam os dendritos e apenas um terminal de saída y , que representava o axônio, conforme pode ser observado na Figura 2.3. Para simular o comportamento da sinapse, os terminais de entrada possuem pesos acoplados w_1, w_2, \dots, w_n cujos valores dependem das sinapses inibidoras ou excitatórias. Um efeito particular de uma sinapse é dado por $x_i w_i$. Os pesos determinam o grau de consideração do disparo que ocorre naquela conexão.

Um neurônio biológico dispara quando a soma dos impulsos que ele recebe ultrapassa o seu limiar de excitação (*threshold*). No modelo de McCulloch e Pitts (MCP) a ativação do neurônio é obtida através de uma "função de ativação", que ativa ou não a saída, dependendo do valor das somas ponderadas de suas entradas (BRAGA et al. 2007).

Na Figura 2.3, x_j é o j -ésimo elemento do vetor de entrada, w_{kj} é peso

conectando a j -ésima entrada ao neurônio k , \sum é a função soma, $y_k = \varphi(\cdot)$ é a saída gerada pelo neurônio k , θ_k é o *threshold* aplicado ao neurônio k , para $j = 1, 2, \dots, n$.

O neurônio MCP terá sua saída ativa quando:

$$\sum_{i=1}^n x_i w_{ki} > \theta_k \quad (2.15)$$

Além dos elementos básicos já citados, o neurônio pode ainda apresentar uma entrada adicional, chamada de *bias*, que tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação (HAYKIN, 2001). O *bias* é um meio alternativo de representar o limiar (*threshold*) de um neurônio, ele é tomado fora do corpo do neurônio e conectado usando uma entrada adicional.

2.3.3 Funções de ativação

As funções de ativação fornecem o valor da saída de um neurônio. Elas limitam a saída do neurônio nos intervalos de $[0, 1]$ ou $[-1, 1]$. As principais funções de ativação são listadas a seguir:

Função Limiar (Degrau)

Utilizada no modelo de McCulloch e Pits, a função limiar modela a característica "tudo-ou-nada" deste neurônio. É expressa da seguinte forma:

$$f(v) = \begin{cases} 1, & \text{se } v \geq 0; \\ 0, & \text{se } v < 0. \end{cases} \quad (2.16)$$

Nos neurônios construídos com essa função, a saída y será igual a 0, caso o valor de ativação v seja negativo e 1 nos casos em que o valor de ativação seja positivo.

Deve-se lembrar que o valor de ativação (v) é composto pelo combinador linear e pelo bias:

$$v_j = \sum_{i=1}^n w_{ji}x_i + b \quad (2.17)$$

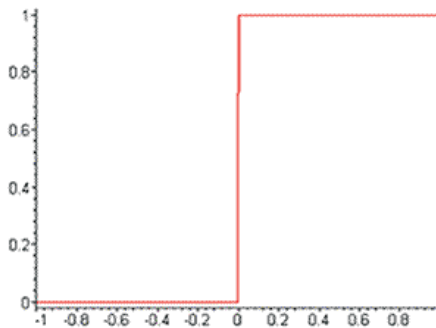


Figura 2.4: Gráfico da função limiar

Função Sigmóide

Esta função, ao contrário da função limiar, pode assumir todos os valores entre 0 e 1. A representação mais utilizada para esta função é a função logística, definida por:

$$f(v) = \frac{1}{1 + \exp(-\alpha v)} \quad (2.18)$$

onde α é o parâmetro de inclinação da função sigmóide e v é o valor de ativação do neurônio.

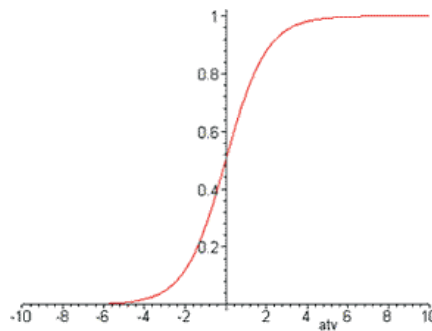


Figura 2.5: Gráfico da função sigmóide

Quando aumenta-se o valor do parâmetro α , tendendo-o ao infinito, esta função comporta-se como uma função de limiar, como podemos observar no gráfico da Figura 2.6.

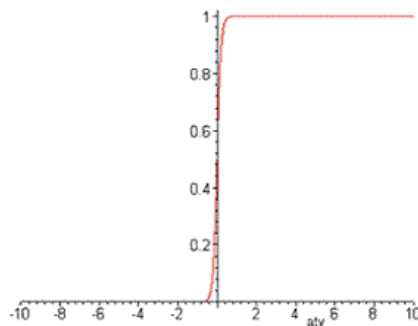


Figura 2.6: Gráfico da função sigmóide, com α tendendo ao infinito

Pode ser necessário, porém, que a função de ativação assuma valores entre 1 e -1. Esta característica traz benefícios analíticos. Para se obter tal intervalo de valores utiliza-se as Funções *Signum*, no caso da Função de Limiar, e a Função *Tangente Hiperbólica*, no caso da Função Sigmóide.

Função Signum

Esta função apresenta as mesmas características da função de limiar, porém está limitada entre 1 e -1. É representada por:

$$f(v) = b \frac{v}{|v|}, v \neq 0 \quad (2.19)$$

onde:

b são os limites inferiores e superiores ($b = |1|$ no gráfico);

v é o valor de ativação da unidade.

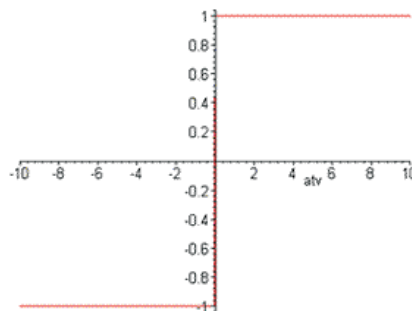


Figura 2.7: Gráfico da função signum

Função Tangente Hiperbólica

Como a Função Logística, também possui forma de "s", assumindo valores entre 1 e -1, sendo representada por:

$$f(x) = \frac{e^{\alpha x} - e^{-\alpha x}}{e^{\alpha x} + e^{-\alpha x}} \quad (2.20)$$

2.3.4 Arquiteturas das redes neurais artificiais

Um neurônio artificial atuando de forma individual possui um poder computacional muito restrito, independente da função de ativação escolhida. Dessa

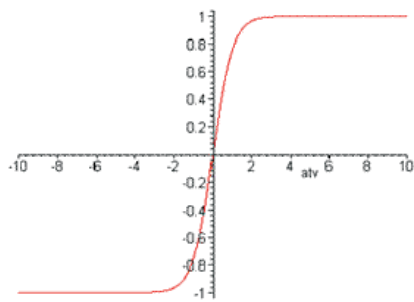


Figura 2.8: Gráfico da função tangente hiperbólica

forma, é necessário conectar vários deles com o intuito de aumentar o poder de processamento. Com isso forma-se uma rede (neural) com capacidade de resolver problemas de alta complexidade.

A forma como os neurônios são interligados varia de acordo com a natureza do problema a ser resolvido e é fator decisivo para a escolha dos algoritmos de aprendizado utilizados. As diversas arquiteturas de redes neurais existentes podem ser classificadas como feedforward e feedback.

Redes feedforward

As redes neurais do tipo feedforward consistem em uma ou mais camadas de unidades de processamento (neurônios) não-lineares. Cada neurônio de uma camada está conectado a todos os neurônios das camadas adjacentes e os neurônios estão dispostos em camadas conectadas por pesos unidirecionais na direção entrada-saída, ou seja, as conexões ocorrem apenas entre as camadas diferentes e subsequentes.

As primeiras redes de feedforward a aparecerem na literatura foram Perceptron (ROSENBLATT, 1962) e Adaline (WINDROW, 1957). No entanto, o potencial das redes feedforward só foi revelado com o surgimento das redes neurais multicamadas. As Figuras 2.9(a), 2.9(b) e 2.9(c) contêm exemplos de redes neurais *feedforward*.

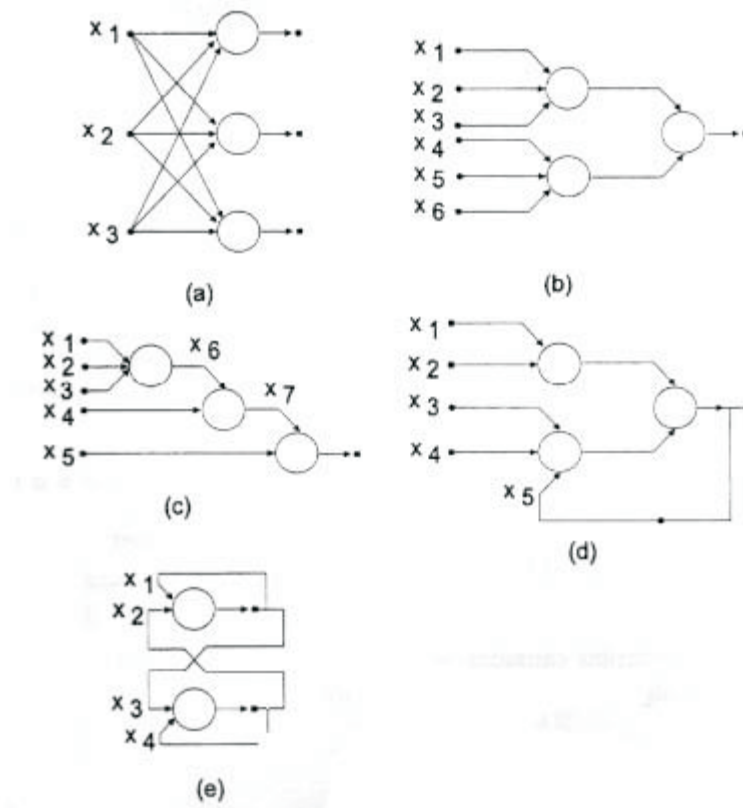


Figura 2.9: Exemplos de arquiteturas de RNA's

Redes feedback

Em redes feedback, a saída de uma unidade de processamento pode conectar-se a uma de suas próprias entradas ou a uma entrada de outra unidade da mesma camada, ao contrário das redes feedforward que não permitem esse tipo de conexão. Nessa classe de arquiteturas, encontram-se as redes neurais recorrentes. Um exemplo de rede neural feedback bastante simples e também muito utilizada é o modelo de Hopfield (HOPFIELD, 1982). As Figuras 2.9(d) e 2.9(e) contém exemplos de redes feedback.

2.3.5 Aprendizado

A principal característica de uma rede neural artificial é sua capacidade de aprender com exemplos. Aprender, nesse contexto, é o processo pelo qual uma rede neural adquire a habilidade de responder corretamente a estímulos, através do ajuste de parâmetros internos de acordo com algum esquema de aprendizado. O aprendizado ocorre quando a rede neural atinge uma solução generalizada para uma classe de problemas.

Segundo Braga et al. (2007), o conceito de aprendizado está relacionado à melhoria do desempenho da rede segundo algum critério preestabelecido. O erro quadrático médio da resposta da rede em relação ao conjunto de dados fornecido pelo ambiente, por exemplo, é utilizado como critério de desempenho pelos algoritmos de correção de erros.

O processo de aprendizado se dá com a atualização dos pesos sinápticos a cada iteração da rede, conforme a equação 2.21.

$$w(t + 1) = w(t) + \Delta w(t) \quad (2.21)$$

onde $w(t)$ e $w(t+1)$ representam os valores dos pesos nos instantes t e $t+1$, respectivamente, e $\Delta w(t)$ é o ajuste aplicado ao peso a cada iteração.

O aprendizado em redes neurais pode ser classificado, de acordo com presença ou ausência de realimentação explícita do mundo exterior, em supervisionado e não supervisionado (BARRETO, 2002). Os dois métodos são descritos a seguir:

Aprendizado supervisionado

A aprendizagem supervisionada (aprendizagem com professor) é caracterizada pela utilização de um agente externo que indica à rede a resposta desejada para o padrão de entrada. A rede neural é treinada através da apresentação de pares de entradas e saídas. Para cada entrada a rede produz uma resposta na saída, que é comparada com a resposta desejada. Através da análise de erros, realiza-se o ajuste dos pesos sinápticos. Isto é normalmente utilizado para redes feedforward. O aprendizado supervisionado se aplica à classes de problemas em que se deseja mapear padrões de entrada e saída, como no caso de problemas de classificação. Os algoritmos mais conhecidos para aprendizado supervisionado são a regra delta (WINDROW, 1960) e sua generalização para redes múltiplas camadas, o algoritmo backpropagation (RUMELHART, 1986). A Figura 2.10 (BRAGA et al. 2007) mostra o esquema do aprendizado supervisionado.

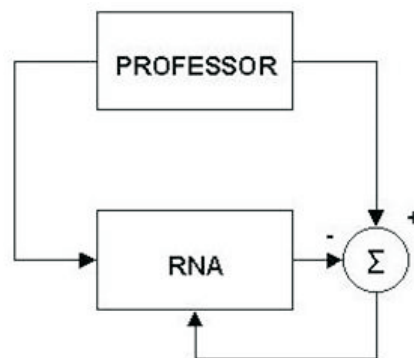


Figura 2.10: Aprendizado supervisionado, adaptado de Braga et al. 2007

Aprendizado não supervisionado

Nesse tipo de aprendizagem, ao contrário da supervisionada, não existe o agente externo (professor ou supervisor) para acompanhar o processo de aprendizado. A rede é treinada através de excitações ou padrões de entrada para, arbitrariamente, organizar os padrões em categorias. Para uma entrada aplicada

à rede, será fornecida uma resposta indicando a classe a qual a entrada pertence. Se o padrão de entrada não corresponde às classes existentes, uma nova classe é gerada. Dessa forma, o aprendizado não supervisionado se aplica à classes de problemas em que se deseja formar grupos de populações semelhantes.



Figura 2.11: Aprendizado não supervisionado

2.3.6 Regras de Aprendizado

Aprender, no contexto de redes neurais, significa ajustar valores para as conexões sinápticas. O algoritmo de aprendizagem ou algoritmo de treinamento mais popular é o algoritmo backpropagation.

Algoritmo backpropagation

É o algoritmo de aprendizado supervisionado mais popular para treinamento de redes perceptron de múltiplas camadas (MLP). Utiliza pares de entrada e saída para, por meio de correção de erros, ajustar os pesos da rede. Durante o treinamento com o algoritmo backpropagation, a rede opera em uma sequência de dois passos, conforme mostrado na Figura 2.12.

Primeiro, um padrão é apresentado à camada de entrada da rede. A resposta de uma unidade é propagada como entrada para as unidades na camada seguinte, até a camada de saída, onde é obtida a resposta da rede e o erro é calculado. No segundo passo, o erro é propagado a partir da camada de saída até a camada de entrada, e os pesos das conexões das unidades das camadas internas vão sendo modificados conforme o erro é retropropagado.

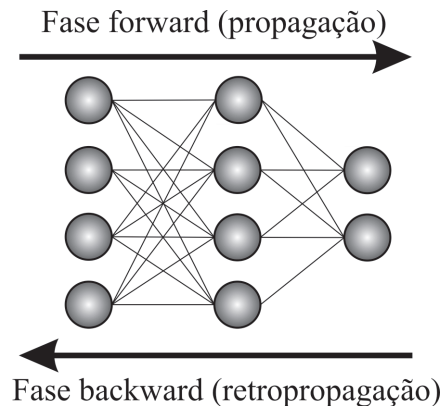


Figura 2.12: Fluxo de processamento do algoritmo backpropagation

2.4 Taxa de erro aparente

Uma vez obtidas as funções discriminantes foi avaliada a sua eficácia, que é dependente do grau de dissimilaridade entre as populações analisadas e, principalmente, da quantidade e qualidade das variáveis consideradas na discriminação. Como funções discriminantes são obtidas a partir de análises prévias de observações que se supõe serem, de fato, pertencentes às populações consideradas, pode-se calcular a probabilidade de má classificação, reclassificando toda observação até então disponível. A classificação de uma observação pertencente a uma população π_j em outra é indicativo de menor eficiência da função discriminante estimada, contribuindo para o acréscimo na taxa de erro aparente.

Um método alternativo para estimar a taxa de erro, neste estudo, foi a validação cruzada. Neste método omite-se o primeiro indivíduo da análise e gera-se a função discriminante usando os $(\sum_{j=1}^p n_j - 1)$ indivíduos restantes. Classifica-se então o indivíduo omitido. Este procedimento é repetido para todos os indivíduos $(\sum_{j=1}^p n_j)$. Finalmente, o número de classificações erradas para cada população é contabilizado e a taxa de erro de cada população é computada. A taxa de erro aparente (global) pode ser obtida com as médias ponderadas destas proporções, considerando as probabilidades a priori como pesos. Com este procedimento tem-se um menor viés.

2.5 Simulação

Uma das grandes contribuições da estatística é viabilizar o estudo de fenômenos, por meio da simulação de uma situação complexa, em que são estabelecidos parâmetros e restrições, de forma que os efeitos de certos fatores controláveis possam ser convenientemente estudados. A simulação tem sido definida como a maneira de imitar, por meio de recursos computacionais, o comportamento de um sistema real, para estudar seu funcionamento em condições alternativas (DACHS, 1988), envolvendo certos tipos de modelos lógicos, que permitam descrever, da melhor forma possível, o sistema natural (NAYLOR, 1971). Quando se faz estudos baseados em simulação é necessário considerar que nenhuma parte do universo é tão simples que possa ser compreendida e controlada sem abstração. A abstração consiste em substituir a parte do universo por um modelo semelhante, porém com estrutura mais simples. Portanto, outro aspecto importante na simulação é a modelagem. O modelo deve ser suficientemente simples para ser operacionalizado e interpretado adequadamente, mas seu desempenho deve ser comparável com o modelo real e, se a defasagem for grande, ele deve ser eliminado ou refinado. McNitt (1985) lembra que a simulação envolve modelos que representam a entidade a ser investigada, porém é mais que um simples modelo, é uma metodologia para avaliação destes. A simulação tem sido de grande utilidade em estudos genéticos sob vários contextos, incluindo estudos de populações, do indivíduo ou do próprio genoma.

3. Materiais e Métodos

3.1 Obtenção e preparação dos dados

Neste estudo foram utilizadas informações de 10 populações π_j , $j = 1, 2, \dots, p = 10$, constituídas por indivíduos ($n_j = 500$ para todo j) mensurados em relação a $v = 10$ características quantitativas, contínuas, com distribuição normal, tendo média e variância previamente conhecidas. Para cada população simulada, considera-se igualdade das matrizes de variância e covariâncias Σ dada pela equação 2.13, uma vez que sem essa pressuposição perde-se a linearidade das funções discriminantes. Todas as populações foram simuladas com o uso do aplicativo computacional Genes (CRUZ, 2006).

Para a simulação das populações referentes a esse estudo, foram utilizados os dados do Ensaio Nacional de Cultivares de Milho, Centro - 2007/2008 Média Minas Gerais, que podem ser encontrados em Embrapa (2008).

Os ensaios Nacionais de milho são coordenados pela Embrapa Milho e Sorgo e sua condução no campo é feita cooperativamente, pelos melhoristas e técnicos da cultura de milho no Brasil. No ano agrícola 2007/2008, os Ensaios Nacionais foram conduzidos nas principais regiões produtoras de milho, nos seguintes estados: Goiás, Mato Grosso do Sul, Minas Gerais, Pará, Paraná e São Paulo. Na rede de ensaios Centro foram avaliadas 49 cultivares, conforme Figura 3.1 e os ensaios foram instalados em 28 municípios, sendo o delineamento expe-

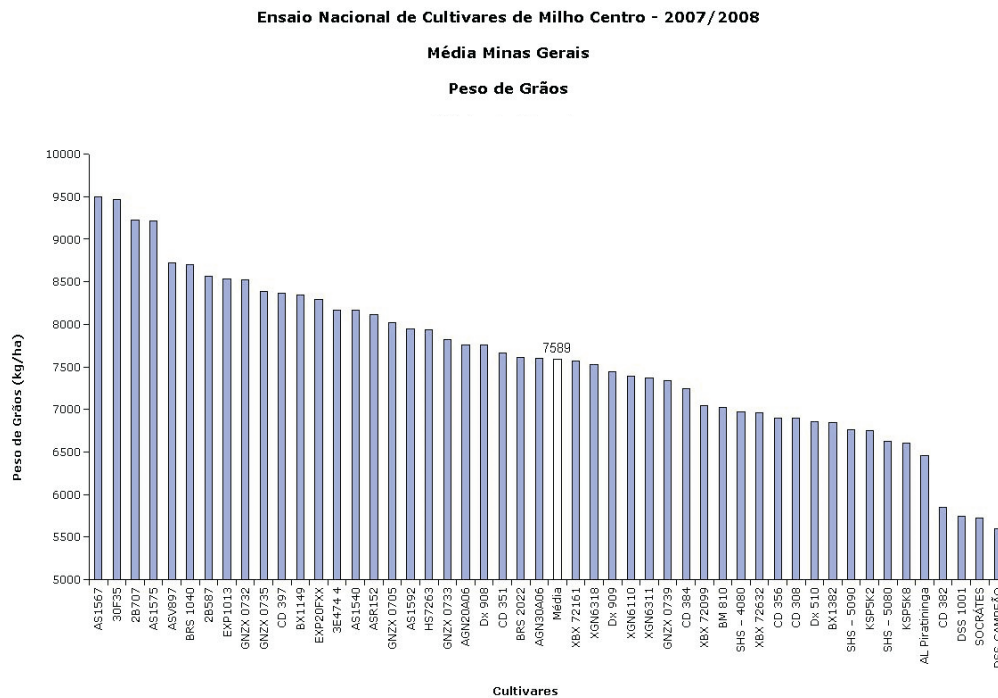


Figura 3.1: Média dos 49 cultivares de milho em relação à variável Peso de Grãos - adaptado de Embrapa, 2008

rimental látice 7x7. Todos os ensaios tiveram duas repetições, as parcelas foram constituídas em duas fileiras, com espaçamento entre linhas e o estande final foi de aproximadamente 52 plantas por parcela.

As características agronômicas de cada cultivar de milho apresentadas pelo Ensaio Nacional foram:

- x_1 : peso de grãos (kg/ha)
- x_2 : florescimento masculino (dias)
- x_3 : altura de planta (cm)
- x_4 : altura de espiga (cm)
- x_5 : plantas acamadas e quebradas (%)
- x_6 : estande final
- x_7 : número de espigas
- x_8 : espigas doentes (%)
- x_9 : umidade dos grãos (%)
- x_{10} : prolificidade

Foram sorteados 10 cultivares de milho dos 49 encontrados no Ensaio Nacional, utilizando o software estatístico R (R Development Core Team, 2006). Com esses 10 cultivares sorteados, foram simuladas as 10 populações que foram utilizadas nesse trabalho, utilizando o software computacional Genes (CRUZ, 2006).

Os valores médios para as dez características agronômicas dos dez cultivares de milho são dados na Tabela 3.1.

Tabela 3.1: Médias e desvios padrões (σ) das características agrônômicas dos dez cultivares sorteados para simulação das populações

População	Cultivar	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1	CD397	8731	64	247	137	4	56	48	13	15	0,8571
2	2B587	8569	65	223	121	6	55	52	2	16	0,9455
3	DSS-CAMPEÃO	5599	66	256	145	7	52	47	2	16	0,9038
4	AL Piratininga	6458	68	261	155	7	54	48	3	16	0,8888
5	AS1592	7946	66	246	140	3	51	48	4	16	0,9411
6	GNZX0732	8518	68	260	148	4	56	54	3	16	0,9642
7	2B707	9225	66	237	125	2	56	50	11	17	0,8929
8	AGN20A06	7760	65	231	126	4	55	47	8	16	0,8545
9	XGN6110	7393	65	229	129	7	53	49	0	17	0,9245
10	CD356	6899	66	248	136	7	52	46	4	17	0,8846
σ		1534,40	7,67	23,93	16,37	2,74	3,61	7,06	1,16	1,34	0,14

¹ x_1 : peso de grãos (kg/ha), x_2 : florescimento masculino (dias), x_3 : altura de planta (cm), x_4 : altura de espiga (cm),

x_5 : plantas acamadas e quebradas (%), x_6 : estande final, x_7 : número de espigas, x_8 : espigas doentes (%),

x_9 : umidade dos grãos (%), x_{10} : prolificidade

A partir dos dados apresentados na Tabela 3.1, foram simulados 500 indivíduos de cada cultivar para compor as 10 populações.

Definidas as populações procurou-se construir cenários constituídos de três populações cada, conforme Tabela 3.2, diferidos pelo grau de dissimilaridade entre as populações. Para construção de cada cenário, foram considerados os métodos de agrupamento de Tocher e UPGMA cuja matriz de distâncias utilizada, baseada na distância de Mahalanobis (D^2), é dada abaixo:

$$\begin{pmatrix} 0 & 98,51 & 101,95 & 83,84 & 66,96 & 81,09 & 7,10 & 21,48 & 120,19 & 69,67 \\ 98,51 & 0 & 9,62 & 10,77 & 8,69 & 7,40 & 66,30 & 29,67 & 3,25 & 8,57 \\ 101,95 & 9,62 & 0 & 1,92 & 7,99 & 8,11 & 76,05 & 34,85 & 5,43 & 4,82 \\ 83,84 & 10,77 & 1,92 & 0 & 5,96 & 4,57 & 61,54 & 26,63 & 9,25 & 3,47 \\ 66,96 & 8,69 & 7,99 & 5,96 & 0 & 4,25 & 42,78 & 15,48 & 11,40 & 3,41 \\ 81,09 & 7,40 & 8,11 & 4,57 & 4,25 & 0 & 54,37 & 24,69 & 10,09 & 7,11 \\ 7,10 & 66,30 & 76,05 & 61,54 & 42,78 & 54,37 & 0 & 9,24 & 84,61 & 45,89 \\ 21,48 & 29,67 & 34,85 & 26,63 & 15,48 & 24,69 & 9,24 & 0 & 41,76 & 16,49 \\ 120,19 & 3,25 & 5,43 & 9,25 & 11,40 & 10,09 & 84,61 & 41,76 & 0 & 8,71 \\ 69,67 & 8,57 & 4,82 & 3,47 & 3,41 & 7,11 & 45,89 & 16,49 & 8,71 & 0 \end{pmatrix}$$

Tabela 3.2: Definição dos cenários utilizados no estudo para aplicação das funções discriminantes e da rede neural

Cenário	Populações		
C_1	P_{11}	P_{21}	P_{31}
C_2	P_{12}	P_{22}	P_{32}
C_3	P_{13}	P_{23}	P_{33}

Baseados nas informações dos agrupamentos propostos pelos métodos de Tocher e UPGMA, pode-se construir os cenários utilizados neste estudo,

seguindo o critério apresentado pela Tabela 3.3.

Tabela 3.3: Critério para definição dos cenários utilizados no estudo com base nos diferentes graus de similaridade das populações

Cenário	Critério
C_1	Três populações menos divergentes
C_2	Três populações com divergência intermediária
C_3	Três populações mais divergentes

3.2 Análise discriminante

As funções discriminantes de Anderson e de Fisher utilizadas para classificar novos indivíduos desconhecidos em alguma das três populações definidas em cada um dos três cenários estabelecidos foram estimadas usando o software computacional Genes (CRUZ, 2006), com um arquivo de entrada contendo 1050 indivíduos (350 indivíduos de cada população).

Conhecidas as funções discriminantes, utilizou-se a taxa de erro aparente para avaliar a sua eficácia quanto a classificação de novos indivíduos em populações previamente conhecidas. Essa taxa também foi utilizada para medir a eficiência da classificação utilizando redes neurais. A taxa de erro aparente é dada em função da probabilidade de má classificação, que para cada população, é dada por

$$p_j = \frac{m_j}{n_j}, \quad (j = 1, 2, \dots, 10) \quad (3.22)$$

onde m_j é o número de observações retiradas da população π_j , que foram, por

meio da técnica avaliada, classificadas em outra população $\pi_{j'}$, sendo $j \neq j'$ e n_j é o número de indivíduos da população j . Assim, temos que

$$TEA = \frac{1}{N} \sum_{j=1}^g m_j \quad (3.23)$$

em que N é o número total de observações avaliadas e g é o número de populações consideradas.

3.3 Redes Neurais Artificiais

As redes neurais utilizadas nesse estudo foram simuladas utilizando o software MATLAB R2008a, através do seu toolkit nntoll que contém funções específicas e pacotes matemáticos para elaboração de redes neurais. Foram testadas diversas arquiteturas, a fim de encontrar aquela que minimizasse o número de classificações erradas, definida neste trabalho pela taxa de erro aparente.

O treinamento de redes feedforward é do tipo supervisionado. Neste tipo de treinamento é preciso possuir um conjunto de dados para treinamento, ou seja, uma série de pares de entradas e saídas desejadas. As entradas são apresentadas à rede e seus pesos são alterados de modo que a saída se aproxime da saída desejada. Pode-se dizer que a rede neural aprende a fazer seu trabalho observando uma série de exemplos que lhe são exibidos. Para alterar os pesos de forma adequada é necessária uma regra. A regra de treinamento mais utilizada para o treinamento de redes neurais feedforward é a Error Backpropagation (retropropagação de erros). A idéia deste algoritmo é atualizar os pesos utilizando as derivadas dos erros em relação aos pesos.

Neste trabalho, as redes neurais utilizadas foram as redes perceptron múltiplas camadas, com uma camada escondida e o número de neurônios em

cada camada variando em cada cenário. O algoritmo de treinamento usado foi o Backpropagation (Retropropagação), que minimiza o erro gerado na comparação entre a saída da rede e a saída desejada.

O conjunto de dados foi dividido em três partes, sendo 70% dos dados utilizado no conjunto de treinamento da rede, 20% usados no conjunto de validação da rede e os 10% restantes, usados para teste. A estrutura de cada arquivo utilizado pela rede neural é apresentada na Tabela 3.4.

As funções de ativação utilizadas foram a linear para as camadas de entrada e de saída e a função tangente hiperbólica para a camada escondida. Essa escolha foi feita para preservar a não-linearidade da rede.

Tabela 3.4: Estrutura dos arquivos apresentados à rede neural, com suas respectivas composições

Arquivo	Número de indivíduos	Composição
Treinamento	1050	(350 indivíduos de cada população)
Validação	300	(100 indivíduos de cada população)
Teste	150	(50 indivíduos de cada população)

O número de ciclos de treinamento foi fixado em 2000 épocas, para o caso em que as populações eram muito divergentes (cenário 3) e em 7000 épocas para os demais, devido ao alto grau de similaridade entre as populações. Teve-se o cuidado de limitar o número de iterações, para que esse não se tornasse excessivo, o que poderia levar à perda do poder de generalização. Por outro lado, esse número não pode ser muito pequeno, evitando assim, que a rede não atinja o seu melhor desempenho.

Definidos os parâmetros das redes, a configuração daquela que foi utilizada pelo presente estudo é apresentada na Tabela 3.5.

Tabela 3.5: Parâmetros utilizados na rede neural que apresentou melhor resultado em relação à taxa de erro aparente

Parâmetro	Descrição
Número de camadas	3 (entrada, camada escondida, saída)
Número de Neurônios por camada	10 (entrada), 6 (escondida), 3 (saída)
Função de ativação	<i>purelin</i> (entrada e saída) e <i>tansig</i> (escondida)
Função de treinamento	<i>traingdx</i> - Backpropagation
Função de performance	<i>mse</i> - Erro Quadrático Médio
Número de épocas de treinamento	2000 (cenário 3) e 7000 (cenários 1 e 2)
Número de dados para treinamento	1050
Número de dados para validação	300
Número de dados para teste	100

Geralmente, em sistemas de reconhecimento de padrões, a decisão de aceitar ou não determinado padrão depende do nível de contribuição que aquele padrão exerce no poder de separação do indivíduo. Em problemas de classificação, um padrão (indivíduo) é considerado ambíguo, se ele não puder ser associado a uma determinada classe ou população com determinada certeza. Dessa forma, sentiu-se a necessidade de se obter uma nova taxa para representar o percentual de indivíduos cuja classificação foi rejeitada.

4. Resultados e Discussões

Definidas as populações, utilizou-se os métodos de agrupamento de Tocher e UPGMA, a fim de definir os cenários propostos diferidos pelo grau de dissimilaridade entre as populações.

A correspondente representação gráfica dos agrupamentos propostos pelo método UPGMA (dendrograma) é mostrada na Figura 4.1. Um corte feito à distância 11,84, possibilita estabelecer três grupos hierárquicos: o primeiro constituído pela população 8, o segundo pelas populações 1 e 7 e o terceiro pelas demais populações.

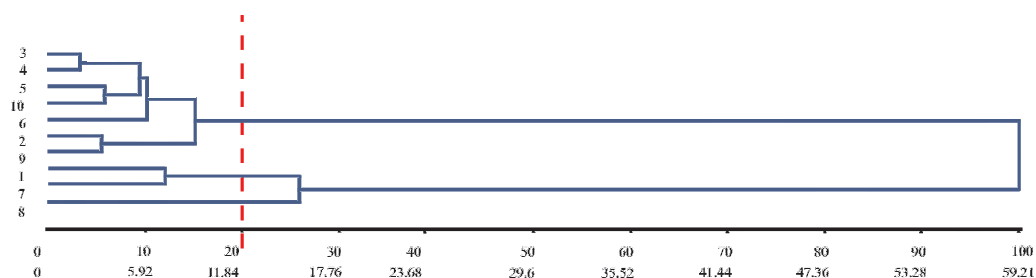


Figura 4.1: Método de Agrupamento: Ligação Média Entre Grupos (UPGMA)

Os agrupamentos estabelecidos pelo método de Tocher são apresentados na Tabela 4.1.

Dessa forma, os cenários foram estabelecidos conforme a Tabela 4.2.

Os coeficientes estimados das funções discriminantes de Anderson de cada variável para os três cenários estabelecidos são apresentados na Tabela 4.3

Tabela 4.1: Grupos de populações estabelecidos pelo método de Tocher com base na dissimilaridade expressa pela distância de Mahalanobis

Grupos	Populações
1	3 4 10 5 6 9 2
2	1 7
3	8

Tabela 4.2: Composição dos cenários utilizados no estudo com base nos diferentes graus de similaridade das populações

Cenário	Populações
C_1	4 5 10
C_2	1 7 8
C_3	1 8 9

Tabela 4.3: Estimativa dos coeficientes das variáveis consideradas na função discriminante de Anderson para os três cenários estabelecidos, com probabilidade a priori $p_i = 0,333$

Variável	Cenário C_1			Cenário C_2			Cenário C_3		
	4	5	10	1	7	8	1	8	9
Constante	-281,746	-303,347	-290,092	-347,599	-355,738	-337,253	-382,353	-371,997	-378,342
x_1	0,002	0,002	0,001	0,001	0,001	0,001	0,002	0,002	0,002
x_2	0,942	0,956	0,937	1,056	1,074	1,063	1,239	1,248	1,240
x_3	0,399	0,415	0,406	0,451	0,445	0,445	0,446	0,432	0,444
x_4	0,313	0,321	0,303	0,535	0,514	0,524	0,480	0,452	0,475
x_5	0,380	0,541	0,733	1,689	1,531	1,654	0,907	0,714	0,901
x_6	2,265	2,467	2,349	4,095	4,121	4,084	4,050	4,070	4,044
x_7	0,466	0,532	0,463	0,757	0,774	0,743	0,944	0,919	0,941
x_8	5,800	5,392	5,741	0,504	0,227	-0,328	0,750	0,551	0,294
x_9	9,118	9,170	9,443	7,701	8,279	7,855	10,345	10,429	10,395
x_{10}	33,896	34,680	32,861	43,595	44,759	43,950	51,321	49,841	51,473

¹ x_1 : peso de grãos (kg/ha), x_2 : florescimento masculino (dias), x_3 : altura de planta (cm), x_4 : altura de espiga (cm),

x_5 : plantas acamadas e quebradas (%), x_6 : estande final, x_7 : número de espigas, x_8 : espigas doentes (%),

x_9 : umidade dos grãos (%), x_{10} : prolificidade

As funções discriminantes de Fisher foram obtidas para cada par de populações dentro de cada cenário e as estimativas dos coeficientes de cada característica dos indivíduos dentro de cada cenário são apresentados na Tabela 4.7.

Conhecidas as funções discriminantes de Anderson e Fisher e obtida a rede neural utilizada no processo de classificação, a comparação das técnicas se deu através da taxa de erro aparente. As Tabelas 4.4, 4.5 e 4.6 indicam a porcentagem de classificação correta e incorreta em cada grupo para cada técnica de classificação para os diferentes cenários propostos.

Tabela 4.4: Porcentagem de classificação correta e incorreta para cada técnica de classificação referente às populações do cenário C_1

População	FDA			FDF			RNA			
	4	5	10	4	5	10	4	5	10	R
4	75,43	11,71	12,86	75,43	11,71	12,86	69,42	10,85	7,73	12,00
5	14,86	78,57	6,57	14,86	78,57	6,57	6,85	79,73	2,28	11,14
10	18,57	4,57	76,86	18,57	4,57	76,86	8,85	3,14	76,00	12,00

¹FDA: Função Discriminante de Anderson, FDF: Função Discriminante de Fisher, RNA: Rede Neural Artificial,

R: Porcentagem de indivíduos rejeitados pela rede neural

Tabela 4.5: Porcentagem de classificação correta e incorreta para cada técnica de classificação referente às populações do cenário C_2

População	FDA			FDF			RNA		
	1	7	8	1	7	8	1	7	8
1	86,57	13,14	0,29	86,57	13,14	0,29	85,43	7,71	0,29
7	9,71	83,43	6,86	9,71	83,43	6,86	4,85	86,57	2,29
8	0,57	9,14	90,29	0,57	9,14	90,29	0	1,71	98,29
									R
									6,57
									6,29
									0

¹FDA: Função Discriminante de Anderson, FDF: Função Discriminante de Fisher, RNA: Rede Neural Artificial,

R: Porcentagem de indivíduos rejeitados pela rede neural

Tabela 4.6: Porcentagem de classificação correta e incorreta para cada técnica de classificação referente às populações do cenário C_3

População	FDA			FDF			RNA		
	1	8	9	1	8	9	1	8	9
1	84,86	15,14	0	84,86	15,14	0	99,15	0,85	0
8	13,71	84,58	1,71	13,71	84,58	1,71	0,29	99,71	0
9	0	4,29	95,71	0	4,29	95,71	0	0	100

¹FDA: Função Discriminante de Anderson, FDF: Função Discriminante de Fisher, RNA: Rede Neural Artificial,

R: Porcentagem de indivíduos rejeitados pela rede neural

Tabela 4.7: Estimativa dos coeficientes das variáveis consideradas na função discriminante de Fisher para os três cenários propostos com seus respectivos pontos médios (m_{it})

Variável	Cenário C_1					Cenário C_2			Cenário C_3		
	4 e 5	4 e 10	5 e 10	1 e 7	1 e 8	7 e 8	1 e 8	1 e 8	1 e 8	1 e 9	8 e 9
x_1	-0,00006	0,00024	0,00030	-0,00027	0,00006	0,00034	0,00012	0,00003	0,00008	-0,00008	
x_2	-0,01376	0,00574	0,01951	-0,01717	-0,00685	0,01032	-0,00926	-0,00074	0,00851		
x_3	-0,01592	-0,00633	0,00959	0,00549	0,00550	0,00005	0,01431	0,00183	-0,01248		
x_4	-0,00856	0,00919	0,01775	0,02079	0,01101	-0,00978	0,02756	0,00442	-0,02314		
x_5	-0,16137	-0,35287	-0,19150	0,15760	0,03423	-0,12336	0,19287	0,00649	-0,18637		
x_6	-0,20196	-0,08483	0,11713	-0,02550	0,01153	0,03703	-0,02053	0,00575	0,02629		
x_7	-0,06541	0,00304	0,06845	-0,01718	0,01409	0,03127	0,02552	0,00305	-0,02247		
x_8	0,40786	0,05897	-0,34888	0,27701	0,83300	0,55598	0,19843	0,45528	0,25685		
x_9	-0,05208	-0,32496	-0,27287	-0,57847	-0,15426	0,42421	-0,08376	-0,04946	0,03430		
x_{10}	-0,78468	1,03445	1,81913	-1,16433	-0,35479	0,80953	1,48045	-0,15199	-1,63245		
m_{it}	-21,60	-8,34	13,25	-8,13	10,34	18,48	10,35	4,01	-6,34		

¹ x_1 : peso de grãos (kg/ha), x_2 : florescimento masculino (dias), x_3 : altura de planta (cm), x_4 : altura de espiga (cm),

x_5 : plantas acamadas e quebradas (%), x_6 : estande final, x_7 : número de espigas, x_8 : espigas doentes (%),

x_9 : umidade dos grãos (%), x_{10} : prolificidade

Dessa forma, a taxa de erro aparente obtida das funções discriminantes e da rede neural, além da taxa de rejeição da rede são dadas pela Tabela 4.8.

Tabela 4.8: Taxa de Erro Aparente (%) e Taxa de Rejeição (R) da rede neural (%) nos três cenários

Cenário	FDA	FDF	RNA	
	TEA	TEA	TEA	R
C_1	23,04	23,04	13,23	11,71
C_2	13,23	13,23	5,61	4,28
C_3	11,61	11,61	0,38	0

¹FDA: Função Discriminante de Anderson, FDF: Função Discriminante de Fisher, RNA: Rede Neural Artificial, TEA: Taxa de Erro Aparente, R: Porcentagem de indivíduos rejeitados pela rede neural

A taxa de erro aparente calculada da forma proposta para as funções discriminantes é subestimada uma vez que utiliza-se o artifício de empregar os mesmos dados para a obtenção da função discriminante e da probabilidade de má classificação. Quando a taxa de erro aparente é alta deve-se concluir que:

- As populações analisadas não são suficientemente diferenciadas para que possam ser distinguidas por meio das funções discriminantes.
- As populações analisadas, apesar de serem diferenciadas, não puderam ser distinguidas em razão da quantidade e qualidade das variáveis consideradas na discriminação.

Com a avaliação da taxa de erro aparente calculada utilizando os procedimentos usual e o de validação cruzada, apresentada na Tabela 4.9, pode-se perceber que a rede neural apresenta um desempenho de classificação superior às funções discriminantes nos três cenários propostos. À medida que a similaridade entre as populações vai aumentando, percebe-se o aparecimento de uma taxa de rejeição, que são aqueles indivíduos que não podem ser associados

Tabela 4.9: Taxa de Erro Aparente (%) e Taxa de Rejeição (R) da rede neural (%) nos três cenários considerando o procedimento de validação cruzada

Cenário	FDA	FDF	RNA	
	TEA	TEA	TEA	R
C_1	26,67	-	13,23	11,71
C_2	14,00	-	5,61	4,28
C_3	11,52	-	0,38	0

¹FDA: Função Discriminante de Anderson, FDF: Função Discriminante de Fisher, RNA: Rede Neural Artificial, TEA: Taxa de Erro Aparente, R: Porcentagem de indivíduos rejeitados pela rede neural

à uma determinada população por serem ambíguos em relação às suas características discriminatórias. Esta taxa de rejeição surge como vantagem das redes neurais sobre as funções discriminantes, pois minimiza a probabilidade de um indivíduo ser classificado erroneamente em uma das populações consideradas. Examinando somente aqueles indivíduos que efetivamente possuem poder discriminatório em suas características, observa-se uma menor taxa de erro de classificação quando se utiliza as redes neurais.

Verifica-se que as redes neurais podem ser consideradas uma ferramenta de suma importância para a solução de problemas de classificação. Criar uma rede neural para a resolução de um problema é uma tarefa que exige atenção quanto a alguns detalhes. O primeiro deles é a definição da sua forma, quantas camadas ela deve possuir e quais devem ser seus tamanhos. Segundo Braga et al. (2007), qualquer problema pode ser resolvido por uma rede neural feedforward com duas camadas ocultas. Na prática, o mais comum é utilizar apenas uma camada oculta, que é suficiente na maioria dos casos. A determinação do tamanho das camadas de entrada e de saída não é problemática, já que eles têm uma relação direta com o formato dos dados que se utiliza nas entradas e os que se deseja obter nas saídas. Determinar o tamanho da camada oculta é um processo de tentativa e erro, segundo Braga et al. (2007). Sabe-se que se ela for muito pe-

quena não terá poder de processamento suficiente para resolver o problema; por outro lado, se for muito grande, perderá sua capacidade de generalização e atuará como uma memória. Também é preciso ter em mente que o conjunto de exemplos utilizados no treinamento é crítico. Ele deve conter exemplos que representem o maior número de casos possíveis, para que o sistema seja capaz de "aprender" a resolver o problema nas mais diversas situações.

5. Conclusão

As redes neurais artificiais podem substituir as funções discriminantes na tarefa de classificar indivíduos em grupos previamente conhecidos, desde que seja estabelecida uma RNA adequada. Porém, esse novo paradigma de computação apresenta como desvantagem a impossibilidade de se identificar qual variável contribui de forma efetiva na saída da rede. Na análise discriminante, isso pode ser feito facilmente, avaliando os coeficientes de cada variável.

A grande desvantagem das redes neurais está na dificuldade de se encontrar um algoritmo ótimo, de forma a conseguir uma maximização da taxa de classificação correta e a minimização da taxa de rejeição de um indivíduo atribuído como ambíguo. O processo de se encontrar essa melhor arquitetura da rede é por tentativa e erro, o que se torna um trabalho árduo e duradouro.

Referências Bibliográficas

- [1] ABREU, F.B. Análise multivariada na determinação da divergência genética em feijão-vagem (*Phaseolus vulgaris L.*) de crescimento indeterminado. 2001. 73f. Dissertação (Mestrado em Produção Vegetal). Curso de Pós-graduação em Produção Vegetal, Universidade Estadual do Norte Fluminense Darcy Ribeiro.
- [2] ALMEIDA, F. C. Novo modelo organizacional baseado no cérebro humano. Revista de Administração, 1995, p.46-56 ALMEIDA, F. C. Desvendando o uso de Redes Neurais em Problemas de Administração de Empresas. Revista de Administração de Empresas, São Paulo, 1995, p. 46-55.
- [3] ANDERSON T.W. An Introduction to Multivariate Statistical Analysis. New York: John Wiley & Sons, 1958, 345 p.
- [4] ANDREWS, H. C. Mathematical Techniques in Pattern Recognition. Canada, John Wiley & Sons, inc, 1972.
- [5] ASSIS, G.M.L. et al. Discriminação de espécies de *Brachiaria* baseada em diferentes grupos de caracteres morfológicos. Revista Brasileira de Zootecnia, Viçosa, v.32, n.3, p.576-584, 2003.
- [6] BARBOSA, V. C. Redes Neurais e "Simulated Annealing" como Ferramentas para Otimização Combinatória, Investigación Operativa, 1989, vol. 1, n. 2, p.125-142.

- [7] BARRETO, J. M. (2002). Introdução às Redes Neurais Artificiais. Disponível em <http://www.inf.ufsc.br/.barreto/tutoriais/Survey.pdf>. Acessado em Novembro de 2008.
- [8] BATISTA, J. L. M. Análise Multivariada no Planejamento de Extensão Florestal : Subsídios para uma Política de Reocupação de Áreas. Curitiba, Universidade Federal do Paraná, 1990. (Dissertação de mestrado).
- [9] BENNETT, K. P. & MANGASARIAN, O. L. Neural Network Training Via Linear Programming, in P. M. Pardalos (Ed.), *Advances in Optimization and Parallel Computing*, North Holland, Amsterdam 1992a, p.56-67.
- [10] BENNETT, K. P. & MANGASARIAN, O. L. Robust Linear Programming Discrimination of Two Linearly Inseparable Sets, *Optimization Methods and Software*, 1992b, vol.1, p. 23-34.
- [11] BIALKOWSKI, W. L. Integration of paper machine operations, *Tappi Journal*, 1988, p.65-71.
- [12] BICKEL, P. J. K. A. & DOCKSOM. *Mathematical Statistics*. San Francisco, Hoden-Day, inc, 1977.
- [13] BISHOP, C.M. *Pattern Recognition and Machine Learning*. Springer, 2007
- [14] BLAMEY, S. L., FEARON, K. C. H., GILMORE, W. H., OSBORNE, D. H. & CARTER, D. C. Prediction of risk in biliary surgery, 1983.
- [15] BRAGA, A. de P., CARVALHO, A.P.de L.F. de, LUDERMIR, T.B., *Redes Neurais Artificiais - Teoria e Aplicações*, 2.ed. Rio de Janeiro, LTC, 2007, 226p.
- [16] CARNIERI C., YAMAMOTO, T. , SILVA, A. L. e PATIAS, N. Um algoritmo eficiente para Reconhecimento de Padrão via Programação Linear. Uberlândia, M. G., CNMAC, 1993.
- [17] CARVALHO, L. A. V., BARBOSA, V. C., MENDES, S. T., EIZIRIK, L. M. e FRANÇA, F. G. , UFRJ., *Redes Neurais - A volta do cérebro*, *Ciência Hoje*, 1994, vol. 12, n. 70, p. 12-21.

- [18] CHAMPION, H. R., SACCO, W. J. & HUNT, T. K. Trauma Severity Scoring to Predict Mortality, *World J. Surg.*, 1983, 7, p. 4-11.
- [19] CIPRA, B. A. A Chaotic Walk on Wall Street, *SIAM News*, may 1992.
- [20] CORDEIRO, G. M., Modelos Lineares Generalizados, VII SINAPE, Campinas, S. P., 1986.
- [21] CRUZ, C.D.; CARNEIRO, P.C.S. Modelos Biométricos Aplicados ao Melhoramento Genético - Vol 2. 2.ed. Viçosa: UFV, 2006. 585p.
- [22] CRUZ, C.D; FERREIRA, F.M; PESSONI, L.A. Biometria aplicada ao Estudo da Diversidade Genética. Viçosa Editora UFV, 2008.
- [23] CRUZ, C.D. Programa GENES: versão Windows, Aplicativo Computacional em Genética e Estatística. Análise Multivariada e Simulação. Viçosa: UFV, 2006. 175p.
- [24] DACHS,N., Estatística Computacional. São Paulo. Livros Técnicos e Científicos Editora S.A., 1988, 236.
- [25] CRUZ, C.D; FERREIRA, F.M; PESSONI, L.A. Biometria Aplicada ao Estudo da Diversidade Genética. Viçosa Editora UFV, 2008.
- [26] DEMUTH, H. & BEALE, M. Neural Network Toolbox for use with MATLAB (User's Guide). The Math Works, inc, 1994.
- [27] DOBSON, A. J. Introduction to Statistical Modelling. New York, Chapman and Hall, 1983.
- [28] EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. Centro Nacional de Pesquisa de Milho e Sorgo. Ensaio Nacional de Cultivares do Milho Centro: resultados do ano agrícola 2007/2008 - Minas Gerais. Sete Lagoas, 2008. Disponível em www.cnpms.embrapa.br/ensaio/ensaio2008/centro-1.php. Acessado em 23 de setembro de 2009.

- [29] FADUM, O. Artificial Intelligence : expert systems, fuzzy logic and neural network applications in the paper industry, Pulp & Paper, 1993.
- [30] FERGOSON, R. B. Chemical Process Optimization utilizing Neural Network Systems, Pavilion Technologies, Inc., Austin, Texas, 1992.
- [31] FERREIRA, R. DE P.; CRUZ, C. D.; SEDIYAMA, C. S.; FAGERIA, N. K. Identificação de cultivares de arroz tolerantes à toxidez de alumínio por técnica multivariada. Pesquisa Agropecuária Brasileira, Brasília, v. 30, n. 6, p. 789-795, jun. 1995.
- [32] FERNEDA, E. Redes neurais e sua aplicação em sistemas de recuperação de informação. Ci. Inf., Brasília, v. 35, n. 1, p. 25-30, jan./abr. 2006
- [33] FISHER, R.A. The use of multiple measurements in taxonomic problems. Annals of Eugenics, v.7, p.179-188, 1936.
- [34] GORNI, A. A. Redes Neurais Artificiais - Uma abordagem revolucionária em Inteligência Artificial. Micro Sistemas, São Paulo, 1993.
- [35] GRINOLD, R. C. Mathematical Programming Methods of Pattern Classification, Management Sciences, 1972, 19, p. 272-289.
- [36] GROTSCHHEL, M. & WAKABAYASHI, Y. A Cutting Plane Algorithm for a Clustering Problem, Mathematical Programming, 1989, 45, p. 59 - 96.
- [37] HADJIPROCOPIIS, A. A Neural Network Implementation on a Transputer System and Applications. London, London University, 1993. (Dissertação de mestrado).
- [38] HAIR, J.F., BLACK, W.C, BABIN,B.J., ANDERSON, R.E., TATHAN, R.L., Análise Multivariada de Dados, tradução Adonai Schlup Sant'Anna, 6.ed, Porto Alegre: Bookman, 2009, 688p.
- [39] HAMMERSTROM, D. Neural Networks at work. IEEE Spectrum, 1993, p. 26-32.

- [40] HAYKIN, S. Neural Networks - A Comprehensive Foundation. Macmillan College Publishing, inc., 1994.
- [41] HAYKIN, S. Redes Neurais, Princípios e Prática. Porto Alegre: Bookman, 2001.
- [42] HAYKIN, S. Neural Networks and Learning Machines. 3th ed., Prentice Hall, 2009
- [43] HECHT-NIELSEN, R. Neurocomputing. Addison-Wesley Publishing Company, New York, 1991.
- [44] HERTZ, J., KROGH, A. & PALMER, R. G. Introduction to the Theory of Neural Computation. Reading, Addison-Wesley Publishing Company, 1991.
- [45] HILSTER, D. & CAMPOS, M. B. Modelo da Rede Neural. PC Mundo, 1989, p. 40 - 43.
- [46] HOPFIELD, J.J. Neural networks and physical systems with emergent collective properties. Pro. Nat. Acad. Sci., 79:2554-8,1982
- [47] JOHNSON, R. A. & WICHERN, D. W. Applied Multivariate Statistical Analysis. New Jersey, Prentice-Hall, inc., 1999.
- [48] KHATTREE, R. & NAIK, D.N. Multivariate data reduction and discrimination with SAS software. Cary, NC, USA: SAS Institute Inc., 2000. 558 p.
- [49] KHOURY JR, J.K. Desenvolvimento e avaliação de um sistema de visão artificial para classificação de madeira serrada de eucalipto. 2004. 101 f. Tese (Doutorado em Engenharia Agrícola) - Universidade Federal de Viçosa, Viçosa, 2004.
- [50] KRÖSE, B. J. A. & VAN DER SMAGT, P. P. An Introduction to Neural Networks. Amsterdam, University of Amsterdam, 1993.
- [51] LUGOSI, G. Learning with an Unreliable Teacher, Pattern Recognition, 1992, 25, n.1, p.79-87.

- [52] MANGASARIAN, O. L. Linear and Nonlinear Separation of Patterns by Linear Programming, *Operations Research*, 1965, 13, p. 444-452.
- [53] MANGASARIAN, O. L. Multisurface Method of Pattern Separation, *IEEE Transactions on Information Theory*, 1968, vol. 14, n.6, p. 801-807.
- [54] MANGASARIAN, O. L., SETIONO, R. & WOLBERG, W. H. Pattern Recognition via Linear Programming : Theory and Application to Medical Diagnosis, in : *Large-Scale Numerical Optimization*, Thomas F. Coleman and Yuying Li, (Eds.), SIAM, Philadelphia 1990a, p.22-30.
- [55] MARDÍA, K. V., KENT, J. P. & BIBBY, J. M. *Multivariate Analysis*, London, Academic Press, 1979
- [56] MINGOTI, S.A., *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Belo Horizonte, Editora UFMG, 2005
- [57] McCULLOCH, W. and PITTS, W. A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bull. Math. Biophysics*, 1943, 5:115.133.
- [58] MCNITT, L.L., *Simulação em Basic*. Rio de Janeiro, Livros Técnicos e Científicos Editora S.A., 1985, 348p.
- [59] MOURA ZANINE, A. de et al. Uso de funções discriminantes para comparação de cultivares dos gêneros *Cynodon* e *Digitaria* quanto à produção de matéria seca e teores de macronutrientes. *Asociación Latinoamericana de Producción Animal*, v.15, n.4, p.152-156, 2007.
- [60] NASCIMENTO, C. A. O. e YAMAMOTO, C. I. Modelagem de Processos Químicos via Redes Neurais, 1993, *I Workshop em Redes Neurais*, IPT, São Paulo.
- [61] NAKANO, K.. Application of neural networks to the color grading of apples. *Computers and Electronics in Agriculture* 18 (1997) 105-116.

- [62] NAYLOR, T.H., BALINTFY, J.L., BURDICK, D.S., CHU, K. Técnicas de simulação em computadores. São Paulo, Vozes, 1971, 401p.
- [63] NELDER, J. A. & WEDDERBURN, R. W. M. Generalized Linear Models, J. R. Statist. Soc., A, n. 135, 1972, p. 370-384.
- [64] PALADINI, E. P. Controle de Qualidade - Uma Abordagem Abrangente. São Paulo, Ed. Atlas S.A., 1990.
- [65] PATUWO, E., HU, M. Y. & HUNG, M. S. Two-Group Classification using Neural Networks, Decision Sciences, 1993, vol 24, n. 4.
- [66] PEREIRA, J.J., Análise de agrupamento e discriminante no melhoramento genético - aplicação na cultura do arroz (*Oryza sativa L.*) 1999. 191 f. Tese(doutorado)
- [67] R Development Core Team (2006). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [68] RAMOS, J.P.S. Redes Neurais Artificiais na Classificação de Frutos: Cenário Bidimensional, Ciênc. agrotec., Lavras. V.27, n.2, p.356-362, mar.-abr., 2003.
- [69] REBOUÇAS, L., Rede neural vai detectar fraudes com cartões, Gazeta Mercantil, 1993.
- [70] REGAZZI, A.J. Análise Multivariada, Notas de Aula EST 746, Departamento de Informática da Universidade Federal de Viçosa, v.2, 2006.
- [71] ROEHRIG C., CARNIERI, C. e STEINER, M. T. A. Diagnóstico Diferencial da Causa da Icterícia Obstrutiva por P.L. Discriminante não Convexa, por Partes, XI Seminário Bras. de Endoscopia Digestiva, Curitiba, out. 1993
- [72] ROSENBLATT, F. Principles of Neurodynamics: Perceptrons and the theory of brain mechanisms. Spartan Books, New York, 1962

- [73] RUDD, J. B. Using a neural network system for advanced process control, Tappi Journal, 1991, p. 153-159.
- [74] RUMELHART, D.E., HINTON, G.E., WILLIAMS, R.J., Learning representations by back-propagating errors. Nature, 323:533-536, 1986
- [75] SCHMIDT, R. L. The effect of wood chip inventory rotation policies on storage costs, chip quality, and chip variability, Tappi Journal, nov. 1990, p. 211-216.
- [76] SENAI - Departamento Regional de São Paulo, IPT - Centro Técnico em Celulose e Papel, Tecnologia de fabricação da pasta celulósica, vol. I, Coordenação Técnica de M. L. O. D'Almeida, São Paulo, 1981b.
- [77] SIMÕES, A.S. e COSTA, A.H.R., Classificação de Laranjas Baseada em Padrões Visuais. Anais do Simpósio Brasileiro de Automação Inteligente, 2003.
- [78] SHARDA, R. Neural Networks for the MS/OR Analyst : An Application Bibliography, Interfaces, 1994, vol 24, n. 2, p. 116-130.
- [79] SMITH, F. W. Pattern Classifier Design by Linear Programming, IEEE Transactions on Computers, 1968, vol. C-17, n. 4, p.367-372.
- [80] STANGE, P. Programação Matemática no Ajustamento de Pontos e no Reconhecimento de Padrões. Florianópolis, Universidade Federal de Santa Catarina, 1988.
- [81] STEINER, M. T. A., CARNIERI, C., ROHERIG, C. Programação Linear e Métodos Estatísticos em Reconhecimento de Padrão, Campinas, S. P., XXV SBPO, 1993b.
- [82] STEINER, M. T. A. Verificação da Eficiência de Métodos Estatísticos no R. de Padrão via Simulação, XXVI SBPO, Florianópolis, dez. 1994.
- [83] STEINER, M. T. A., CARNIERI, C. Redes Neurais na Indústria de Papel, II Congresso Brasileiro de Redes Neurais, Curitiba, out. 1995.

- [84] STEPHENS, G. H. Advanced Process Management - a vision for the not too distant future, Tappi Journal, 1992.
- [85] STOVES, G. D. & FLANDERS, L. N. An Analytical Method for Determining the Lowest Cost Fiber to the Pulp Mill, Tappi Journal, 1992, p. 143-150.
- [86] SUDRÉ, C.P. et al. Variáveis multicategóricas na determinação da divergência genética entre acessos de pimenta e pimentão. Horticultura Brasileira, Campinas, v.24, n.1, p.88-93, 2006.
- [87] TAM, K. Y. & KIANG, M. Y. Managerial Applications of Neural Networks : The Case of Bank Failure Predictions, Management Sciences 38 n. 7, 1992, p. 926-947.
- [88] VARELLA, C.A.A. Estimativa da produtividade e do estresse nutricional da cultura do milho usando imagens digitais. 2004. 92 f. Tese (Doutorado em Engenharia Agrícola) - Universidade Federal de Viçosa, Viçosa, 2004.
- [89] WANG, S. The Unpredictability of Standard Back Propagation Neural Networks in Classification Applications, Management Science, 1995, vol. 41, n. 3, p. 555-559.
- [90] WINDOM Jr., D. R., COOK, D. F. & MASSEY, J. G. Integrating the art and science of papermaking in an expert system, Tappi Journal, 1991, p. 85-88.
- [91] WINDROW,B., HOFF,M.E., Adaptative switching circuits. Institute os Radio Engineers, Western Eletronic Show and Convention, 1960
- [92] WINDROW,B. The original adaptive neural net broom-balancer. In Proceedings of the IEEE International Symposium on Circuits and Systems, pages 351.357, Philadelphia, PA., 1957
- [93] ZAHEDI, F. An Introduction to Neural Networks and a Comparison with Artificial Intelligence and Expert Systems, 1991, Interfaces, vol 21, n. 2, p.25-38.

Apêndice A

Código fonte da rede neural

```
clear, close all, clc;

%echo on
%clear
numEntradas = 10; numEscondidos1 = 8;
%numEscondidos2 = 10;
numSaidas = 3; numTr = 1050; numVal = 300; numTeste = 149;
%echo off

lim_pop1 = 100; lim_pop2 = 200; lim_pop3 = 300;

tam_pop = 100;

Class_pop1 = []; Class_pop2 = []; Class_pop3 = [];

Pop1 = []; Pop2 = []; Pop3 = [];
```

```
TEAmenor = 100;

N = 5;

certo1 = 0; certo2 = 0; certo3 = 0; errado12 = 0; errado13 = 0;
errado21 = 0; errado23 = 0; errado31 = 0; errado32 = 0;

arquivoTreinamento = fopen('Treinamento.dat','rt');
arquivoValidacao = fopen('Validação.dat','rt'); arquivoTeste =
fopen('Teste.dat','rt');

dadosTreinamento = fscanf(arquivoTreinamento,'%f',[(numEntradas
+ numSaidas),numTr]); entradasTreinamento =
dadosTreinamento(1:numEntradas, 1:numTr); saidasTreinamento =
dadosTreinamento((numEntradas + 1):(numEntradas +
numSaidas),1:numTr);

dadosValidacao = fscanf(arquivoValidacao,'%f',[(numEntradas
+ numSaidas),numVal]); entradasValidacao =
dadosValidacao(1:numEntradas, 1:numVal); saidasValidacao =
dadosValidacao((numEntradas + 1):(numEntradas +
numSaidas),1:numVal)';

dadosTeste = fscanf(arquivoTeste,'%f',[(numEntradas), numTeste]);
entradasTeste = dadosTeste(1:numEntradas, 1:numTeste);
%saidasTeste = dadosTeste((numEntradas + 1):(numEntradas + numSaidas),
1:numTeste);

fclose(arquivoTreinamento); fclose(arquivoValidacao);
fclose(arquivoTeste);
```

```
[Pnorm, minP, maxP] = premmx(entradasTreinamento); [Tnorm, minT,
maxT] = premmx(saidasTreinamento);

[Qnorm, minQ, maxQ] = premmx(entradasValidacao); [Rnorm, minR,
maxR] = premmx(saidasValidacao);

[Snorm, minS, maxS] = premmx(entradasTeste);

%Para MLP de uma camada
rede = newff(minmax(Pnorm), [numEntradas numEscondidos1 numSaidas],
{'purelin', 'tansig', 'purelin'}, 'traingdx', 'learngd', 'mse');

%Para MLP de duas camada
%rede = newff(minmax(Pnorm), [numEntradas numEscondidos1 numEscondidos2
numSaidas], {'purelin', 'logsig', 'logsig', 'purelin'}, 'traingdx', 'learngd',
'sse');

fprintf('\nInicio do treinamento: ')

DATESTR(NOW)

fprintf('\n');

fprintf('\nTreinando ... \n')

for i=1:N

rede = init(rede);
```

```
rede.trainParam.epochs = 7000;
%rede.trainParam.lr = 0.7;
rede.trainParam.goal = 0.01;
%rede.trainParam.max_fail = 30;
%rede.trainParam.min_grad = 0;
rede.trainParam.show = 50;
%rede.trainParam.time = inf;
%rede.trainParam.mem_reduc = 14;

redeNova = train(rede,Pnorm,Tnorm);

Y = hardlim(sim(redeNova,Qnorm))';

Class_pop1 = [Y(1:lim_pop1,:)]; Class_pop2 =
[Y(lim_pop1+1:lim_pop2,:)]; Class_pop3 =
[Y(lim_pop2+1:lim_pop3,:)];

Pop1 = [saidasValidacao(1:lim_pop1,:)]; Pop2 =
[saidasValidacao(lim_pop1+1:lim_pop2,:)]; Pop3 =
[saidasValidacao(lim_pop2+1:lim_pop3,:)];

Erro_pop1 = length(Pop1) - sum(Pop1(1:tam_pop,1) ==
Class_pop1(1:tam_pop,1) & Pop1(1:tam_pop,2) ==
Class_pop1(1:tam_pop,2) & Pop1(1:tam_pop,3) ==
Class_pop1(1:tam_pop,3)); Erro_pop2 = length(Pop2) -
sum(Pop2(1:tam_pop,1) == Class_pop2(1:tam_pop,1) &
Pop2(1:tam_pop,2) == Class_pop2(1:tam_pop,2) & Pop2(1:tam_pop,3)
== Class_pop2(1:tam_pop,3)); Erro_pop3 = length(Pop3) -
sum(Pop3(1:tam_pop,1) == Class_pop3(1:tam_pop,1) &
Pop3(1:tam_pop,2) == Class_pop3(1:tam_pop,2) & Pop3(1:tam_pop,3)
```



```
== Class_pop3(1:tam_pop,3));

Nao_sei1 = length(Pop1) -
sum(Class_pop1(1:tam_pop,1)==Class_pop1(1:tam_pop,2)
==Class_pop1(1:tam_pop,3));
Nao_sei2 = length(Pop2) -
sum(Class_pop2(1:tam_pop,1)==Class_pop2(1:tam_pop,2)
==Class_pop2(1:tam_pop,3));
Nao_sei3 = length(Pop3) -
sum(Class_pop3(1:tam_pop,1)==Class_pop3(1:tam_pop,2)
==Class_pop3(1:tam_pop,3));

Erro_real1 = Erro_pop1 - Nao_sei1; Erro_real2 = Erro_pop2 -
Nao_sei2; Erro_real3 = Erro_pop3 - Nao_sei3;

certo1 = length(Pop1) - Erro_pop1; certo2 = length(Pop2) -
Erro_pop2; certo3 = length(Pop3) - Erro_pop3;

errado12 = sum(Pop1(1:tam_pop,1) ~= Class_pop1(1:tam_pop,1) &
Pop1(1:tam_pop,2) ~= Class_pop1(1:tam_pop,2) & Pop1(1:tam_pop,3)
== Class_pop1(1:tam_pop,3)); errado13 = sum(Pop1(1:tam_pop,1) ~=
Class_pop1(1:tam_pop,1) & Pop1(1:tam_pop,2) ==
Class_pop1(1:tam_pop,2) & Pop1(1:tam_pop,3) ~=
Class_pop1(1:tam_pop,3));

errado21 = sum(Pop2(1:tam_pop,1) ~= Class_pop2(1:tam_pop,1) &
Pop2(1:tam_pop,2) ~= Class_pop2(1:tam_pop,2) & Pop2(1:tam_pop,3)
== Class_pop2(1:tam_pop,3)); errado23 = sum(Pop2(1:tam_pop,1) ==
Class_pop2(1:tam_pop,1) & Pop2(1:tam_pop,2) ~=
Class_pop2(1:tam_pop,2) & Pop2(1:tam_pop,3) ~=
```

```
Class_pop2(1:tam_pop,3));

errado31 = sum(Pop3(1:tam_pop,1) ~= Class_pop3(1:tam_pop,1) &
Pop3(1:tam_pop,2) == Class_pop3(1:tam_pop,2) & Pop3(1:tam_pop,3)
~= Class_pop3(1:tam_pop,3)); errado32 = sum(Pop3(1:tam_pop,1) ==
Class_pop3(1:tam_pop,1) & Pop3(1:tam_pop,2) ~=
Class_pop3(1:tam_pop,2) & Pop3(1:tam_pop,3) ~=
Class_pop3(1:tam_pop,3));

%Taxa de Erro Aparente

TEA = 100 * (Erro_real1 + Erro_real2 +
Erro_real3)/length(entradasValidacao);

NaoClass = 100 * (Nao_sei1 + Nao_sei2 +
Nao_sei3)/length(entradasValidacao);

fprintf('\n');

fprintf(strcat(num2str(i),'a repetição\n'));

fprintf('\n');

fprintf('Taxa de indivíduos classificados erradamente: %f%%\n', TEA);

fprintf('\n');

fprintf('Taxa de indivíduos não classificados: %f%%\n', NaoClass);

fprintf('\n');
```

```
resultado1 = [Erro_real1 Erro_real2 Erro_real3 Nao_sei1 Nao_sei2
Nao_sei3];

xlswrite(strcat(num2str(i) , 'TEA.xls'), TEA);

xlswrite(strcat(num2str(i) , 'Nao_Classificado.xls'), NaoClass);

if TEA < TEAmenor
    TEAmenor = TEA;
    resultadoMelhor = resultado1;
    melhorRede = redeNova;
    Matriz=[certo1, errado12, errado13, Nao_sei1; errado21, certo2,
    errado23, Nao_sei2; errado31, errado32, certo3, Nao_sei3];
end

xlswrite(strcat(num2str(i) , 'Matriz.xls'), Matriz);

end

fprintf('\n');
fprintf('\nTestando ... \n');
Yt=hardlim(sim(melhorRede, Snorm));

fprintf('\n');
fprintf('\nFim do treinamento: ')
DATESTR(NOW)
diary off
```

Apêndice B

Descrição de algumas funções do *ntoll* do MATLAB

B.1 Funções de Inicialização

São funções onde definimos fatores como o tipo da rede a ser usada, número de camadas, número de neurônios por camadas, funções de ativação, etc. Para cada tipo de rede, existe uma função específica, conforme pode ser visto abaixo:

- `network`: cria uma rede neural customizada
- `newc`: cria uma rede competitiva
- `newcf`: cria uma rede backpropagation cascade-forward
- `newelm`: cria uma rede backpropagation Elman
- `newff`: cria uma rede backpropagation feed-forward multicamada
- `newfftd`: cria uma rede backpropagation feed-forward com retardo de entrada

- `newgrnn`: cria uma rede de regressão generalizada
- `newhop`: cria uma rede recorrente de hopfield

A rede neural utilizada nesse trabalho foi uma rede perceptron multicamadas com treinamento através do algoritmo backpropagation e sua criação se deu utilizando a função `newff`. Sua sintaxe é dada a seguir:

```
net = newff(PR, [S1 S2 ... SN], TF1 TF2 ... TFN, BTF, BLF, PF)
```

em que:

PR - matriz $R \times 2$ de valores min e max para os elementos de entrada R

Si - número de neurônios da i-ésima camada, para N camadas

TFi - função de transferência da i-ésima camada

BTF - função de treinamento backpropagation da rede

BLF - função de aprendizado dos pesos

PF - função de performance

B.2 Funções de Transferência (Ativação)

As principais são:

- `hardlim`: função de transferência Hard limit
- `logsig`: função de transferência Log sigmoide
- `poslin`: função de transferência Linear Positiva

- `purelin`: função de transferência Linear
- `tansig`: função de transferência Tangente Hiperbólica

B.3 Funções de Treinamento

As principais são:

- `trainbr`: regulação Bayesiana
- `traingd`: backpropagation de gradiente descendente
- `traingdm`: backpropagation de gradiente descendente com momentum
- `traingda`: backpropagation de gradiente descendente com taxa de aprendizagem adaptativa
- `trainlm`: backpropagation Levenberg-Marquardt

B.4 Funções de Aprendizado de Pesos

As principais são:

- `learngd`: função de aprendizagem Gradient descent weight/bias
- `learngh`: função de aprendizagem Gradient descent w/momentum weight/bias
- `learnh`: função de aprendizagem Hebb weight

B.5 Funções de Performance de Erro

As principais são:

- `mae`: função de performance de erro médio absoluto
- `mse`: função de performance de erro médio quadrático
- `msereg`: função de performance de erro médio quadrático com regulação
- `sse`: função de performance de soma de erro quadrático

B.6 Treinamento da Rede

A função responsável pelo treinamento da rede é a função `train`. Alguns parâmetros, dados a seguir, devem ser configurados:

- `net.trainParam.epochs`: número máximos de iterações
- `net.trainParam.show`: intervalos de iterações a serem mostrados
- `net.trainParam.goal`: erro máximo a ser atingido
- `net.trainParam.time`: tempo máximo de treinamento
- `net.trainParam.min_grad`: mínimo gradiente de erro
- `net.trainParam.mc`: taxa de momento
- `net.trainParam.lr`: taxa de aprendizagem

B.7 Utilizando a Rede

Após a inicialização e treinamento da rede, devemos simulá-la usando o comando *sim*:

```
Y = sim(net, P)
```

em que:

net - rede treinada

P - entrada da rede (a ser testada)