

UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM  
ENGENHARIA ELÉTRICA

DISSERTAÇÃO DE MESTRADO

Aplicação de Redes Neurais para Diagnóstico de  
Poluição em Isoladores de Alta Tensão



RECIFE, JUNHO DE 2008.

UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

APLICAÇÃO DE REDES NEURAIS PARA  
DIAGNÓSTICO DE POLUIÇÃO EM ISOLADORES DE  
ALTA TENSÃO

por

GABRIELA SEABRA MELO E SANTOS

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Pernambuco como parte dos requisitos para obtenção do grau de Mestre em Engenharia Elétrica.

**ORIENTADOR: Ronaldo Ribeiro Barbosa de Aquino, D.Sc.**

**CO-ORIENTADOR: José Maurício de Barros Bezerra, D.Sc.**

Recife, Junho de 2008.

**S237a Santos, Gabriela Seabra Melo e.**

Aplicação de redes neurais para diagnóstico de poluição em isoladores de alta tensão / Gabriela Seabra Melo e Santos. – Recife: O Autor, 2008.

xviii, 106 f.; il., tabs.

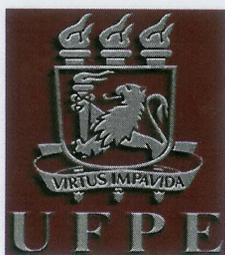
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia Elétrica, 2008.

Inclui Referências Bibliográficas.

**1. Engenharia Elétrica. 2. Inteligência Artificial. 3. Poluição em Isoladores. 4. Classificador de Padrões. 5. Ultra-som. 6. Descargas Superficiais. I. Título.**

**621.3 CDD (22.ed.)**

**UFPE/BCTG/2008-165**



**Universidade Federal de Pernambuco**

***Pós-Graduação em Engenharia Elétrica***

PARECER DA COMISSÃO EXAMINADORA DE DEFESA DE  
DISSERTAÇÃO DO MESTRADO ACADÊMICO DE

**GABRIELA SEABRA MELO E SANTOS**

TÍTULO

**“APLICAÇÃO DE REDES NEURAIS PARA DIAGNÓSTICO  
DE POLUIÇÃO EM ISOLADORES DE ALTA TENSÃO”**

A comissão examinadora composta pelos professores: RONALDO RIBEIRO BARBOSA DE AQUINO, DEESP/UFPE, JOSÉ MAURÍCIO DE BARROS BEZERRA, DEESP/UFPE, MARCOS TAVARES DE MELO, DES/UFPE e WASHINGTON LUIZ ARAÚJO NEVES, DEE/UFCG sob a presidência do primeiro, consideram a candidata **GABRIELA SEABRA MELO E SANTOS**  
**APROVADA.**

Recife, 19 de junho de 2008.

**EDUARDO FONTANA**  
Coordenador do PPGEE

**RONALDO RIBEIRO BARBOSA DE AQUINO**  
Orientador e Membro Titular Interno

**JOSÉ MAURÍCIO DE BARROS BEZERRA**  
Co-Orientador e Membro Titular Externo

**MARCOS TAVARES DE MELO**  
Membro Titular Interno

**WASHINGTON LUIZ ARAÚJO NEVES**  
Membro Titular Externo

Dedico este trabalho ao meu avô, Orlando,  
homem que sempre vibrou e se orgulhou  
com minhas conquistas.

# AGRADECIMENTOS

Agradeço primeiramente a Deus por ter me ajudado nesta difícil caminhada.

Ao professor Ronaldo, pela amizade, orientação e disponibilidade em todos os momentos.

Ao professor Maurício, pela co-orientação e pelos dados fornecidos sem os quais este trabalho não seria realizado.

Aos professores Afonso, Milde e Aida, pelas palavras de conforto momentos antes da defesa.

Gostaria também de agradecer o incentivo e o apoio dado pela minha família, aproveitando para desculpar-me por momentos de ausência e estresse, que vocês suportaram por amor. Mãe, tia Magaly, tia Marly e Vovó, registro aqui o meu "muito obrigada".

Ao meu irmão João Lucas, que amo incondicionalmente, e estava sempre de braços abertos com um lindo sorriso no rosto para me receber independentemente do meu humor.

Ao amigo Gustavo que se fez extremamente presente na elaboração deste trabalho, me ajudando sempre, não só no momento em que estava no Brasil, mas também da Espanha.

Às pessoas que mesmo longe, em João Pessoa ou em Brasília, se fazem sempre presente em momentos importantes da minha vida. Tia Carmita, tio Nonato, Let, Doca, Diéi, tia Quitéria, muito obrigada pelo incentivo e ajuda.

Ao amigo de jornada Josinaldo, que enfrentou assim como eu a dificuldade de trabalhar fora de Recife e escrever a dissertação nas noites e finais de semana. Além de compartilharmos o mesmo momento de ansiedade para a defesa.

À minha irmãzinha de coração, Marizinha, pessoa pura e de princípios invejáveis, que nunca desistiu da nossa amizade e gastou horas e horas do seu tempo comigo. "Para sempre!!"

Ao amigo Cadu, que em toda e qualquer situação me faz sorrir mesmo que eu não queira.

Às minhas amigas do Colégio Damas que são alicerces fundamentais em minha vida. Maria, Taxa, Kah, muito obrigada por tudo que vocês fizeram e fazem por mim.

Aos amigos Dani e Milton, que sempre me encorajam a lutar pelo que quero, por mais que seja difícil.

Aos meus colegas de trabalho, tanto da Saelpa como da Compesa, que sempre incentivaram e foram flexíveis, viabilizando a conclusão deste trabalho.

À tia Maria, Nina, Onca, Gorda e família, obrigada pela torcida. Amo a eterna alegria de vocês!

Enfim, a todos que me ajudaram, direta ou indiretamente, meus mais sinceros agradecimentos.

GABRIELA SEABRA MELO E SANTOS

*Universidade Federal de Pernambuco*

*19 de Junho de 2008*

Resumo da Dissertação apresentada à UFPE como parte dos requisitos necessários para a obtenção do grau de Mestre em Engenharia Elétrica

## **APLICAÇÃO DE REDES NEURAIS PARA DIAGNÓSTICO DE POLUIÇÃO EM ISOLADORES DE ALTA TENSÃO**

**Gabriela Seabra Melo e Santos**

Junho/2008

**Orientador:** Ronaldo Ribeiro Barbosa de Aquino, D.Sc.

**Área de Concentração:** Processamento de Energia

**Palavras-chaves:** Inteligência Artificial, Poluição em Isoladores, Classificador de padrões, Ultra-som, Descargas superficiais

**Número de páginas:** 106

Este trabalho se propõe a apresentar uma metodologia para classificar o nível de poluição presente na superfície dos isoladores aéreos de alta tensão, ou seja, diagnosticar o estado do isolamento do sistema elétrico de transmissão através da técnica de reconhecimento de padrões dos sinais de ultra-som emitidos por descargas superficiais nos isoladores a serem classificados via Redes Neurais Artificiais (RNA). Visando este objetivo, foram utilizadas neste trabalho técnicas de extração de atributos dos sinais de entrada da RNA visando viabilizar a execução computacional do treinamento e simulação da rede. Pode-se citar como técnicas de extração utilizadas: o Centróide, Parâmetros Estatísticos e no Domínio da Freqüência. Na aplicação das RNA além da Rede Perceptrons de Múltiplas Camadas (MLP), também foi simulada a Rede de Funções de Bases Radiais (RBF) com o objetivo de realizar uma análise comparativa dos resultados das redes na montagem de processos de classificação, utilizando a técnica de extração de atributos que resultou um melhor desempenho para as redes MLP.

Abstract of Dissertation presented to UFPE as a partial fulfillment of the requirements for the degree of Master in Electrical Engineering

**APPLICATION OF NEURAL NETWORKS FOR DIAGNOSIS OF  
POLLUTION IN INSULATORS OF HIGHT TENSION**

**Gabriela Seabra Melo e Santos**

June/2008

**Supervisor:** Ronaldo Ribeiro Barbosa de Aquino, D.Sc.

**Area of Concentration:** Energy processing

**Keywords:** Artificial intelligence, Insulators's pollution, Pattern recognition, Ultrasound, Superficial discharges

**Number of pages:** 106

This work proposes a methodology to classify the current pollution level on insulators surface, in other words, to diagnose the state of the electrical system isolation through pattern recognition technique on the ultrasonic signals obtained from surface discharges on outdoor insulators to be classified via Artificial Neural Networks (ANN). Pattern extraction techniques on the inputs signals from ANN were used in this work in order to turn the computational execution feasible in the training and simulation of the network. It can be point out as extractions techniques: the Centroid, Statistics Parameters and the Frequency Domain. In the use of ANN, besides Multilayer Perceptron Networks (MLP), the Radial Basis Function Network (RBF) was also used to make possible a comparison of the performance of the networks with pattern recognition based on features extraction techniques that led to best performance to the MLP networks.

# SUMÁRIO

<b>LISTA DE FIGURAS</b>	<b>xii</b>
<b>LISTA DE TABELAS</b>	<b>xiv</b>
<b>NOMENCLATURA</b>	<b>xvii</b>
<b>1 INTRODUÇÃO E OBJETIVO</b>	<b>1</b>
<b>1.1 Poluição em Isoladores</b> . . . . .	1
<b>1.2 Reconhecimento de Padrões</b> . . . . .	7
<b>1.3 Sinopse dos Capítulos</b> . . . . .	9
<b>2 BASE TEÓRICA - REDES NEURAIS ARTIFICIAIS</b>	<b>10</b>
<b>2.1 Reconhecimento de Padrões</b> . . . . .	10
<b>2.2 Redes Neurais Artificiais</b> . . . . .	12
2.2.1 Funções de Ativação . . . . .	14
2.2.2 Perceptron e Adaline . . . . .	17
2.2.3 Redes Perceptron de Múltiplas Camadas . . . . .	21
2.2.4 Algoritmo do Gradiente Descendente . . . . .	23
2.2.5 Algoritmo Resilient Propagation . . . . .	24
2.2.6 Algoritmo Levenberg-Marquardt . . . . .	24
<b>2.3 Redes de Funções de Base Radiais</b> . . . . .	26
2.3.1 Rede Neural Probabilística . . . . .	27
2.3.2 Rede RBF Eficiente . . . . .	28
2.3.3 Rede RBF Exata . . . . .	29
2.3.4 Síntese do Capítulo . . . . .	29

<b>3</b>	<b>TRATAMENTO DOS SINAIS DE ENTRADA</b>	<b>30</b>
3.1	Montagens Experimentais . . . . .	30
3.2	Constituição das Amostras e Estabelecimento das Classes . . .	31
3.3	Digitalização de Sinais . . . . .	32
3.3.1	Amostragem . . . . .	33
3.4	Normalização . . . . .	35
3.5	Extração de Atributos . . . . .	35
3.5.1	Centróide . . . . .	36
3.5.2	Parâmetros Estatísticos . . . . .	39
3.5.3	Sinais no Domínio da Freqüência . . . . .	43
3.5.4	Síntese do Capítulo . . . . .	46
<b>4</b>	<b>MODELAGEM DAS REDES PERCEPTRON DE MÚLTIPLAS CAMADAS</b>	<b>47</b>
4.1	Separação dos Dados de Entrada . . . . .	48
4.2	Codificação das Classes . . . . .	49
4.3	Seleção das Melhores Arquiteturas . . . . .	50
4.3.1	Melhores Arquiteturas - Centróide - Modelo 1 . . . . .	50
4.3.2	Melhores Arquiteturas - Centróide - Modelo 2 . . . . .	51
4.3.3	Melhores Arquiteturas - Parâmetros Estatísticos - Modelo 1 . . .	52
4.3.4	Melhores Arquiteturas - Parâmetros Estatísticos - Modelo 2 . . .	53
4.3.5	Melhores Arquiteturas - Domínio da Freqüência - Modelo 1 . . .	54
4.3.6	Melhores Arquiteturas - Domínio da Freqüência - Modelo 2 . . .	55
4.4	Seleção dos Melhores Experimentos . . . . .	56
4.4.1	Melhores Experimentos - Centróide - Modelo 1 . . . . .	58
4.4.2	Melhores Experimentos - Centróide - Modelo 2 . . . . .	58
4.4.3	Melhores Experimentos - Parâmetros Estatísticos - Modelo 1 . .	59
4.4.4	Melhores Experimentos - Parâmetros Estatísticos - Modelo 2 . .	60
4.4.5	Melhores Experimentos - Domínio da Freqüência - Modelo 1 . . .	61
4.4.6	Melhores Experimentos - Domínio da Freqüência - Modelo 2 . . .	62
4.5	Síntese do Capítulo . . . . .	63

<b>5</b>	<b>SIMULAÇÃO DAS REDES PERCEPTRON DE MÚLTIPLAS CAMADAS</b>	<b>65</b>
5.1	<b>Classificação</b>	65
5.1.1	Classificador com Um Especialista	65
5.1.2	Classificador com Três Especialistas	66
5.2	<b>Seleção das Melhores Inicializações</b>	66
5.2.1	Melhores Inicializações - Centróide - Modelo 1	67
5.2.2	Melhores Inicializações - Centróide - Modelo 2	69
5.2.3	Melhores Inicializações - Parâmetros Estatísticos - Modelo 1	71
5.2.4	Melhores Inicializações - Parâmetros Estatísticos - Modelo 2	71
5.2.5	Melhores Inicializações - Domínio da Freqüência - Modelo 1	73
5.2.6	Melhores Inicializações - Domínio da Freqüência - Modelo 2	75
5.3	<b>Resultados das Simulações com Um Especialista</b>	76
5.3.1	Resultados das Simulações - Centróide - Modelo 1	76
5.3.2	Resultados das Simulações - Parâmetros Estatísticos - Modelo 1	77
5.3.3	Resultados das Simulações - Domínio da Freqüência - Modelo 1	77
5.3.4	Resultados das Simulações - Centróide - Modelo 2	78
5.3.5	Resultados das Simulações - Parâmetros Estatísticos - Modelo 2	78
5.3.6	Resultados das Simulações - Domínio da Freqüência - Modelo 2	79
5.4	<b>Resultados das Simulações com Três Especialistas</b>	79
5.4.1	Resultados das Simulações - Centróide - Modelo 3	80
5.4.2	Resultados das Simulações - Parâmetros Estatísticos - Modelo 3	81
5.4.3	Resultados das Simulações - Domínio da Freqüência - Modelo 3	82
5.4.4	Resultados das Simulações - Centróide - Modelo 4	84
5.4.5	Resultados das Simulações - Parâmetros Estatísticos - Modelo 4	85
5.4.6	Resultados das Simulações - Domínio da Freqüência - Modelo 4	86
5.4.7	Síntese do Capítulo	88
<b>6</b>	<b>MODELAGEM E SIMULAÇÃO DAS REDES DE FUNÇÕES DE BASES RADIAIS</b>	<b>89</b>
6.1	<b>Introdução Teórica</b>	89
6.2	<b>Seleção dos Melhores Experimentos</b>	91

6.2.1	Melhores Experimentos - RBF Exata - Modelo 1 . . . . .	91
6.2.2	Melhores Experimentos - RBF Exata - Modelo 2 . . . . .	92
6.2.3	Melhores Experimentos - RBF Eficiente - Modelo 1 . . . . .	93
6.2.4	Melhores Experimentos - RBF Eficiente - Modelo 2 . . . . .	93
6.2.5	Melhores Experimentos - RBF Probabilística - Modelo 1 . . . . .	94
<b>6.3</b>	<b>Seleção da Melhor Constante do <i>SPREAD</i> . . . . .</b>	<b>95</b>
6.3.1	Melhor <i>SPREAD</i> - RBF Exata - Modelo 1 . . . . .	95
6.3.2	Melhor <i>SPREAD</i> - RBF Exata - Modelo 2 . . . . .	95
6.3.3	Melhor <i>SPREAD</i> - RBF Eficiente - Modelo 1 . . . . .	96
6.3.4	Melhor <i>SPREAD</i> - RBF Eficiente - Modelo 2 . . . . .	97
6.3.5	Melhor <i>SPREAD</i> - RBF Probabilística - Modelo 1 . . . . .	97
<b>7</b>	<b>CONCLUSÕES E TRABALHOS FUTUROS . . . . .</b>	<b>98</b>
7.1	Conclusões . . . . .	98
7.2	Trabalhos Futuros . . . . .	102
	<b>REFERÊNCIAS BIBLIOGRÁFICAS . . . . .</b>	<b>103</b>

# LISTA DE FIGURAS

1.1	Bicos injetores . . . . .	2
1.2	Representação elétrica de isoladores . . . . .	4
1.3	Linha de fuga no isolador . . . . .	5
1.4	Percentual de falhas nas linhas de transmissão da CHESF . . . . .	6
1.5	Percentual mensal de falhas devido à poluição nas linhas de transmissão da CHESF . . . . .	7
1.6	Reconhecimento de padrões . . . . .	8
2.1	Diagrama de um classificador de padrões . . . . .	11
2.2	Neurônio biológico . . . . .	13
2.3	Neurônio de McCulloch e Pitts . . . . .	13
2.4	Rede neural artificial . . . . .	14
2.5	Função linear . . . . .	15
2.6	Função rampa . . . . .	15
2.7	Função degrau . . . . .	16
2.8	Função tangente hiperbólica . . . . .	16
2.9	Função sigmóide logística . . . . .	17
2.10	Grafo de fluxo de sinais do Perceptron . . . . .	18
2.11	Grafo de fluxo de sinais do Perceptron . . . . .	19
2.12	Rede neural múltiplas camadas - MLP . . . . .	22
2.13	Propagação e retropropagação de sinais na rede . . . . .	23
2.14	Arquitetura de redes RBF . . . . .	26
2.15	Arquitetura da rede neural probabilística . . . . .	28
3.1	Exemplificação de um sinal de ultra-som . . . . .	32

3.2	Diagrama do sistema de digitalização de sinais . . . . .	32
3.3	Exemplificação de <i>aliasing</i> . . . . .	34
3.4	Extração de atributos do centróide . . . . .	36
3.5	Área positiva do Centróide . . . . .	37
3.6	Área negativa do Centróide . . . . .	37
3.7	Exemplificação do cálculo do centróide . . . . .	38
3.8	Extração de atributos estatísticos . . . . .	40
3.9	Extração de atributos no domínio da freqüência . . . . .	46
4.1	Treinamento com parada antecipada . . . . .	48
5.1	Classificação com um especialista . . . . .	66
5.2	Classificação com três especialista . . . . .	66
6.1	Diagrama de um classificador de padrões . . . . .	90
6.2	Função de base radial . . . . .	91
7.1	Faixa de variação do EC da rede combinada . . . . .	100

# LISTA DE TABELAS

1.1	Níveis de ESDD para cada classe . . . . .	3
4.1	Dimensão de entrada da RNA . . . . .	49
4.2	Codificação da camada de saída da RNA . . . . .	49
4.3	Seleção da melhor arquitetura - Centróide - Modelo 1 . . . . .	51
4.4	Arquitetura das melhores redes - Centróide - Modelo 1 . . . . .	51
4.5	Seleção da melhor arquitetura - Centróide - Modelo 2 . . . . .	52
4.6	Arquitetura das melhores redes - Centróide - Modelo 2 . . . . .	52
4.7	Seleção da melhor arquitetura - Parâmetros Estatísticos - Modelo 1 . . . . .	53
4.8	Arquitetura das melhores redes - Parâmetros Estatísticos - Modelo 1 . . . . .	53
4.9	Seleção da melhor arquitetura - Parâmetros Estatísticos - Modelo 2 . . . . .	54
4.10	Arquitetura das melhores redes - Parâmetros Estatísticos - Modelo 2 . . . . .	54
4.11	Seleção da melhor arquitetura - Domínio da Freqüência - Modelo 1 . . . . .	55
4.12	Arquitetura das melhores redes - Domínio da Freqüência - Modelo 1 . . . . .	55
4.13	Seleção da melhor arquitetura - Domínio da Freqüência - Modelo 2 . . . . .	56
4.14	Arquitetura das melhores redes - Domínio da Freqüência - Modelo 2 . . . . .	56
4.15	Validação cruzada . . . . .	57
4.16	EC - Centróide - Modelo 1 . . . . .	58
4.17	EC - Centróide - Modelo 2 . . . . .	59
4.18	EC - Parâmetros Estatísticos - Modelo 1 . . . . .	60
4.19	EC - Parâmetros Estatísticos - Modelo 2 . . . . .	61
4.20	EC - Domínio da Freqüência - Modelo 1 . . . . .	62
4.21	EC - Domínio da Freqüência - Modelo 2 . . . . .	63
4.22	Síntese dos Resultados da Modelagem das Redes . . . . .	64

5.1	EC - Centróide - Modelo 1 - Experimento 3 . . . . .	67
5.2	EC - Centróide - Modelo 1 - Experimento 4 . . . . .	68
5.3	EC - Centróide - Modelo 1 - Experimento 9 . . . . .	68
5.4	EC - Centróide - Modelo 2 - Experimento 1 . . . . .	69
5.5	EC - Centróide - Modelo 2 - Experimento 5 . . . . .	70
5.6	EC - Centróide - Modelo 2 - Experimento 7 . . . . .	70
5.7	EC - Parâmetros Estatísticos - Modelo 1 - Experimento 4 . . . . .	71
5.8	EC - Parâmetros Estatísticos - Modelo 2 - Experimento 4 . . . . .	72
5.9	EC - Parâmetros Estatísticos - Modelo 2 - Experimento 5 . . . . .	73
5.10	EC - Domínio da Frequência - Modelo 1 - Experimento 2 . . . . .	74
5.11	EC - Domínio da Frequência - Modelo 1 - Experimento 7 . . . . .	74
5.12	EC - Domínio da Frequência - Modelo 2 - Experimento 5 . . . . .	75
5.13	EC - Domínio da Frequência - Modelo 2 - Experimento 6 . . . . .	76
5.14	EC das redes combinadas - Centróide - Modelo 3 . . . . .	80
5.15	EC das redes combinadas - Parâmetros Estatísticos - Modelo 3 . . . . .	81
5.16	EC das redes combinadas - Domínio da Frequência - Modelo 3 . . . . .	83
5.17	EC das redes combinadas - Centróide - Modelo 4 . . . . .	84
5.18	EC das redes combinadas - Parâmetros Estatísticos - Modelo 4 . . . . .	85
5.19	EC das redes combinadas - Domínio da Frequência - Modelo 4 . . . . .	87
5.20	Síntese dos Resultados da Modelagem das Redes . . . . .	88
6.1	Seleção do melhor experimento - RBF Exata - Modelo 1 . . . . .	92
6.2	Seleção do melhor experimento - RBF Exata - Modelo 2 . . . . .	92
6.3	Seleção do melhor experimento - RBF Eficiente - Modelo 1 . . . . .	93
6.4	Seleção do melhor experimento - RBF Eficiente - Modelo 2 . . . . .	94
6.5	Seleção do melhor experimento - RBF Probabilística - Modelo 1 . . . . .	94
6.6	Seleção do SPREAD - RBF Exata - Modelo 1 . . . . .	95
6.7	Seleção do SPREAD - RBF Exata - Modelo 2 . . . . .	96
6.8	Seleção do SPREAD - RBF Eficiente - Modelo 1 . . . . .	96
6.9	Seleção do SPREAD - RBF Eficiente - Modelo 2 . . . . .	97
6.10	Seleção do SPREAD - RBF Probabilística - Modelo 1 . . . . .	97

7.1	Resumo dos resultados das simulações das redes MLP . . . . .	99
7.2	Resumo dos resultados das simulações das redes RBF . . . . .	101

# Nomenclatura

$\alpha$	Inclinação da Reta
$\epsilon$	Erro
$\eta$	Taxa de Aprendizagem
$\gamma$	Valor Máximo/Mínimo de Saída
$\kappa$	Curtose
$\Lambda$	Assimetria
$\mu$	Média
$\sigma$	Desvio Padrão
$\varphi$	Função de Ativação
$b$	Bias
$C$	Padrões Classificados Corretamente
$C_v$	Coefficiente de Variação
$d$	Saída Desejada
$E$	Valor Esperado
$EC$	Erro de Classificação
$I$	Padrões Classificados Incorretamente
$MSE$	Erro Médio Quadrático
$N$	Todos os Padrões Classificados
$Q$	Quantidade de Exemplos
$T$	Parâmetro de Suavidade da Curva
$t$	Tempo
$u$	Vetor Centro da Função Radial

<i>w</i>	Pesos Sinápticos das Redes Neurais Artificiais
<i>x</i>	Entrada da Rede Neural Artificial
<i>y</i>	Saída da Rede Neural Artificial

# 1

## INTRODUÇÃO E OBJETIVO

O capítulo atual visa apresentar o principal propósito do presente trabalho que é a detecção do nível de poluição em isoladores aéreos de alta tensão. Serão apresentados os níveis de poluição utilizados na classificação, assim como a metodologia utilizada para a realização da poluição artificial. No referido capítulo também foi abordada a problemática existente em empresas do ramo de energia elétrica envolvendo falhas em isoladores ocasionadas pelo elevado índice de poluição.

### 1.1 Poluição em Isoladores

A contaminação de isoladores, utilizados em linhas de transmissão de alta e média tensão, é responsável por uma parcela considerável de falhas de um sistema de potência. Visando uma maior continuidade e confiabilidade do sistema elétrico faz-se necessário o desenvolvimento de técnicas que auxiliem na detecção do elevado grau de poluição depositada na superfície do isolador. A deposição de poluição na superfície de isoladores preocupa bastante por reduzir suas qualidades dielétricas, podendo, em situações críticas não suportar a tensão de operação, ocasionando a descarga superficial completa (*flashover*) [1].

O desempenho das instalações elétricas está diretamente relacionado à performance dos seus isoladores [2]. Haja vista que um percentual significativo de falhas no sistema

elétrico são decorrentes da poluição em isoladores.

Preocupado com os problemas de continuidade do fornecimento de energia ocasionados pela poluição dos isoladores, este trabalho propõe um sistema de classificação que viabilize a monitoração dos sinais de ultra-som relacionados com o nível de contaminação de isoladores para ajudar na redução de falhas no sistema de potência [3].

A poluição atmosférica pode ser generalizada, ou seja, pode atingir em longo trecho de isoladores da linha, ou pontual onde seus efeitos são em áreas de pequena extensão, suas origens são respectivamente naturais e industriais. Nos ensaios realizados por [4] no Laboratório de Alta Tensão da UFCG a névoa foi simulada por uma câmara de composta por seis pares de bicos injetores. A névoa é formada a partir do encontro dos fluidos (água e ar comprimido) como mostrado na Figura 1.1. Os isoladores foram previamente submetidos à deposição de camada poluente com valores de ESDD (*Equivalent Salt Deposit Density* - Densidade equivalente de sal depositado) estabelecidos como apresentado na Tabela 1.1.

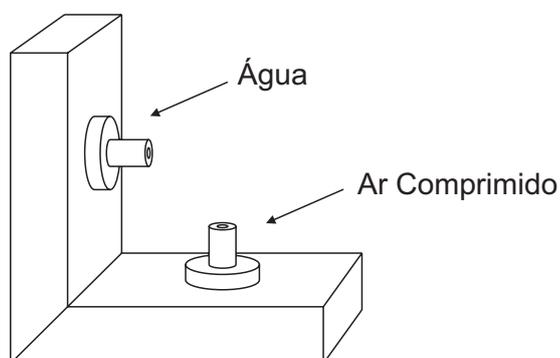


Figura 1.1: Bicos injetores

Tabela 1.1: Níveis de ESDD para cada classe

<b>Classes</b>	<b>Distúrbios</b>	<b>Faixas de ESDD (<math>mg/cm^2</math>)</b>
A	Atmosfera sem poluição	-
B	Atmosfera limpa	0,00 - 0,032
C	Contaminação muito leve	0,032 - 0,038
D	Contaminação leve	0,038 - 0,056
E	Contaminação pesada	0,056 - 0,123
F	Contaminação muito pesada	Maior que 0,123

Em seguida foram aplicados níveis de tensão [5] como estabelecido em trabalhos anteriores que procuram avaliar as condições da poluição e tensão aplicada que conduzirão à propagação do arco elétrico iniciado [6] [7].

A técnica aqui abordada visa a utilização de detecção de sinais de ultra-som, um método não destrutivo, sendo desnecessária a retirada do isolador da linha de transmissão para análise, não interrompendo o fornecimento de energia e nem exigindo uma elevada quantidade de homem-hora. A referida técnica de diagnóstico procura aprimorar as ações preditivas voltadas para alertar quanto à necessidade de manutenção, antes que equipamentos ou instalações venham a sofrer danos ou interrupções de suas condições operativas normais, assegurando a confiabilidade do sistema elétrico.

Sendo dada ênfase, neste trabalho, a isoladores aéreos de alta tensão, presentes em subestações e linhas de transmissão. Ficou evidenciado que quando uma cadeia de isoladores contaminada é energizada em estado seco não existe evidência de cintilação [4]. A corrente de fuga é tão pequena que praticamente toda corrente do isolador permanece capacitiva. Nesta ocasião o isolador possui a maior parte de sua suportabilidade, como se estivesse limpo e seco. Com o surgimento da neblina, se forma um filme de umidade na superfície contaminada, ocorrendo, por conseguinte, um aumento gradual na amplitude da corrente de fuga. A fase da corrente de fuga muda para um tipo mais resistivo. Como o contaminante começa a conduzir, centelhas se

formam no contorno da porcelana e do cimento do pino. Algumas vezes essas centelhas também se formam na saia interna do isolador e na cavidade entre a campânula e a porcelana. As representações elétricas das referidas situações, contaminado sem umidade ou limpo, e contaminado com umidade são respectivamente ilustradas nas Figura 1.2(a) e Figura 1.2(b).

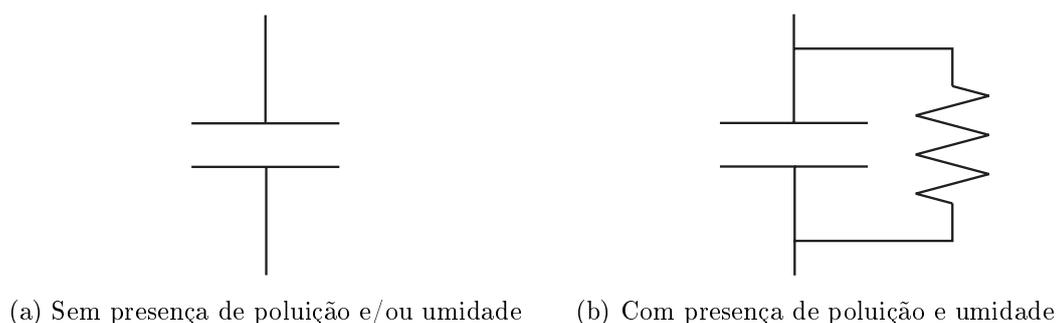


Figura 1.2: Representação elétrica de isoladores

A condutividade superficial, acontece devido à ionização que ocorre quando as moléculas de gás dissociam-se em ions carregados negativamente (elétrons livres) e positivamente. A descarga superficial total é também conhecida como flashover, ilustrada na Figura 1.3. Essa condutividade superficial é favorecida pela deposição de poeiras e sais na superfície do isolador, juntamente com a presença da umidade. Pode-se citar, como fatores que dificultam o surgimento do *Flash Over*, os seguintes itens:

- Forma do isolador;
- Limpeza dos isoladores;
- Elevada incidência pluviométrica.

A linha de fuga é o caminho mais curto que, seguindo a superfície do dielétrico, liga duas partes metálicas do isolador, Figura 1.3. As correntes de fuga foram analisadas em trabalhos anteriores para determinar os períodos para manutenção preventiva baseado no nível máximo da corrente de fuga que um isolador pode suportar até que ocorra

a descarga [8], uma técnica distinta da apresentada neste trabalho mas com o mesmo objetivo de reduzir as falhas no sistema elétrico devido a poluição em isoladores.

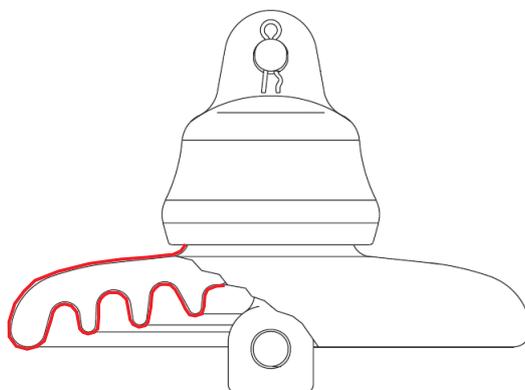


Figura 1.3: Linha de fuga no isolador

As falhas em isolamentos elétricos são as principais responsáveis por interrupções no fornecimento de energia elétrica aos diversos consumidores supridos por um sistema de potência. Para o caso específico de linhas de transmissão, como dito anteriormente, se o desempenho das instalações está diretamente relacionado com o desempenho dos seus isoladores, a introdução de tecnologias confiáveis que possibilitem alertar às equipes de manutenção quanto ao processo degenerativo de isolamentos representa, uma ação estratégica importante, tanto para a operação normal do sistema quanto para os consumidores que serão poupados do infortúnio da falta intempestiva da energia.

As análises aqui empreendidas buscam introduzir técnicas de diagnóstico que possam ser utilizadas em ações preditivas voltadas a alertar quanto à necessidade de manutenções especiais antes que os equipamentos ou instalações venham a sofrer danos e/ou interrupção de suas condições normais de operação.

A suportabilidade dos isoladores reduz drasticamente quando os mesmos são submetidos a determinadas condições climáticas como chuva e alta umidade, principalmente em conjunto com um ambiente poluído. A presença apenas da poluição, sem umidade, não reduz a suportabilidade dielétrica. Porém se a superfície do isolador for umidificada, de modo a umidificar os sais, mas não removê-los, a probabilidade da

ocorrência de um arco na superfície do isolador aumentará substancialmente podendo ocasionar uma descarga total sobre a superfície, ou seja, falhas no sistema elétrico.

Como pode-se ver na Figura 1.4 o percentual de falhas ocasionadas pela poluição nas linhas de transmissão da Chesf foi de 6%, no horizonte de 20 anos (1986-2006) de operação do sistema [9]. Apesar de não ser o foco deste trabalho destaca-se ainda um percentual significativo de falhas no sistema elétrico devido à poluição em níveis de média tensão [10].

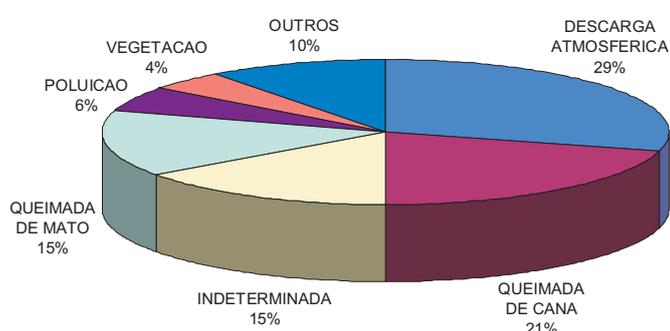


Figura 1.4: Percentual de falhas nas linhas de transmissão da CHESF

Apesar do percentual de falhas ocasionadas pela poluição nos isoladores nas linhas de transmissão da Chesf ser pouco representativo no horizonte de 20 anos, 6%, deve-se levar em consideração a sazonalidade de ocorrências destas referidas faltas. Na Figura 1.5 serão apresentados os percentuais das ocorrências das falhas mensais ocasionadas apenas pela poluição para os anos de 1987 até 2005. Onde pode-se destacar o elevado percentual no mês de Novembro de 1998 que apresentou valores superiores a 60% e o mês de Julho de 2005 que atingiu valores superiores a 50%, motivo pelo qual há tamanha preocupação com a detecção do grau de poluição presente na superfície dos isoladores proposta neste trabalho. Evidenciando então a importância do presente trabalho ao Sistema Elétrico.

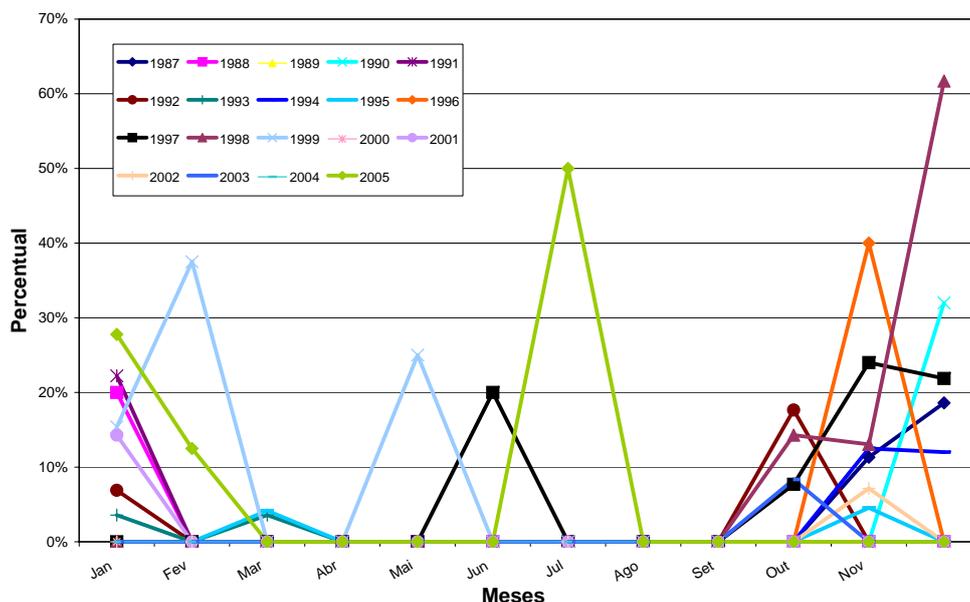


Figura 1.5: Percentual mensal de falhas devido à poluição nas linhas de transmissão da CHESF

## 1.2 Reconhecimento de Padrões

A análise dos dados colhidos através de sensores utilizados para diagnosticar o estado de isolantes é uma tarefa difícil que requer o suporte de ferramentas matemáticas e computacionais, uma vez que os dados a serem analisados se constituem de sinais, os quais possuem padrões específicos que requerem o tratamento através de técnicas tradicionalmente aceitas e que automatize o diagnóstico diante da necessidade de monitorações seguras em tempo real. Essas dificuldades têm conduzido a tratar o problema através de técnicas de reconhecimento de padrões [11].

Um sistema de reconhecimento de padrões [4] pode ser representado, simplificada, em diagrama de blocos, conforme ilustra a Figura 1.6, na qual o padrão de entrada (sinais adquiridos através de sensores de ultra-som, no caso deste trabalho) constitui o vetor  $r$  ( $r_1, r_2, \dots, r_M$ ), o qual é submetido à rotina de extração de atributos. Os atributos extraídos através da rotina possibilitam a formação do vetor  $x$  ( $x_1, x_2, \dots, x_N$ ) denominado vetor de atributos, o qual é submetido ao classificador que

propiciará a tomada de decisão necessária ao processo. Os atributos são aferições feitas sobre o padrão de entrada obtido, com o intuito de levantar características intrínsecas que viabilize a classificação dos padrões originais.

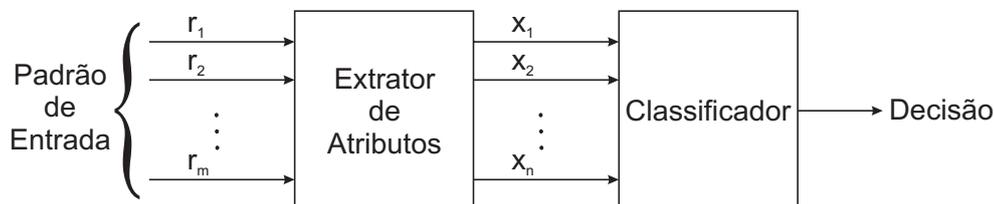


Figura 1.6: Reconhecimento de padrões

Os atributos podem ser simbólicos, numéricos ou ambos. As regras básicas estabelecidas [12] para a escolha dos atributos são:

- que sejam computacionalmente práticos;
- que conduzam a bons resultados no processo de reconhecimento de padrões;
- que reduzam o processamento a uma quantidade tratável de dados sem descartar informações que ainda incorporem valor.

O processo de seleção de atributos envolve uma análise crítica. Em alguns casos existem ferramentas matemáticas que ajudam na seleção. Em outros casos são tradicionalmente feitas simulações para identificar a contribuição final de cada atributo analisado. A construção do sistema de reconhecimento, ilustrado na Figura 1.6, requer, portanto, toda uma análise prévia no sentido de dotá-lo de condições necessárias a minimizar erros de decisão. Uma das etapas fundamentais está associada à caracterização dos atributos que compõem o vetor  $x$  [13]. Uma das contribuições deste trabalho é corroborar a citada etapa.

Para o classificador, cujo objetivo é distinguir os isoladores em grupos distintos de poluição, foi utilizada a técnica de Redes Neurais Artificiais (RNA). A Rede Neural apresenta uma capacidade de classificação de um padrão desconhecido. O referido classificador passa por uma etapa em que o número de acertos, para classes conhecidas, deve ser observado, validando ou não a rede analisada.

### 1.3 Sinopse dos Capítulos

O Capítulo 1 tem como principal objetivo descrever de forma sucinta e objetiva o propósito da dissertação apresentada. Que foca basicamente a detecção de níveis distintos de poluição em isoladores, através de sinais de ultra-som, utilizando como classificador Redes Neurais Artificiais. O referido capítulo também visa realizar uma abordagem da problemática existente, em empresa do ramo de energia elétrica, que envolva falhas em isoladores ocasionadas pelo elevado índice de poluição.

No Capítulo 2 será apresentada a parte conceitual e teórica de Redes Neurais e de técnicas de extração de atributos, utilizadas na dissertação.

O Capítulo 3 descreve, de forma breve, a metodologia experimental utilizada para aquisição dos sinais de entrada da Rede Neural. Posteriormente é apresentada a técnica de digitalização dos sinais de ultra-som, assim como do tratamento desses sinais, que constituirão a entrada da Rede. O tratamento de sinais citado anteriormente envolve atividades, tais como: suavização da curva, normalização e extração de atributos.

O Capítulo 4 abordará o procedimento utilizado para modelagem da Rede Neural Artificial. Dentre as etapas do procedimento citado podemos destacar a separação dos dados para treinamento e avaliação (Conjunto de validação, treinamento e teste), seleção da melhor representação dos padrões de saída, da melhor arquitetura e finalmente da melhor rede.

No Capítulo 5 serão mostrados os resultados obtidos, erro de classificação, das simulações das redes modeladas e apresentadas no capítulo anterior, visando a seleção da melhor rede para o problema proposto.

O Capítulo 6 destina-se a conclusões finais e propostas para trabalhos futuros.

# 2

## BASE TEÓRICA - REDES NEURAIS ARTIFICIAIS

Será apresentado no presente capítulo um resumo de conceitos e teorias de técnicas utilizadas como base no projeto realizado.

### 2.1 Reconhecimento de Padrões

A técnica de reconhecimento de padrões é definida, formalmente, como sendo o processo pelo qual um sinal recebido é atribuído a uma classe, dentre as classes anteriormente determinadas [28]. Uma Rede Neural realiza o reconhecimento de padrões passando inicialmente por uma etapa de treinamento, durante a qual se apresenta repetidamente à rede um conjunto de padrões de entrada junto com a categoria à qual cada padrão pertence. O número de padrões de treinamento para cada uma das classes deve ser idealmente o mesmo. Em uma outra etapa apresenta-se à rede um novo padrão que não foi visto antes, mas que pertence ao mesmo conjunto de padrões utilizado para treinar a rede. É esperado que a rede seja capaz de identificar a classe daquele padrão particular por causa da informação que ela extraiu dos dados de treinamento. Para que o desempenho do classificador possa ser avaliado, para novos padrões, o classificador passa por uma fase de teste, onde deve ser observado o número de acertos realizados para um conjunto de padrões cujas classes sejam conhecidas. Nenhum dos padrões do conjunto de teste pode ter sido utilizado no conjunto de

treinamento. O desempenho da rede é avaliado através do Erro de Classificação (EC), deve-se apresentar a proporção de padrões classificados corretamente e incorretamente, calculado da seguinte forma:

$$\begin{aligned}
 EC[\%] &= \frac{I}{N} \cdot 100 \\
 EC[\%] &= \frac{I}{C + I} \cdot 100
 \end{aligned}
 \tag{2.1}$$

Onde,

I → Padrões classificados incorretamente;

C → Padrões classificados corretamente;

N → Todos os padrões classificados.

A maioria do foco dado em pesquisas sobre Redes Neurais é em classificação de padrões [27]. Poderemos apresentar um classificador de padrões de forma sucinta e generalizada através do diagrama apresentado na Figura 2.1.

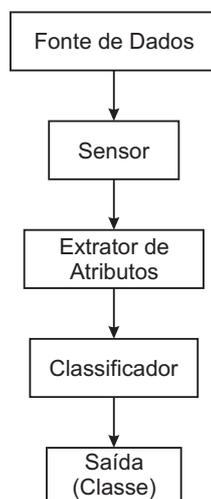


Figura 2.1: Diagrama de um classificador de padrões

O Erro Médio Quadrático (MSE - *Mean Square Error*), função de performance utilizada para redes *feedforward*, é a média dos erros ao quadrado entre a saída da rede

e a saída desejada.

$$MSE = \frac{1}{Q} \sum_{k=1}^Q \epsilon(k)^2 = \frac{1}{Q} \sum_{k=1}^Q (d(k) - y(k))^2 \quad (2.2)$$

Onde,

Q  $\rightsquigarrow$  Quantidade de exemplos;

$\epsilon$   $\rightsquigarrow$  Erro;

d  $\rightsquigarrow$  Saída desejada;

y  $\rightsquigarrow$  Saída da rede.

Como uma solução perfeita não existe é adotado como critério de parada do treinamento um valor do MSE. Neste momento a rede encontrará uma solução o mais próxima possível da saída desejada

## 2.2 Redes Neurais Artificiais

As Redes Neurais Artificiais são sistemas paralelos compostos por unidades de processamento simples, neurônios, que calculam determinadas funções matemáticas, lineares ou não-lineares [18]. Os referidos neurônios são dispostos em uma ou mais camadas, interligadas por um grande número de conexões. Na maioria dos modelos estas conexões estão associadas a pesos, os quais armazenam o conhecimento representado no modelo e servem para ponderar a entrada recebida por cada neurônio da rede. O funcionamento destas redes é inspirado em uma estrutura física concebida pela natureza: o cérebro humano, composto de aproximadamente 100 bilhões de neurônios biológicos [23].

O neurônio biológico, Figura 2.2, é, de forma simplista, a unidade básica da estrutura do cérebro. A membrana exterior de um neurônio é chamada dendritos que são dispositivos de entrada, cujo objetivo é conduzir os sinais das extremidades para o corpo celular. Por outro lado o axônio é o dispositivo de saída, que transmite o sinal do corpo celular para suas extremidades. As extremidades do axônio são

conectadas com os dendritos de outros neurônios através das sinapses. As sinapses possuem neurotransmissores que permitem a propagação dos pulsos nervosos, que podem ser excitatórias ou inibitórias. Em outras palavras o neurônio biológico é uma célula que ao ser estimulada em uma entrada gera uma saída que é enviada a outros neurônios. Essa saída irá depender da força de cada uma das entradas e da magnitude da conexão associada a cada entrada (sinapse). Uma sinapse pode excitar um neurônio, aumentando sua saída, ou inibi-lo, diminuindo sua saída.

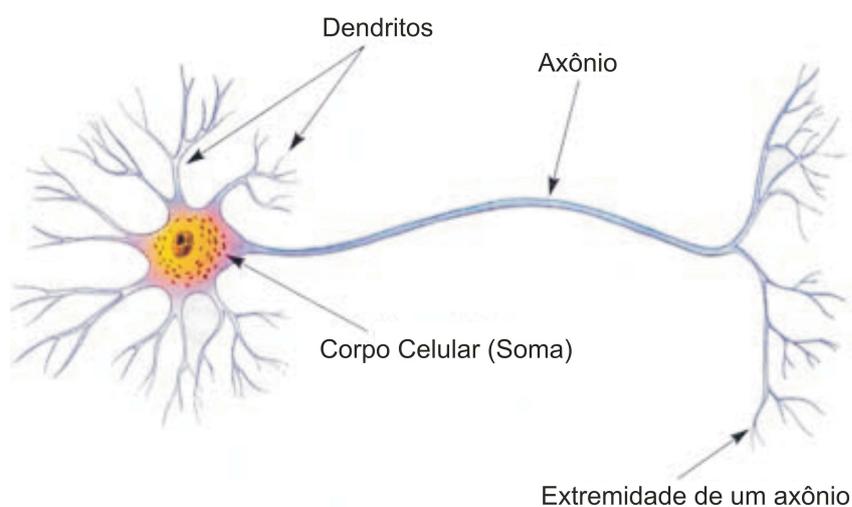


Figura 2.2: Neurônio biológico

Em 1943, Warren McCulloch e Walter Pitts publicaram um trabalho pioneiro no qual introduziram a idéia de redes neurais como máquinas computacionais. O neurônio de McCulloch e Pitts, Figura 2.3, foi proposto com pesos fixos, isto é, não-ajustáveis, e só conseguem implementar funções linearmente separáveis, limitando sua utilização [32] [24].

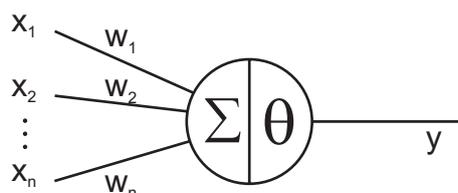


Figura 2.3: Neurônio de McCulloch e Pitts

Após várias pesquisas chegou-se a uma rede neural artificial constituída de vários neurônios artificiais que estão altamente conectados. O neurônio artificial apresentado na Figura 2.4 é composto por dois módulos de processamento. O primeiro módulo, denominado regra de propagação, executa uma soma ponderada das entradas  $x_i$  multiplicadas pelos pesos sinápticos,  $w_i$ , associados a cada entrada. O segundo módulo de processamento, denominado função de ativação, aplica uma função ao resultado da regra de propagação. O resultado da função de ativação é a saída  $y_k$  do neurônio artificial [17].

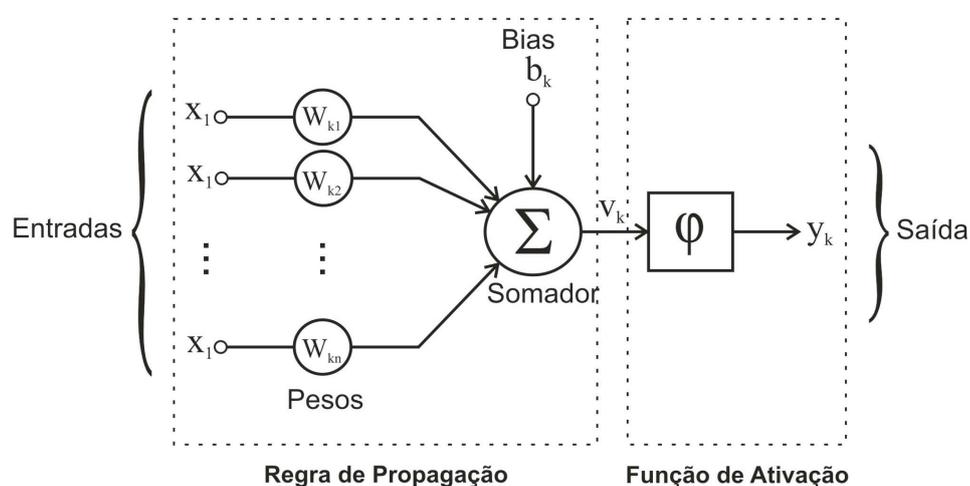


Figura 2.4: Rede neural artificial

### 2.2.1 Funções de Ativação

As funções de ativação normalmente são funções não-lineares, que transformam as redes neurais em sistemas computacionais não-lineares capazes de representarem problemas mais complexos. São apresentadas abaixo as funções de ativação mais utilizadas:

- Função Linear

$$f(x) = \alpha \cdot x \quad \forall \alpha \in \mathfrak{R} \quad (2.3)$$

Onde,

$\alpha \rightsquigarrow$  Define a inclinação da reta.

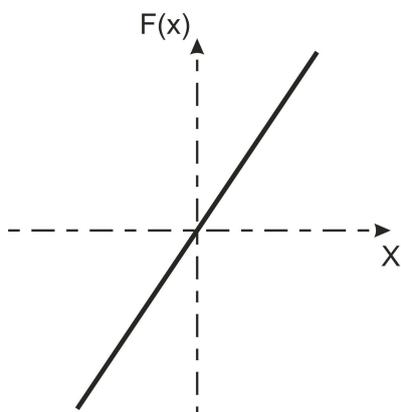


Figura 2.5: Função linear

- Função Rampa

$$f(x) = \begin{cases} +1, & \text{se } x \geq +\gamma \\ \alpha \cdot x, & \text{se } -\gamma < x < \gamma \\ 0 \text{ ou } -1, & \text{se } x \leq -\gamma \end{cases} \quad (2.4)$$

Onde,

$\gamma \rightsquigarrow$  Valor máximo/mínimo de saída;

$\alpha \rightsquigarrow$  Define a inclinação da reta.

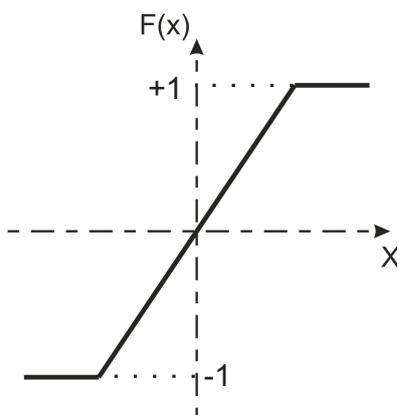


Figura 2.6: Função rampa

- Função Degrau

$$f(x) = \begin{cases} +1, & \text{se } x > 0 \\ 0 \text{ ou } -1, & \text{se } x \leq 0 \end{cases} \quad (2.5)$$

Onde,

$\gamma \rightsquigarrow$  Valor máximo/mínimo de saída.

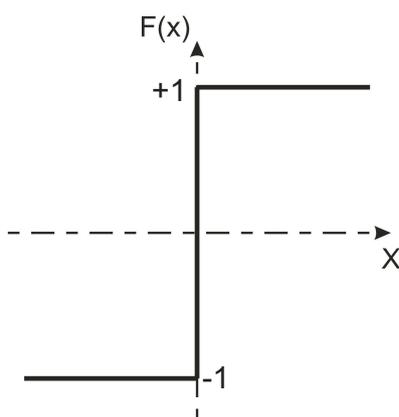


Figura 2.7: Função degrau

- Função Tangente Hiperbólica

$$f(x) = \tanh(x) \quad (2.6)$$

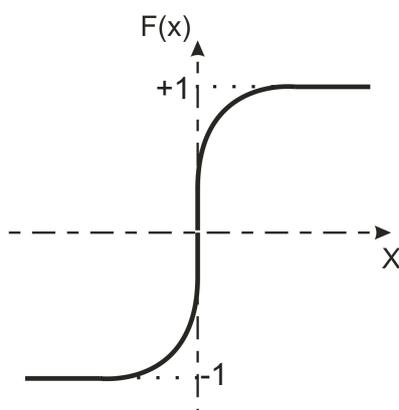


Figura 2.8: Função tangente hiperbólica

- Função Sigmóide Logística

$$f(x) = \frac{1}{1 + e^{-\frac{x}{T}}} \quad (2.7)$$

Onde,

$T \rightsquigarrow$  Parâmetro que determina a suavidade da curva.

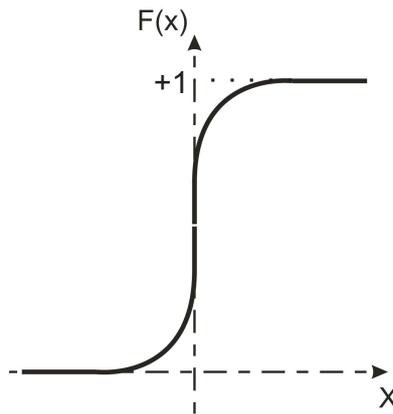


Figura 2.9: Função sigmóide logística

### 2.2.2 Perceptron e Adaline

Tanto o modelo Adaline como o modelo Perceptron seguem o princípio de comparação entre os resultados obtidos, através de operações de soma ponderada e um valor limiar (*threshold*), ou seja, o neurônio de McCulloch e Pitts. Basicamente percebe-se que as regras de aprendizado do Perceptron e do Adaline são idênticas, porém as equações de ajuste são obtidas após a aplicação da função de ativação e para a saída linear do nodo, respectivamente.

Para o Perceptron, [17] durante a fase de aprendizagem, o vetor peso é atualizado com um valor de incremento ( $\Delta w$ ), da seguinte forma:

$$w(t + 1) = w(t) + \Delta w, \quad (2.8)$$

que objetiva uma maior aproximação do resultado de saída da rede ao resultado

desejado. Logo, os algoritmos de aprendizagem visam uma técnica que otimize a obtenção desse incremento do vetor peso. Pode-se representar o fluxo de sinal do Perceptron através do grafo apresentado na Figura 2.10.

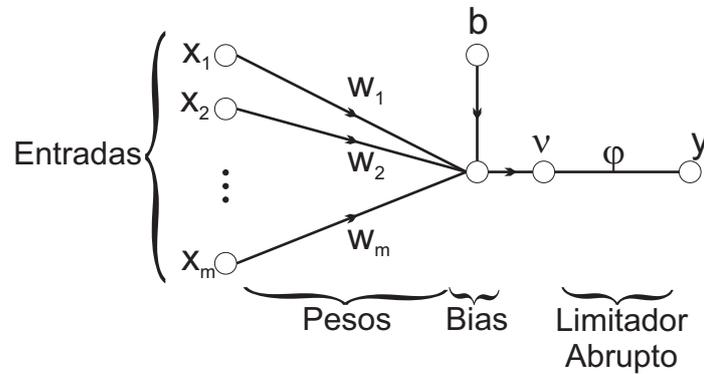


Figura 2.10: Grafo de fluxo de sinais do Perceptron

Logo, a função que descreve o grafo, Figura 2.10, é a seguinte:

$$v = \sum_{i=1}^m w_i \cdot x_i + b \quad (2.9)$$

O bias ( $b$ ) tem como objetivo deslocar a fronteira de decisão em relação à origem. O grafo de fluxo de sinais do Perceptron pode também ser representado como mostra a Figura 2.11, sendo apenas mais conveniente para trabalhar com a seguinte equação, equivalente a Equação 2.9:

$$v = \sum_{i=0}^m w_i \cdot x_i. \quad (2.10)$$

Nessa representação o bias é tratado como um peso sináptico acionado por uma entrada fixa igual a 1. Pode-se notar que o elemento  $w_0$  representa o bias.

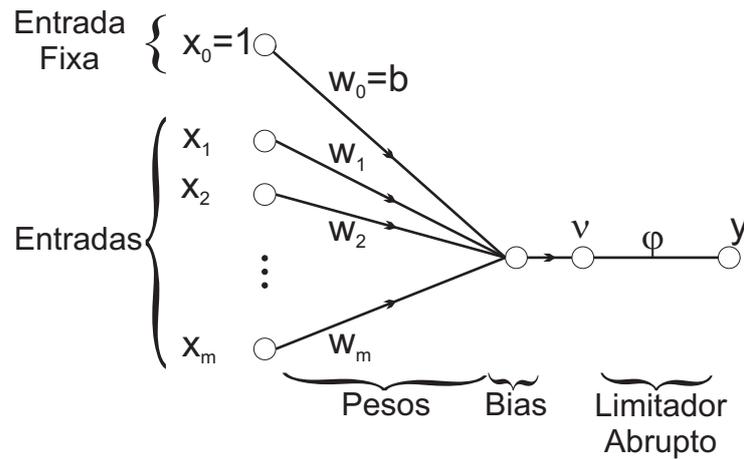


Figura 2.11: Grafo de fluxo de sinais do Perceptron

A regra de aprendizagem pela correção do erro pode ser descrita através da equação

$$w(n+1) = w(n) + \eta \cdot \epsilon(n) \cdot x(n). \quad (2.11)$$

Tem-se,

$$\epsilon(n) = d(n) - y(n). \quad (2.12)$$

Substituindo a Equação (2.12) na Equação (2.11) tem-se:

$$w(n+1) = w(n) + \eta \cdot [d(n) - y(n)] \cdot x(n). \quad (2.13)$$

Na condição da taxa de aprendizagem ( $\eta$ ) fixa teremos a chamada *Regra de Adaptação com Incremento Fixo*. O valor da taxa de aprendizagem deve ser positivo e menor ou igual a 1 ( $0 < \eta \leq 1$ ).

O modelo Adaline é baseado no algoritmo de treinamento do mínimo quadrado médio (*Least Mean Square* - LMS) que utiliza a função custo:

$$\xi(\omega) = \frac{1}{2} \cdot \epsilon^2(n), \quad (2.14)$$

que será diferenciada, através da regra da cadeia, em relação ao vetor peso

$$\frac{\delta\xi(\omega)}{\delta\omega(n)} = \epsilon(n) \cdot \frac{\delta\epsilon(n)}{\delta\omega(n)}. \quad (2.15)$$

Tem-se,

$$y = x^T(n) \cdot \omega(n), \quad (2.16)$$

substituindo na Equação 2.12 é encontrada:

$$\epsilon(n) = d(n) - x^T(n) \cdot \omega(n), \quad (2.17)$$

e derivando-a obteremos:

$$\frac{\delta\epsilon(n)}{\delta\omega(n)} = -x(n). \quad (2.18)$$

Logo, substituindo a equação encontrada acima na Equação (2.15), tem-se:

$$\frac{\delta\xi(\omega)}{\delta\omega(n)} = -x(n) \cdot \epsilon(n). \quad (2.19)$$

Para o referido algoritmo, a atualização do vetor peso é definida como sendo contrária ao vetor gradiente  $\left(\frac{\delta\xi(\omega)}{\delta\omega(n)}\right)$ , que sempre aponta para o valor máximo, ou seja:

$$\Delta\omega(n) \propto -\frac{\delta\xi(\omega)}{\delta\omega(n)} \quad (2.20)$$

$$\Delta\omega \propto x(n) \cdot \epsilon(n). \quad (2.21)$$

Finalmente pode-se concluir que a atualização do vetor peso é descrita da seguinte forma

$$\Delta\omega(n) = \eta \cdot x(n) \cdot \epsilon(n), \quad (2.22)$$

onde  $\eta$  é o parâmetro da taxa de aprendizagem. Onde, se escolhermos um valor muito pequeno para a taxa de aprendizagem o processo adaptativo progredirá lentamente. Então conclui-se que:

$$\omega(n+1) = \omega(n) + \Delta\omega(n) \quad (2.23)$$

$$\omega(n+1) = \omega(n) + \eta \cdot x(n) \cdot \epsilon(n) \quad (2.24)$$

O Adaline, que possui saídas binárias bipolares ( $y \in [-1, +1]$ ), tem seus pesos adaptados em função do erro de sua saída linear ( $y = \sum_i w_i \cdot x_i$ ), antes da aplicação da função de ativação.

### 2.2.3 Redes Perceptron de Múltiplas Camadas

A rede Perceptrons de Múltiplas Camadas (*Multilayer Perceptron* - MLP) é constituída da camada de entrada, da(s) camada(s) oculta(s) e da camada de saída. A rede MLP contém uma ou mais camadas de neurônios ocultos, que não são a camada de entrada nem a camada de saída da rede [33]. Esses neurônios ocultos capacitam a rede a aprender tarefas complexas extraindo progressivamente as características mais significativas dos padrões de entrada.

As redes MLP têm sido muito utilizadas para resolver diversos problemas de difícil solução, através do seu treinamento de forma supervisionada, com o algoritmo de retropropagação do erro (*Error Back-Propagation*) que é baseado na regra de aprendizagem por correção do erro.

As redes neurais utilizadas neste trabalho são de múltiplas camadas (MLP, *MultiLayer Perceptron*), esse tipo de rede é constituído de uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída, Figura 2.12.

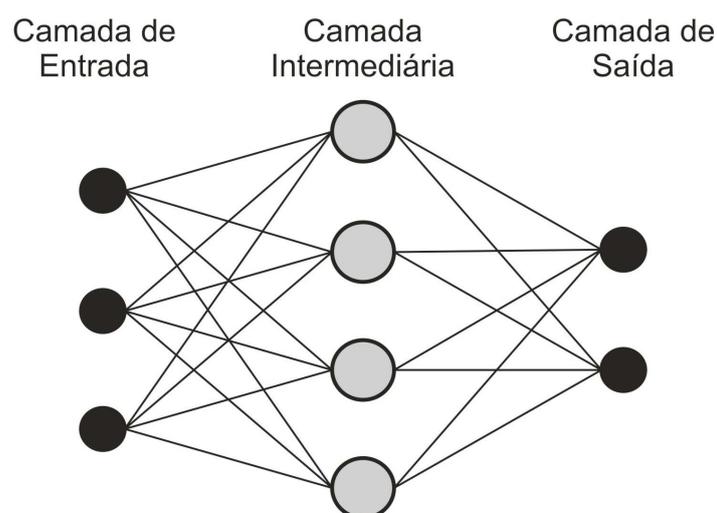


Figura 2.12: Rede neural múltiplas camadas - MLP

Existem atualmente vários algoritmos para treinar redes MLP, esses algoritmos são geralmente do tipo supervisionado [18]. De acordo com os parâmetros que eles atualizam, os algoritmos para treinamento das redes do tipo MLP podem ser classificados em:

- Estáticos
- Dinâmicos

Enquanto os algoritmos estáticos não alteram a estrutura da rede, variando apenas os valores de seus pesos, os algoritmos dinâmicos podem tanto diminuir como aumentar o tamanho da rede (número de camadas, número de neurônios nas camadas ocultas e número de conexões). Quando o aprendizado estático é utilizado, a mesma regra de aprendizado é empregada para redes do tipo MLP com diferentes tamanhos e formatos. É interessante observar que topologias diferentes podem resolver o mesmo problema.

O algoritmo *back-propagation* é um algoritmo supervisionado que utiliza pares, entrada - saída desejada, para, por meio de um mecanismo de correção de erros, ajustar os pesos da rede. O treinamento ocorre em duas fases, em cada fase percorre a rede em um sentido. Essas duas fases são chamadas de fase *forward* e fase *backward*. A fase *forward* é utilizada para definir a saída da rede para um dado padrão de entrada ,

onde o vetor de entrada que é aplicado aos nodos da rede e propaga seu efeito camada por camada, nessa etapa todos os pesos sinápticos da rede são fixos. A fase *backward* utiliza a saída desejada e a saída fornecida pela rede para atualizar os pesos de suas conexões, onde a resposta real da rede é subtraída da resposta desejada para produzir um sinal de erro. Esse sinal de erro é então propagado para trás através da rede, contra a direção das conexões sinápticas ajustando os pesos sinápticos objetivando que a resposta real da rede se aproxime da resposta desejada. Como ilustrado na rede MLP na Figura 2.13.

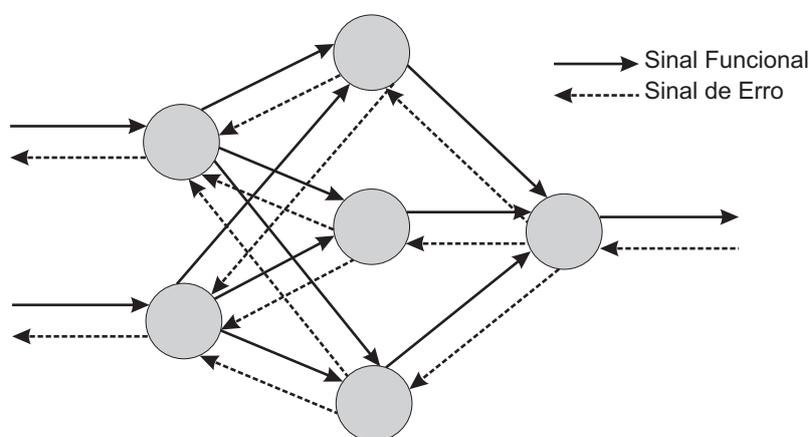


Figura 2.13: Propagação e retropropagação de sinais na rede

#### 2.2.4 Algoritmo do Gradiente Descendente

O algoritmo do gradiente descendente (GD) é baseado no algoritmo de retropropagação de erro (criado a partir de uma generalização da regra delta proposta por Widrow e Hoff em 1960) sendo composto por duas fases: a fase *forward*, na qual um padrão de dados é apresentado à rede e esta processa os dados, produzindo respostas, e a fase *backward*, que utiliza o erro obtido entre o resultado da execução *forward* e o resultado desejado conhecido para determinar os ajustes a serem feitos nos pesos, dos neurônios da rede.

### 2.2.5 Algoritmo Resilient Propagation

O algoritmo *Resilient Propagation* (Rprop) [18] possui uma técnica de aprendizagem eficiente. O Rprop apresenta uma adaptação direta dos pesos baseada na informação do gradiente local. A adaptação do Rprop não é interferida pelo comportamento da magnitude do gradiente da função erro.

Este algoritmo objetiva a eliminação da influência negativa do valor da derivada parcial na definição do ajuste dos pesos. A referida influência negativa ocorre porque quando a saída de um nodo for próxima de 0 e a saída desejada for próxima de 1, ou vice-versa, a derivada será próxima de zero, fazendo com que os pesos deste nodo recebam um ajuste mínimo. No caso do algoritmo Rprop não se tem este problema, pois o mesmo utiliza apenas o sinal da derivada para o ajuste dos pesos, indicando simplesmente a direção do ajuste, e não o valor. Esse valor da variação dos pesos ( $\Delta_{ji}$ ) é:

$$\Delta\omega_{ji} = \begin{cases} -\Delta_{ji}, & \text{se } \frac{\delta\epsilon}{\delta\omega_{ji}} > 0 \\ +\Delta_{ji}, & \text{se } \frac{\delta\epsilon}{\delta\omega_{ji}} < 0 \\ 0, & \text{se } \frac{\delta\epsilon}{\delta\omega_{ji}} = 0 \end{cases} \quad (2.25)$$

O valor ( $\Delta_{ji}$ ) é definido por um processo de adaptação que depende do sinal da derivada do erro com relação ao peso a ser ajustado da seguinte forma:

$$\Delta_{ji}(t) = \begin{cases} \eta^+ \cdot \Delta_{ji}(t-1), & \text{se } \frac{\delta\epsilon(t-1)}{\delta\omega_{ji}} \cdot \frac{\delta\epsilon(t)}{\delta\omega_{ji}} > 0 \\ \eta^- \cdot \Delta_{ji}(t-1), & \text{se } \frac{\delta\epsilon(t-1)}{\delta\omega_{ji}} \cdot \frac{\delta\epsilon(t)}{\delta\omega_{ji}} < 0 \\ \Delta_{ji}(t-1), & \text{Caso contrário} \end{cases} \quad (2.26)$$

Onde  $0 < \eta^- < 1 < \eta^+$

### 2.2.6 Algoritmo Levenberg-Marquardt

Esse algoritmo é considerado o método mais rápido para treinamento de redes *feedforward backpropagation* [32], desde que a rede possua uma quantidade moderada de

pesos sinápticos a fim de tornar o problema exequível. Para acelerar o treinamento da rede, esse algoritmo baseia-se na determinação de derivadas de segunda ordem do erro quadrático em relação aos pesos, diferindo do algoritmo *backpropagation* tradicional que considera derivadas de primeira ordem. Este método apresenta convergência em menos iterações, mas requer mais cálculos por iteração devido ao cálculo de matrizes inversas.

O referido algoritmo se baseia no método de otimização de Newton. Será necessário para a utilização deste algoritmo o cálculo da matriz Hessiana:

$$[H] = \begin{bmatrix} \delta^2 \varepsilon \\ \delta \omega_{kj}^2 \end{bmatrix} \quad (2.27)$$

$$\varepsilon = \frac{1}{2} \sum_{k=1}^m \epsilon_k^2 \quad (2.28)$$

Onde  $e_k$  é o erro do neurônio  $k$ .

A matriz Hessiana pode ser obtida a partir da matriz Jacobiana como segue:

$$[H(\omega)] = [J(\omega)]^T \cdot [J(\omega)] \quad (2.29)$$

Onde:

$$[J] = \begin{bmatrix} \delta \varepsilon \\ \delta \omega_{kj} \end{bmatrix} \quad (2.30)$$

Os pesos, para o método de Newton, são atualizados da seguinte forma:

$$\omega_{k+1} = \omega_k + [H]^{-1} q_k \quad (2.31)$$

Onde:

$$q_k = 2[J(\omega)]^T \varepsilon \quad (2.32)$$

Apesar do grande esforço computacional, ele segue sendo o algoritmo de treinamento mais rápido para redes neurais, quando se trabalha com um número moderado de parâmetros na rede. Se esse número é elevado, a utilização desse algoritmo é pouco prática.

## 2.3 Redes de Funções de Base Radiais

As redes de Funções de Base Radiais (RBF) são redes multicamadas em que a ativação de um nodo pode ser função da distância entre seus vetores de entrada e de peso [18]. Diferentemente das funções de ativação aplicadas na maioria das redes multicamadas onde utiliza-se como argumento o produto escalar do vetor de entrada e do vetor de pesos deste nodo, como mostra a Figura 2.14. Os neurônios da camada oculta utilizam funções de base radiais. Também foram utilizadas redes de funções de bases radiais como classificador no presente trabalho.

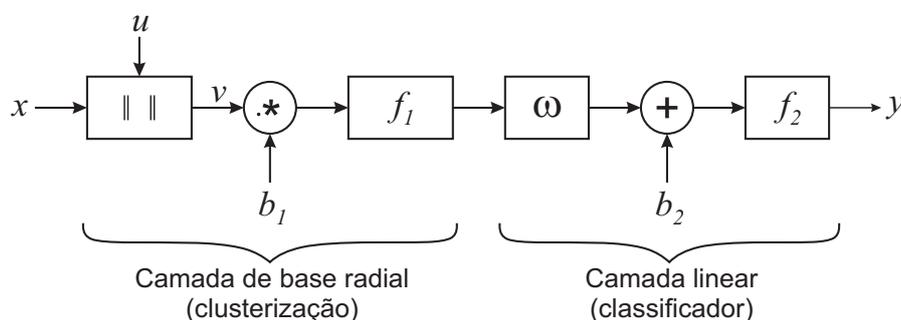


Figura 2.14: Arquitetura de redes RBF

Na Figura 2.14 tem-se:

$v = \|x - u\| \rightsquigarrow$  Distância euclidiana;

$x \rightsquigarrow$  Entrada;

$u \rightsquigarrow$  Vetor centro da função radial;

$.* \rightsquigarrow$  Multiplicação elemento a elemento;

$b \rightsquigarrow$  Bias da primeira camada;

$f_1 \rightsquigarrow$  Função de base radial;

$\omega_1 \rightsquigarrow$  Pesos;

$b \rightsquigarrow$  Bias da segunda camada;

$f_2 \rightsquigarrow$  Função linear;

$\sigma \rightsquigarrow$  Largura da função radial;

Funções radiais têm como característica peculiar a redução ou a ampliação do valor da função em relação à distância de um ponto central. Diferentes funções de base radiais têm sido utilizadas em redes RBF [18]. As mais comuns são:

- Função Gaussiana:

$$f(u) = \exp\left(-\frac{v^2}{2\sigma^2}\right) \quad (2.33)$$

- Função Multiquadrática:

$$f(u) = \sqrt{v^2 + \sigma^2} \quad (2.34)$$

- Função thin-plate-spline:

$$f(u) = v^2 \log(v) \quad (2.35)$$

O treinamento de redes RBF é classificado como híbrido, que geralmente requer um número de neurônios ocultos maior que o treinamento supervisionado [16].

As redes RBF constroem aproximadores locais enquanto as MLP, aproximadores globais. Ou seja, as redes MLP têm uma maior capacidade de generalização para regiões onde não há dados de treinamento.

### 2.3.1 Rede Neural Probabilística

A rede neural probabilística [20] é composta de duas camadas. A primeira calcula as distâncias entre os vetores de entrada e de treinamento e produz um vetor cujos elementos indicam quão próxima a entrada está dos dados de treinamento e a segunda, soma essas contribuições para cada classe de entradas para produzir como saídas da rede um vetor de probabilidades. Finalmente, uma função de transferência compete na saída da segunda camada, escolhendo a máxima dessas probabilidades simbolizando a referida classe com um sinal de saída unitário (1), e nulo (0) para as demais. A rede

neural probabilística pode ser utilizada em problemas de classificação. A arquitetura para este sistema é apresentada da Figura 2.15.

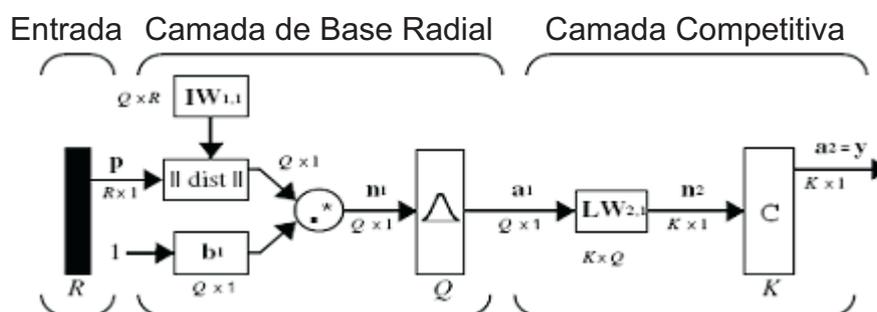


Figura 2.15: Arquitetura da rede neural probabilística

### 2.3.2 Rede RBF Eficiente

Uma rede de base radial pode ser utilizada para aproximar funções [20]. Este tipo de rede adiciona gradativamente neurônios na camada da base radial até atingir um valor de erro médio quadrático suficientemente pequeno. Caso contrário um próximo neurônio será adicionado. Esse procedimento é repetido até o erro objetivo ser satisfeito ou ter sido atingido o número máximo de neurônios.

Quanto menor for o valor do bias da camada de Base Radial mais suave será a função de aproximação. É preferível que a rede de base radial eficiente seja simulada diversas vezes, com valores de bias diferentes, visando encontrar o valor que tenha a melhor performance para o problema estudado.

A função de base radial cria uma rede de duas camadas. Onde na primeira, são utilizados neurônios com funções de Base Radial, cujo objetivo é agrupar os dados de entradas em conjuntos homogêneos com características semelhantes transformando um conjunto não linearmente separável em um conjunto linearmente separável. A camada de saída tem o papel de classificador.

### 2.3.3 Rede RBF Exata

A diferença entre as redes RBF eficiente e a RBF exata é que a primeira cria um neurônio por vez [20]. Já a segunda, RBF exata, cria uma quantidade de neurônios de base radiais igual à quantidade de vetores de entrada podendo então produzir uma rede com erro zero para vetores de treinamento. A rede RBF exata tem como alimentação a matriz de entrada e o vetor da saída desejada, sendo criada uma rede com todos os parâmetros ajustados onde as saídas da rede são exatamente os valores das saídas desejadas para o conjunto de treinamento.

### 2.3.4 Síntese do Capítulo

Neste capítulo foram apresentados conceitos e teorias de técnicas utilizadas para elaboração do projeto proposto. Além da base teórica de Redes Neurais Artificiais foi abordado também o conceito teórico de técnicas de Reconhecimento de Padrões, que é o processo pelo qual um sinal recebido é atribuído a uma classe.

# 3

## TRATAMENTO DOS SINAIS DE ENTRADA

No atual capítulo será abordada a metodologia experimental utilizada para aquisição dos sinais de entrada da Rede Neural, realizada por [4] no laboratório de alta tensão da UFCG, que visa simular as condições reais as quais os isoladores são submetidos. Em seguida serão apresentadas técnicas de tratamento dos sinais de ultra-som coletados dos ensaios realizados que comporão a entrada da Rede Neural Artificial. O referido tratamento de sinais envolve as seguintes atividades: digitalização de sinais, normalização e extração de atributos.

### 3.1 Montagens Experimentais

Foram simuladas, [4], condições de poluição na superfície dos isoladores sob o nível de tensão operacional visando detectar os sinais de ultra-som referentes a determinados níveis de poluição. Os equipamentos utilizados foram:

- Sensor de Ultra-som;

Visa captar os sinais emitidos pelas descargas elétricas na superfície do isolador sob ensaio.

- Sistema de Digitalização;

Visa converter um sinal analógico em um sinal digital discreto.

- Fonte de alta tensão;

Visa submeter o isolador sob ensaio ao nível de tensão operacional.

- Câmara de névoa.

Visa simular as condições de umidade de um ambiente real.

### 3.2 Constituição das Amostras e Estabelecimento das Classes

Durante os ensaios foram obtidos 100 padrões entrada-saída de cada classe, totalizando então 600 padrões entrada-saída. Onde cada padrão de entrada é composto por 32768 pontos, equivalentes aos ruídos de ultra-som gerados durante um período de 1 ciclo (1/60 segundos) a uma taxa de amostragem de 1.966.000 registros por segundo.

$$Dados = \begin{bmatrix} [Classe A]_{100 \times 32768} \\ [Classe B]_{100 \times 32768} \\ [Classe C]_{100 \times 32768} \\ [Classe D]_{100 \times 32768} \\ [Classe E]_{100 \times 32768} \\ [Classe F]_{100 \times 32768} \end{bmatrix}_{600 \times 32768} \quad (3.1)$$

Os dados originais de entrada apresentam este perfil, Figura 3.2.

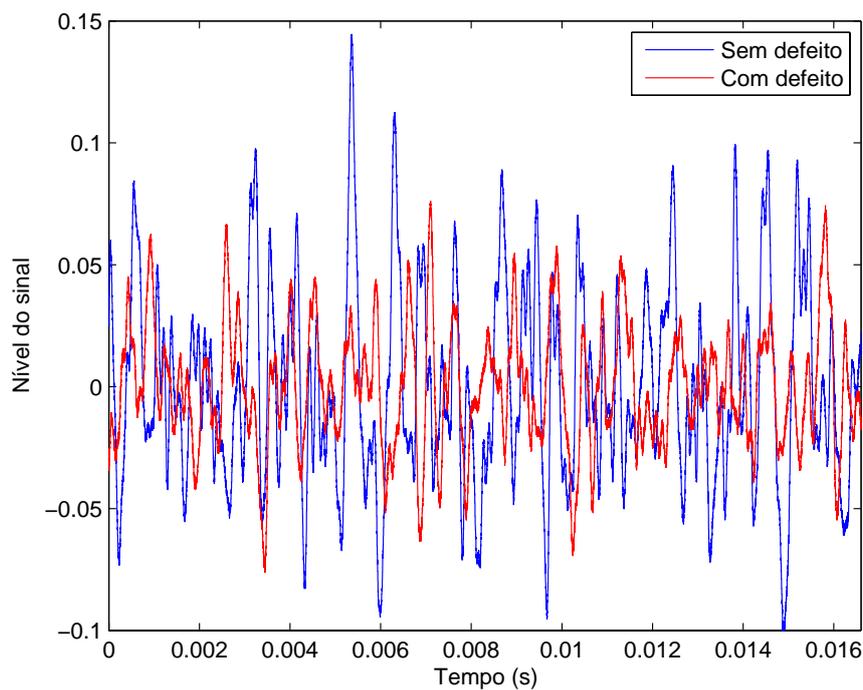


Figura 3.1: Exemplificação de um sinal de ultra-som

Logo, pode-se observar que os sinais de classes extremamente distintas possuem características visuais semelhantes, sendo então necessária a extração de atributos peculiares.

### 3.3 Digitalização de Sinais

A Figura 3.2 ilustra de forma sucinta o processo de digitalização dos sinais de ultra-som sob análise.



Figura 3.2: Diagrama do sistema de digitalização de sinais

Onde:

Ultra-som  $\rightsquigarrow$  Sinal de ultra-som com frequências superiores a 20kHz;

$S_1 \rightsquigarrow$  Sinal de ultra-som na faixa de 35kHz a 45kHz;

$S_2 \rightsquigarrow$  Sinal audível na faixa de frequência de 20Hz a 20kHz;

$S_3 \rightsquigarrow$  Sinal discretizado do sinal  $S_2$  a uma frequência de amostragem de 1,96MHz;

### 3.3.1 Amostragem

O sinal de ultra-som coletado é um sinal analógico, ou seja, contínuo no tempo que será discretizado. De acordo com o Teorema de Nyquist, a quantidade de amostras por unidade de tempo de um sinal, chamada frequência de amostragem, deve ser maior que o dobro da maior frequência do sinal a ser amostrado, para que possa ser reproduzido integralmente sem erro de *aliasing* [35]. A metade da frequência de amostragem é conhecida como a frequência de Nyquist e corresponde ao limite máximo de frequência do sinal que pode ser reproduzido.

$$f_{amostragem} > 2 \cdot f_{sinal} \quad (3.2)$$

Pode-se ilustrar as consequências ocorridas pela não satisfação da Equação 3.2 através da Figura 3.3, onde foi utilizado como sinal amostrado uma senóide.

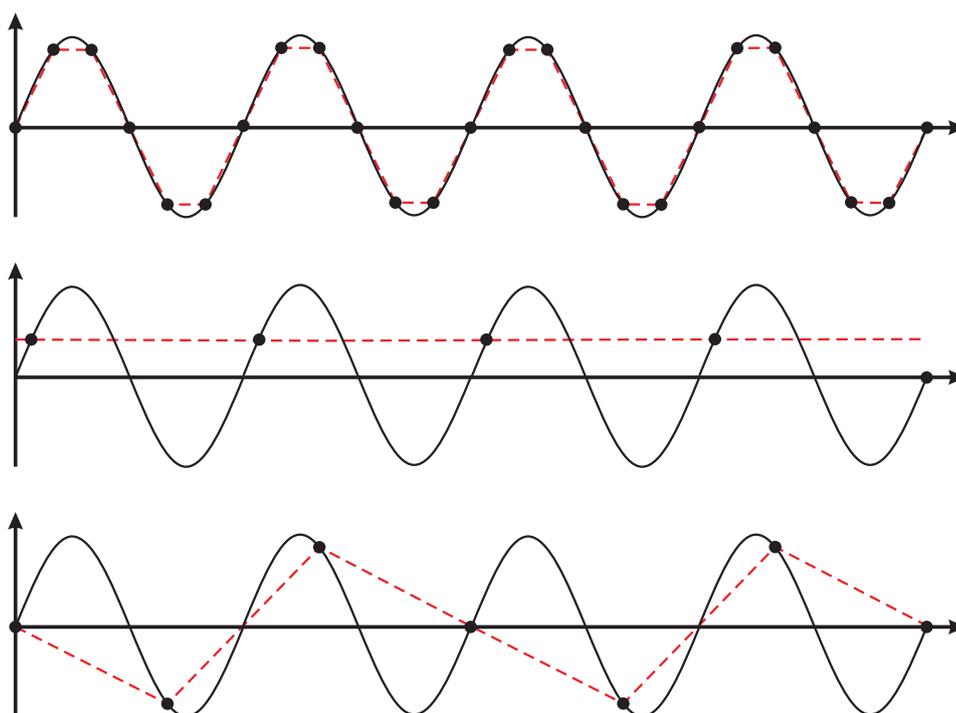


Figura 3.3: Exemplificação de *aliasing*

São apresentadas três situações distintas na Figura 3.3:

$$f_{amostragem} > 2 \cdot f_{sinal}$$

$$f_{amostragem} = 2 \cdot f_{sinal} \quad (3.3)$$

$$f_{amostragem} < 2 \cdot f_{sinal} \quad (3.4)$$

A primeira situação possui uma frequência de amostragem superior à frequência do sinal, assegurando uma reprodução fidedigna do sinal amostrado, como assegura a Equação 3.2. A segunda e a terceira situação não apresentam no sinal digitalizado as características reais do sinal amostrado, por não obedecer a equação de Nyquist, Equação 3.2.

A frequência do sinal audível está na faixa entre 20Hz e 20kHz, ou seja, para satisfazer o teorema de Nyquist para toda faixa de frequência do sinal de audio, deve-se que obedecer a Equação 3.2. Onde  $f_{sinal}$  deverá ser a maior frequência apresentada pelo sinal. No caso dos sinais estudados  $f_{sinal} > 20kHz$ . Logo,

$$\begin{aligned}
 f_{amostragem} &> 2 \cdot 20 \cdot 10^3 \\
 f_{amostragem} &> 40kHz
 \end{aligned}
 \tag{3.5}$$

### 3.4 Normalização

Pode-se descrever de modo claro e objetivo a normalização como sendo uma técnica utilizada para assegurar que a amplitude de determinados atributos estarão compreendidos entre valores previamente estabelecidos.

$$x_{norm} = \frac{a(x_{max} - x) - b(x_{min} - x)}{x_{max} - x_{min}},
 \tag{3.6}$$

onde  $x_{max}$  e  $x_{min}$  são os valores máximo e mínimo, respectivamente, do vetor  $x$ .

Observando que no trabalho proposto as magnitudes das variáveis de entrada da rede são muito diferentes. Logo realizou-se a normalização dos dados de entrada redimensionando-os em um intervalo compreendido entre 0,1 a 0,9. Foi utilizada a expressão de normalização, Equação 3.6, com os valores de  $a = 0,1$  e  $b = 0,9$ , que leva a:

$$x_{norm} = \frac{0,1(x_{max} - x) - 0,9(x_{min} - x)}{x_{max} - x_{min}}
 \tag{3.7}$$

### 3.5 Extração de Atributos

A extração de atributos tem como principal objetivo extrair as características mais importantes de um sinal, evidenciando as diferenças e similaridade entre os padrões de entrada das diferentes classes. Com a utilização do conjunto de entrada apenas normalizado tornou-se inviável a construção de uma rede neural, por apresentar vetores entrada-saída muito extensos, e ocorrer sempre um estouro de memória do computador, logo foi necessária a utilização de alguma técnica de extração de atributos, com o

objetivo de reduzir o sinal de entrada. Foram realizados testes com diversos extratores de atributos: análise dos sinais no domínio da frequência, análise do centróide e outros parâmetros estatísticos dos sinais no domínio do tempo, que serão apresentados neste capítulo.

### 3.5.1 Centróide

O uso da técnica do Centróide possibilitou compactar a informação do sinal de ultra-som, que antes era formado por 32768 pontos em um novo vetor com apenas 9 pontos. Esse novo vetor é composto por 6 entradas formadas pelas coordenadas do eixo das abscissas e ordenadas dos centróides superior, inferior e total do sinal, e 3 entradas formadas pelos valores das áreas superior, inferior e total do sinal. Como ilustra na Figura 3.4.

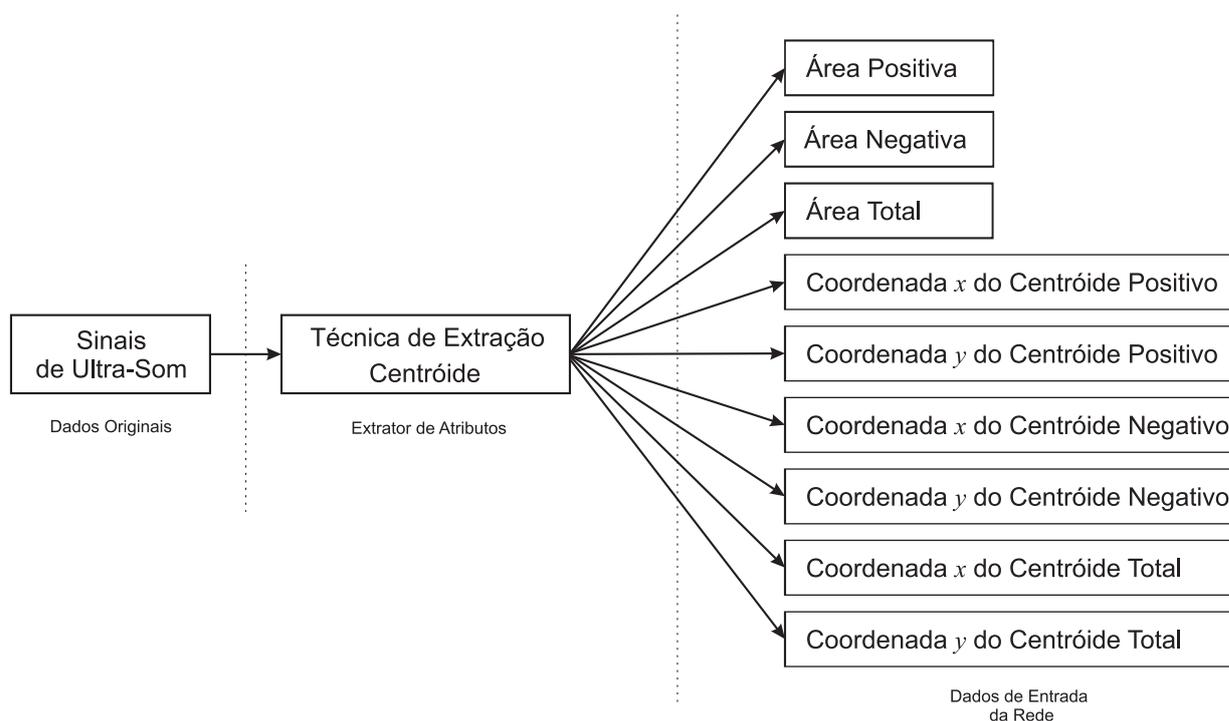


Figura 3.4: Extração de atributos do centróide

- Área Positiva

Tem como propósito calcular toda a área que se encontra sobre o eixo das abscissas. Como exemplificado na Figura 3.5.

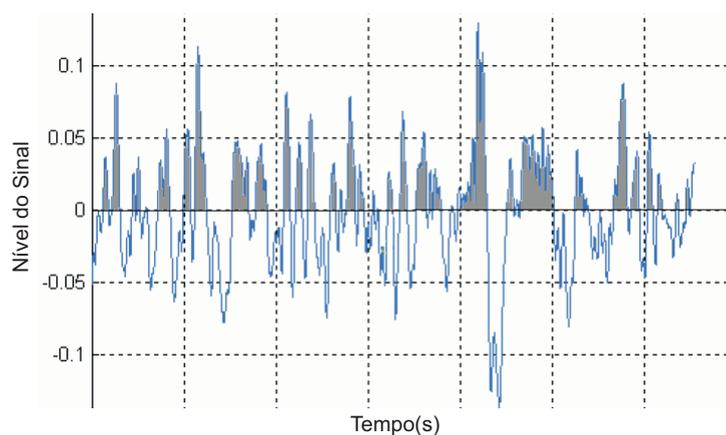


Figura 3.5: Área positiva do Centróide

- Área Negativa

Tem como propósito calcular toda a área que se encontra abaixo o eixo das abscissas. Como exemplificado na Figura 3.6.

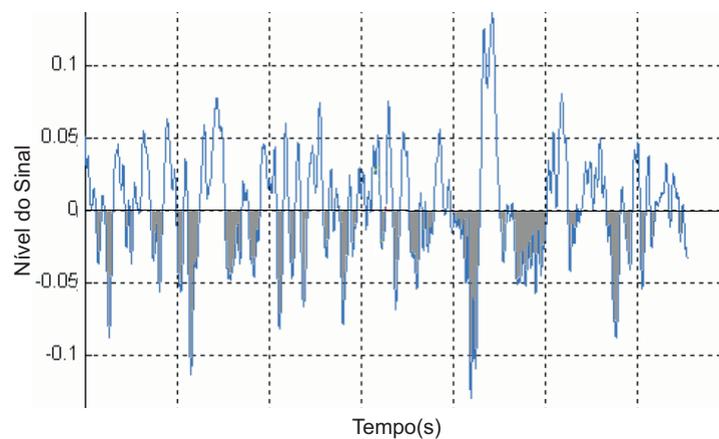


Figura 3.6: Área negativa do Centróide

- Área Total

A área total é o somatório das áreas positiva e negativa.

- Coordenada  $x$  do Centróide Positivo

A referida entrada da rede é a coordenada do eixo  $x$  do centróide calculado, levando apenas em consideração a parte positiva do sinal sob estudo. Como mostrado na Figura 3.7(a).

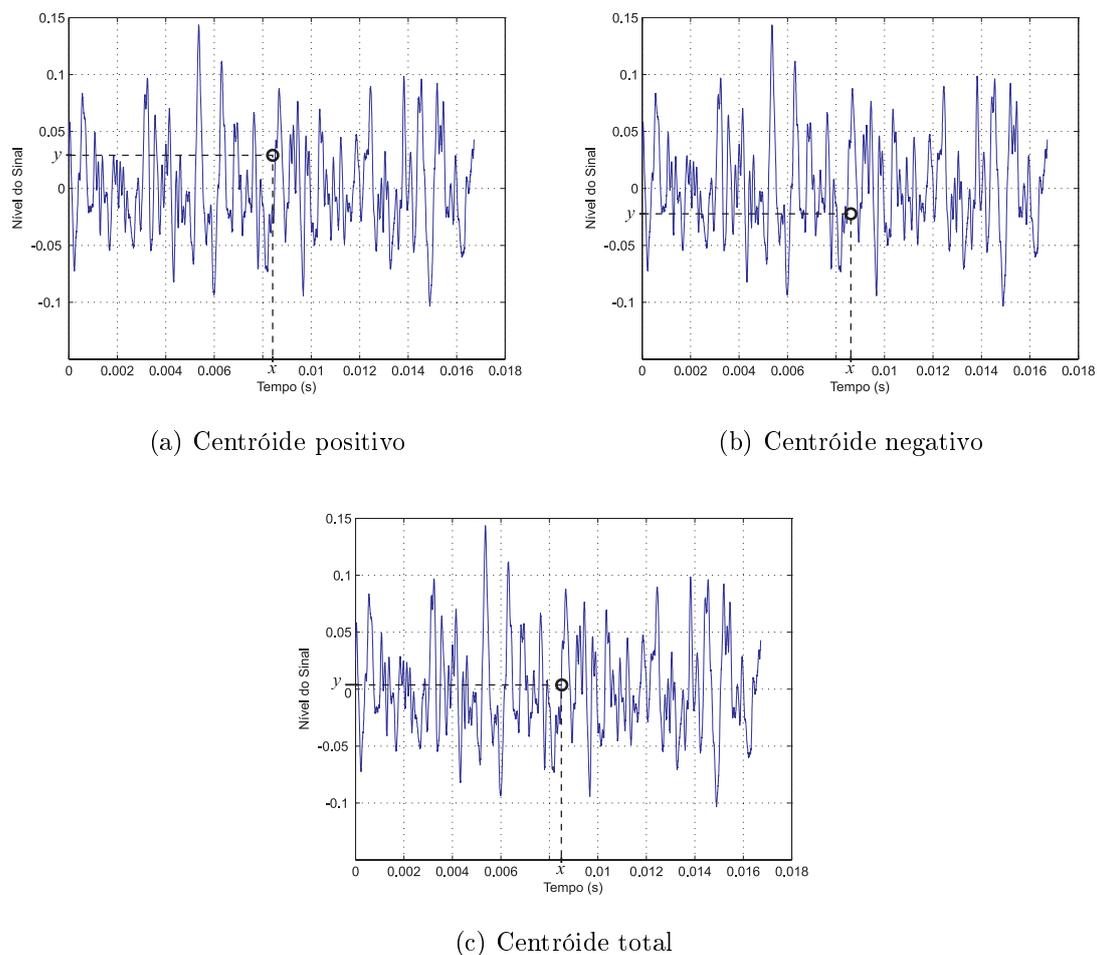


Figura 3.7: Exemplificação do cálculo do centróide

- Coordenada  $y$  do Centróide Positivo  
A referida entrada da rede é a coordenada do eixo  $y$  do centróide calculado, levando apenas em consideração a parte positiva do sinal sob estudo. Como mostrado na Figura 3.7(a).
- Coordenada  $x$  do Centróide Negativo  
A referida entrada da rede é a coordenada do eixo  $x$  do centróide calculado, levando apenas em consideração a parte negativa do sinal sob estudo. Como mostrado na Figura 3.7(b).
- Coordenada  $y$  do Centróide Negativo  
A referida entrada da rede é a coordenada do eixo  $y$  do centróide calculado, levando

apenas em consideração a parte positiva do sinal sob estudo. Como mostrado na Figura 3.7(b).

- Coordenada  $x$  do Centróide Total

A referida entrada da rede é a coordenada do eixo  $x$  do centróide calculado de todo o sinal sob estudo. Como mostrado na Figura 3.7(c).

- Coordenada  $y$  do Centróide Total

A referida entrada da rede é a coordenada do eixo  $x$  do centróide calculado levando apenas em consideração a parte positiva do sinal sob estudo. Como mostrado na Figura 3.7(c).

### 3.5.2 Parâmetros Estatísticos

Os atributos que são extraídos para compor o vetor de atributos trata-se, portanto, de parâmetros estatísticos que buscam caracterizar cada classe. O uso desta técnica possibilitou compactar a informação do sinal de ultra-som, que antes era formado por 32768 pontos, em um novo vetor constituído por apenas 28 pontos. Esse novo vetor é composto por 1 entrada formada pelas médias, 1 pela curtose, 1 pela assimetria, 1 pela passagens por zero, 1 pelo desvio padrão, 1 pelo número de valores negativos, 1 pelo valor eficaz, 1 pelo coeficiente de variação e 20 pelas faixas do histograma [31]. Esses atributos são apresentados na Figura 3.9.

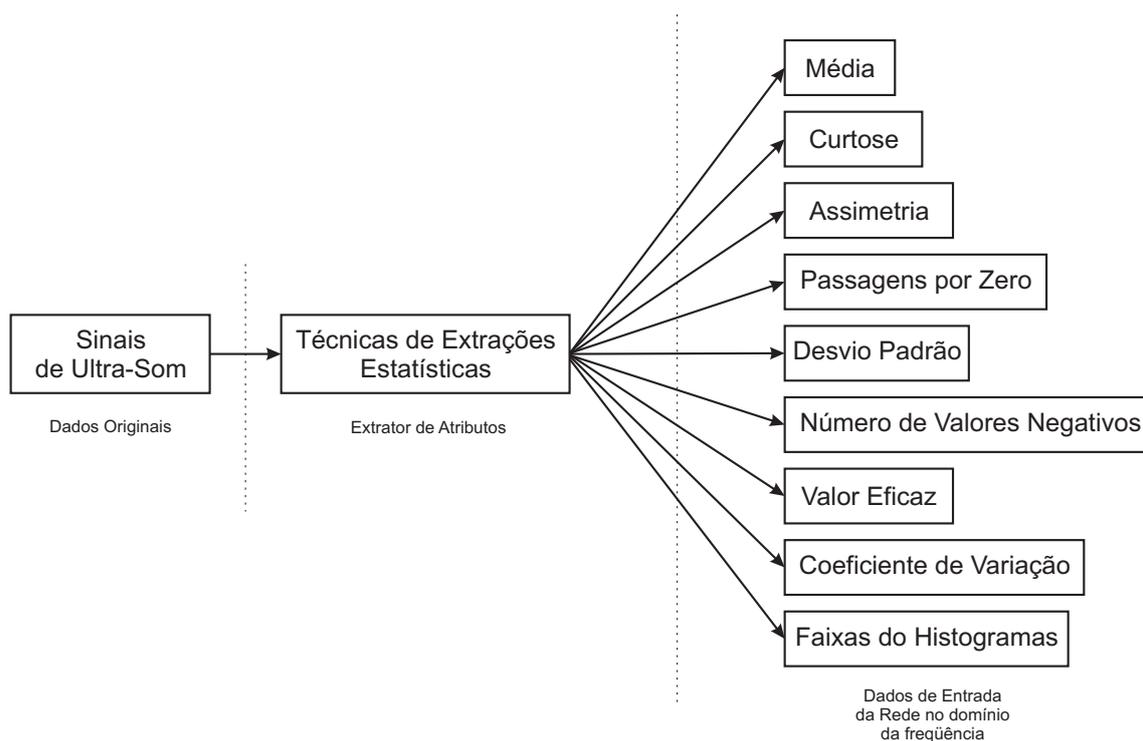


Figura 3.8: Extração de atributos estatísticos

Esses atributos estatísticos foram citados abaixo com uma descrição sucinta de cada:

- Média;

A média de um vetor  $x$  é calculada através da seguinte fórmula:

$$\mu = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$$\mu = \frac{\sum_{i=1}^n x_i}{n}, \quad (3.8)$$

onde  $x = (x_1, x_2, x_3, \dots, x_n)$ .

- Curtose;

É o grau de achatamento de uma distribuição, considerada em relação a uma distribuição normal. Onde, a curtose da distribuição normal é três. A curtose ( $k$ )

de uma distribuição é definida como mostrado na equação 3.9, mostrada abaixo.

$$k = \frac{E(x - \mu)^4}{\sigma^4} \quad (3.9)$$

Onde:

$\sigma$   $\rightsquigarrow$  é o desvio padrão de x;  $E(t)$   $\rightsquigarrow$  é o valor esperado.

- Assimetria;

Assimetria é a medida da desigualdade dos dados com relação à média da amostra. Se a assimetria é negativa, os dados são dispersos mais à esquerda da média do que para direita. Se a assimetria é positiva, os dados são dispersos mais à direita. A assimetria da distribuição normal (ou alguma distribuição perfeitamente simétrica) é zero. A assimetria ( $\Lambda$ ) da distribuição é definida como:

$$\Lambda = \frac{E(x - \mu)^3}{\sigma^3} \quad (3.10)$$

- Passagens pelo zero;

Indica quantas vezes o sinal cruzou o eixo das abscissas, ou seja, passou pelo zero.

A contabilização das passagens pelo zero deverá obedecer às seguintes regras:

- $x(t) < 0$  e  $x(t+1) > 0$ ; ou
- $x(t) > 0$  e  $x(t+1) < 0$ ; ou
- $x(t) < 0$  e  $x(t+1) = 0$  e  $x(t+2) > 0$ ; ou
- $x(t) > 0$  e  $x(t+1) = 0$  e  $x(t+2) < 0$ ;

- Desvio Padrão;

Desvio-padrão é uma medida do grau de dispersão dos resultados em torno da média, isto é, um número que mede o quanto os graus estão mais ou menos dispersos em relação à média. Esse desvio padrão ( $\sigma$ ) é definido pela seguinte

fórmula:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (3.11)$$

- Número de valores negativos;

Quantifica a quantidade de números menores que zero presentes em cada padrão de entrada, que pode ser obtido da seguinte forma:

$$N_{neg} = \sum_{i=1}^n -\frac{1}{2} \left( \frac{x_i}{|x_i|} - 1 \right) \quad (3.12)$$

- Valor eficaz;

O valor eficaz, ou valor quadrático médio (RMS - *Root Mean Square*), pode ser descrito como sendo uma medida estatística da magnitude de uma quantidade variável. O valor eficaz pode ser calculado para uma série de valores discretos ou para uma função variável contínua [15]. O valor quadrático médio calculado para um vetor discreto é:

$$x_{rms} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} \quad (3.13)$$

Para uma função variável e contínua ( $f(t)$ ) o valor eficaz é dado por:

$$x_{rms} = \sqrt{\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} [f(t)]^2 dt} \quad (3.14)$$

onde  $t_1$  é o início do intervalo e  $t_2$  o fim.

- Coeficiente de variação;

O coeficiente de variação ( $C_v$ ) é uma medida de dispersão que visa à comparação de distribuições diferentes [14], conforme apresentado pela equação:

$$C_v[\%] = \frac{\sigma}{\mu} 100 \quad (3.15)$$

- Faixas do Histograma

É a representação gráfica da distribuição de frequência de uma massa de medições, normalmente um gráfico de barras verticais. Foram definidas 20 faixas de valores,

dos quais serão levantados o números de registros da variável sob análise contidos nas referidas faixas.

### 3.5.3 Sinais no Domínio da Frequência

A transformada de Fourier [34] tem como princípio básico a decomposição de qualquer fenômeno periódico em uma série infinita de funções senóidais, com amplitudes variáveis e frequências múltiplas de uma fundamental e fases coincidentes.

No processo da transformação o sinal original é decomposto em uma quantidade, que tende a infinito, de sinais senoidais com diferentes frequências, cujo somatório corresponde ao sinal originalmente apresentado. A transformada de Fourier é a seguinte:

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[ a_n \cos\left(\frac{2\pi nt}{T}\right) + b_n \sin\left(\frac{2\pi nt}{T}\right) \right] \quad (3.16)$$

Onde:

$$\begin{aligned} a_0 &= \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) dt \\ a_n &= \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) \cos\left(\frac{2\pi nt}{T}\right) dt \\ b_n &= \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) \sin\left(\frac{2\pi nt}{T}\right) dt \end{aligned} \quad (3.17)$$

Em sinais discretos no tempo com uma duração finita é possível desenvolver uma representação de Fourier alternativa, a Transformada Discreta de Fourier (DFT - *Discrete Fourier Transform*). A DFT consiste em uma seqüência de uma variável contínua, e correspondentes a amostras, igualmente espaçadas em frequência, da transformada de Fourier do sinal.

Alguns algoritmos eficientes para o cálculo de uma DFT de N pontos receberam a denominação de Transformada Rápida de Fourier (FFT - *Fast Fourier Transform*) [21]. Utilizando um desses algoritmos é possível reduzir significativamente o tempo de

processamento, sendo relevante em muitas aplicações onde o tempo de processamento é crítico.

Para a extração destas características foi utilizada a transformada rápida de fourier que possibilitou compactar a informação do sinal de ultra-som, que antes era formado por 32768 pontos, em um vetor de atributos constituído por apenas 21 pontos.

Antes de ser aplicada a FFT foram realizados alguns tratamento estatísticos, mostrados abaixo, no sinal original, visando a redução da dimensão. Inicialmente o sinal de entrada que era composto por 32768 pontos foi reduzido a um quinto do seu valor (6554 pontos) através do Filtro de Decimação. Este filtro consiste basicamente em amostrar o sinal de entrada a uma taxa menor do que aquela original. O valor que determina a redução da frequência de amostragem inicial é conhecido como Taxa de Decimação (*Decimation Ratio*). O valor da Taxa de Decimação utilizado para esta técnica foi 5. O processo aplicado a um sinal que está sendo amostrado, consiste em descartar n-1 amostras a cada n amostragens, onde n é a Taxa de Decimação, da seguinte forma:

$$\rho(i) = z(1 + 5i) \quad (3.18)$$

Onde:

$z \rightsquigarrow$  Sinal original composto por 32768 pontos;

$\rho \rightsquigarrow$  Sinal reduzido composto por 6554 pontos;

$i \rightsquigarrow$  Contador que varia unitariamente de 1 a 6555.

Logo, podemos afirmar que a frequência de amostragem também reduziu a um quinto da frequência anterior. Ou seja,

$$\begin{aligned} f_{amostragem_2} &= f_{amostragem_1} \cdot \frac{1}{5} = 1,96 \cdot 10^6 \cdot \frac{1}{5} \\ f_{amostragem_2} &= 392kHz \end{aligned} \quad (3.19)$$

que, mesmo assim, satisfaz o teorema de Nyquist, Equação 3.5.

Em seguida o vetor  $\rho$  é convertido para o domínio da frequência, utilizando a FFT

$$z = F(\rho) \quad (3.20)$$

Os sinais  $z$  obtidos, no domínio da frequência, têm a dimensão de 6554 pontos, dos quais os últimos 6294 pontos, apresentam valores muito próximos a zero, podendo então expurgá-los, haja vista que os referidos pontos não agregariam nenhuma característica da classe. Logo o sinal truncado  $\beta$  ficará com 260 entradas,

$$\beta = z(1 : 260) \quad (3.21)$$

o que ainda possui uma quantidade de elementos elevada para ser utilizado no treinamento da rede.

Finalmente subdivide-se estes 260 pontos do vetor  $\beta$  em 20 entradas composta pela soma de 13 pontos adjacentes, representando então os valores das 20 faixas de frequência. Logo o sinal  $x$  apresenta uma dimensão 20, como mostrado abaixo:

$$x(n) = \sum_{k=13(n-1)+1}^{13n} \beta(k) \quad (3.22)$$

Onde:

$n \rightsquigarrow$  Contador que varia unitariamente de 1 a 20.

Além destas 20 entradas ainda será agregada mais uma, cujo valor corresponde à soma das 20 entradas já apresentadas

$$x(21) = \sum_{k=1}^{20} x(k) \quad (3.23)$$

Logo, o sinal extraído pela referida técnica de extração de atributos resultou em um vetor composto por 21 elementos, ou seja, 21 entradas da rede neural.

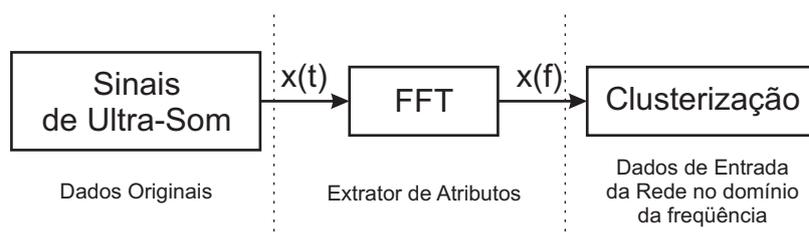


Figura 3.9: Extração de atributos no domínio da frequência

### 3.5.4 Síntese do Capítulo

Neste capítulo foram apresentadas todas as etapas do processo de tratamento dos sinais de entrada das Redes Neurais Artificiais, desde as montagens experimentais até a extração de atributos. Foram estabelecidos os critérios de definição das seis classes de níveis de poluição para aquisição dos sinais que compõem a matriz de dados original. Após a obtenção da matriz de dados original foi constatado que a mesma apresentava elevada dimensão, impossibilitando computacionalmente o treinamento das Redes Neurais Artificiais, logo, faz-se necessário reduzi-la, tornando o problema sob estudo exequível. As técnicas utilizadas para redução da matriz de entrada da RNA foram três métodos de Extração de Atributos (Centróide, Parâmetros Estatísticos, Domínio da Frequência)

# 4

## MODELAGEM DAS REDES PERCEPTRON DE MÚLTIPLAS CAMADAS

O presente capítulo aborda o procedimento utilizado para modelagem da Rede Neural Artificial. Dentre as etapas do procedimento citado podemos destacar a separação dos dados para treinamento e avaliação do desempenho da rede, seleção da melhor representação dos padrões de saída, da melhor arquitetura e finalmente da melhor rede para cada modelo de atributo.

Alguns tipos de RNA têm sido aplicadas com sucesso no desenvolvimento dos classificadores inteligentes, são elas:

- Perceptron Multicamadas - MLP (Multilayer Perceptron);
- Funções Bases Radiais - RBF (Radial Basis Function),
- Mapas Auto-Organizáveis - SOM (Self-Organizing Maps)

Entretanto as redes MLP têm maior capacidade de generalização, pois estas constroem aproximadores globais, enquanto as redes RBF e SOM constroem aproximações e estruturas locais respectivamente. Por isso, os classificadores aqui propostos são desenvolvidos a partir das redes MLP, porém foram também utilizadas Redes RBF, visando uma análise comparativa de resultados, que será mostrada futuramente no Capítulo 6.

Será selecionada a melhor topologia para a classificação de falhas em isoladores, de

acordo com o seu grau de poluição. Por se tratar de um sinal de entrada relativamente pequeno foi decidido utilizar a técnica de treinamento de Levenberg-Marquardt, pois apresenta uma boa precisão, por utilizar derivadas do segundo grau, porém um elevado esforço computacional.

O treinamento da rede deve ser encerrado antes que a rede seja excessivamente ajustada aos dados do treinamento, prejudicando a generalização da rede. O método de treinamento com parada antecipada, *early stop*, consiste em evitar que os resultados obtidos, erros médios quadráticos, pelo conjunto de treinamento divirja dos resultados do conjunto de validação, como ilustrado na Figura 4.1. O erro de validação é medido a cada exemplo do seu subconjunto.

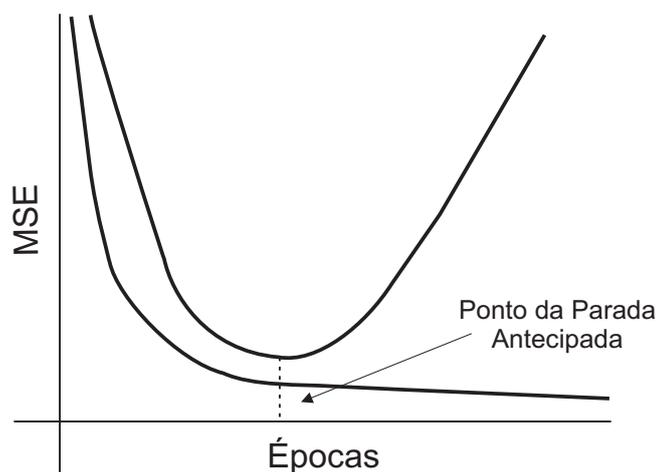


Figura 4.1: Treinamento com parada antecipada

## 4.1 Separação dos Dados de Entrada

Após a etapa da extração de atributos, Seção 3.5, faz-se necessária a divisão do sinal de entrada no conjunto de treinamento, composto por 60% do conjunto total, implicando em uma matriz de dimensão  $n \times 360$ , no conjunto de validação, composto por 20% do conjunto total, implicando em uma matriz de dimensão  $n \times 120$ , e no conjunto de teste, com mesma dimensão do conjunto de validação [22]. Onde  $n$  é a dimensão de entrada da rede, que depende diretamente da extração de atributos

utilizada, como mostrado na Tabela 4.1. Desses conjuntos definidos acima o conjunto de treinamento tem como objetivo ajustar os parâmetros livres da rede neural, o conjunto de validação para o monitoramento dos treinamentos, garantindo a capacidade de generalização da rede e o conjunto de teste utilizado para aferir a performance dos modelos criados, ou seja, erro de classificação.

Tabela 4.1: Dimensão de entrada da RNA

<b>Extração de Atributos</b>	<i>n</i>
Centróide	9
Parâmetros Estatísticos	28
Domínio da frequência	21

## 4.2 Codificação das Classes

Com o intuito de selecionar uma rede com a melhor representação de saída, foram realizados testes para a saída com 1 neurônio e com 6 neurônios. A codificação para a classificação é como segue abaixo:

Tabela 4.2: Codificação da camada de saída da RNA

<b>Classes</b>	<b>Distúrbios</b>	<b>Codificação</b>	
		<b>6 neurônios</b>	<b>1 neurônio</b>
A	Atmosfera sem poluição	0,9 0 0 0 0 0	0,10
B	Atmosfera limpa	0 0,9 0 0 0 0	0,26
C	Contaminação muito leve	0 0 0,9 0 0 0	0,42
D	Contaminação leve	0 0 0 0,9 0 0	0,58
E	Contaminação pesada	0 0 0 0 0,9 0	0,74
F	Contaminação muito pesada	0 0 0 0 0 0,9	0,90

Onde a classificação das redes com um neurônio da saída será definida pelo valor das classes, definidos na Tabela 4.2, que mais se aproximarem em módulo da saída da rede. Já as redes com seis neurônios na camada de saída utilizará a técnica do *winner takes all* (vencedor leva tudo), onde a saída da rede que obtiver o maior valor será a

classe a qual o exemplo analisado pertence, como apresentado na Tabela 4.2.

Foram utilizados 4 modelos de redes distintos. São eles:

- *Modelo 1* - Caracterizado por ter apenas 1 neurônio na camada de saída da RNA;
- *Modelo 2* - Caracterizado por ter 1 neurônio para cada classe, ou seja, 6 neurônios da camada de saída da RNA;
- *Modelo 3* - Formado por três especialistas (redes) a partir do Modelo 1;
- *Modelo 4* - Formado por três especialistas (redes) a partir do Modelo 2.

### 4.3 Seleção das Melhores Arquiteturas

Serão analisadas as duas representações de saída distintas, mostradas na Tabela 4.2. Em cada rede, com as diferentes representações de saída, será variada a quantidade de neurônios da camada oculta. No presente trabalho foi variado de 5 a 15 neurônios, com o intuito de selecionar as três arquiteturas que melhor representam o problema em questão, ou seja, serão avaliadas diversas topologias de rede a partir dos números de neurônios ocultos [19]. Portanto, foram realizados diversos treinamentos com as 11 arquiteturas distintas e 10 inicializações de pesos para cada arquitetura, sendo selecionadas como as três melhores redes aquelas que obtiveram os três menores MSE de validação médio das 10 inicializações como mostrado a seguir [25].

#### 4.3.1 Melhores Arquiteturas - Centróide - Modelo 1

Com base na Tabela 4.3 observa-se que as arquiteturas com 6, 5 e 9 neurônios na camada oculta obtiveram respectivamente o melhor (Rede 1), o segundo melhor (Rede 2) e o terceiro melhor (Rede 3) desempenho durante os treinamentos.

Tabela 4.3: Seleção da melhor arquitetura - Centróide - Modelo 1

Neurônios na camada oculta	MSE médio de validação
5	<b>1,704</b>
<b>6</b>	<b>1,457</b>
7	1,818
8	1,739
<b>9</b>	<b>1,736</b>
10	1,829
11	1,740
12	1,781
13	1,999
14	2,073
15	2,199

Assim as arquiteturas das referidas redes, tendo como sinal de entrada os atributos extraídos pela técnica do centróide, são apresentadas como mostra a Tabela 4.4.

Tabela 4.4: Arquitetura das melhores redes - Centróide - Modelo 1

Rede	Camada de entrada	Camada oculta	Camada de saída
Rede 1	9	6	1
Rede 2	9	5	1
Rede 3	9	9	1

#### 4.3.2 Melhores Arquiteturas - Centróide - Modelo 2

Com base na Tabela 4.5 observa-se que as arquiteturas com 11, 12 e 13 neurônios na camada oculta obtiveram respectivamente o melhor (Rede 1), o segundo melhor (Rede 2) e o terceiro melhor (Rede 3) desempenho durante os treinamentos.

Tabela 4.5: Seleção da melhor arquitetura - Centróide - Modelo 2

Neurônios na camada oculta	MSE médio de validação
5	2,379
6	3,887
7	3,765
8	3,029
9	3,701
10	3,440
<b>11</b>	<b>1,617</b>
<b>12</b>	<b>1,792</b>
<b>13</b>	<b>1,955</b>
14	2,606
15	3,736

Assim as arquiteturas das referidas redes, tendo como sinal de entrada os atributos extraídos pela técnica do centróide, são apresentadas como mostra a Tabela 4.6.

Tabela 4.6: Arquitetura das melhores redes - Centróide - Modelo 2

Rede	Camada de entrada	Camada oculta	Camada de saída
Rede 1	9	11	6
Rede 2	9	12	6
Rede 3	9	13	6

#### 4.3.3 Melhores Arquiteturas - Parâmetros Estatísticos - Modelo 1

Com base na Tabela 4.7 observa-se que as arquiteturas com 12, 10 e 13 neurônios na camada oculta obtiveram respectivamente o melhor (Rede 1), o segundo melhor (Rede 2) e o terceiro melhor (Rede 3) desempenho durante os treinamentos.

Tabela 4.7: Seleção da melhor arquitetura - Parâmetros Estatísticos - Modelo 1

Neurônios na camada oculta	MSE médio de validação
5	7,885
6	7,823
7	7,705
8	7,652
9	7,834
<b>10</b>	<b>7,634</b>
11	7,902
<b>12</b>	<b>7,606</b>
<b>13</b>	<b>7,641</b>
14	7,889
15	7,808

Assim as arquiteturas das referidas redes, tendo como sinal de entrada os atributos extraídos pela técnica de parâmetros estatísticos, são apresentadas como mostra a Tabela 4.8.

Tabela 4.8: Arquitetura das melhores redes - Parâmetros Estatísticos - Modelo 1

Rede	Camada de entrada	Camada oculta	Camada de saída
Rede 1	28	12	1
Rede 2	28	10	1
Rede 3	28	13	1

#### 4.3.4 Melhores Arquiteturas - Parâmetros Estatísticos - Modelo 2

Com base na Tabela 4.9 observa-se que as arquiteturas com 9, 6 e 13 neurônios na camada oculta obtiveram respectivamente o melhor (Rede 1), o segundo melhor (Rede 2) e o terceiro melhor (Rede 3) desempenho durante os treinamentos.

Tabela 4.9: Seleção da melhor arquitetura - Parâmetros Estatísticos - Modelo 2

Neurônios na camada oculta	MSE médio de validação
5	15,242
<b>6</b>	<b>10,014</b>
7	13,260
8	11,128
<b>9</b>	<b>9,334</b>
10	11,549
11	14,285
12	10,473
<b>13</b>	<b>10,362</b>
14	12,191
15	11,088

Assim as arquiteturas das referidas redes, tendo como sinal de entrada os atributos extraídos pela técnica de parâmetros estatísticos, são apresentadas como mostra a Tabela 4.10.

Tabela 4.10: Arquitetura das melhores redes - Parâmetros Estatísticos - Modelo 2

Rede	Camada de entrada	Camada oculta	Camada de saída
Rede 1	28	9	6
Rede 2	28	6	6
Rede 3	28	13	6

#### 4.3.5 Melhores Arquiteturas - Domínio da Frequência - Modelo 1

Com base na Tabela 4.11 observa-se que as arquiteturas com 12, 6 e 5 neurônios na camada oculta obtiveram respectivamente o melhor (Rede 1), o segundo melhor (Rede 2) e o terceiro melhor (Rede 3) desempenho durante os treinamentos.

Tabela 4.11: Seleção da melhor arquitetura - Domínio da Frequência - Modelo 1

Neurônios na camada oculta	MSE médio de validação
<b>5</b>	<b>7,482</b>
<b>6</b>	<b>7,475</b>
7	7,525
8	7,517
9	7,525
10	7,509
11	7,490
<b>12</b>	<b>7,468</b>
13	7,508
14	7,548
15	7,531

Assim as arquiteturas das referidas redes, tendo como sinal de entrada os atributos extraídos pela técnica do domínio da frequência, são apresentadas como mostra a Tabela 4.12.

Tabela 4.12: Arquitetura das melhores redes - Domínio da Frequência - Modelo 1

Rede	Camada de entrada	Camada oculta	Camada de saída
Rede 1	21	12	1
Rede 2	21	6	1
Rede 3	21	5	1

#### 4.3.6 Melhores Arquiteturas - Domínio da Frequência - Modelo 2

Com base na Tabela 4.13 observa-se que as arquiteturas com 14, 6 e 15 neurônios na camada oculta obtiveram respectivamente o melhor (Rede 1), o segundo melhor (Rede 2) e o terceiro melhor (Rede 3) desempenho durante os treinamentos.

Tabela 4.13: Seleção da melhor arquitetura - Domínio da Frequência - Modelo 2

Neurônios na camada oculta	MSE médio de validação
5	14,370
<b>6</b>	<b>10,912</b>
7	12,602
8	11,984
9	14,051
10	11,994
11	14,161
12	18,477
13	14,723
<b>14</b>	<b>10,573</b>
<b>15</b>	<b>11,856</b>

Assim as arquiteturas das referidas redes, tendo como sinal de entrada os atributos extraídos pela técnica do domínio da frequência, são apresentadas como mostra a Tabela 4.14.

Tabela 4.14: Arquitetura das melhores redes - Domínio da Frequência - Modelo 2

Rede	Camada de entrada	Camada oculta	Camada de saída
Rede 1	21	14	6
Rede 2	21	6	6
Rede 3	21	15	6

## 4.4 Seleção dos Melhores Experimentos

### Validação Cruzada

O principal objetivo da validação cruzada, ou Cross Validation, é assegurar a generalização da rede para classificação do problema abordado, ou seja, fornecer uma classificação coerente para uma situação não utilizada no processo de treinamento.

É necessário, para a criação da rede neural, que realizemos a divisão do conjunto

de entrada em conjuntos de treinamento, validação e teste.

A validação cruzada consiste na divisão do conjunto de entrada em  $k$  configurações distintas de treinamento, validação e teste, que devem ser geradas aleatoriamente. Essa técnica é denominada  $k$ -fold Cross Validation. Após a montagem das  $k$  configurações será selecionada a que obteve um melhor desempenho de validação. Neste caso foi utilizada a técnica com  $k$  igual a dez, ou seja, 10-fold.

Foi utilizada a técnica de validação cruzada ( $k$ -fold) com o objetivo de assegurar uma boa generalização da rede. A tabela com os 10 experimentos dos diferentes subconjuntos de treinamento, validação e teste utilizados no presente trabalho foi a Tabela 4.15.

Tabela 4.15: Validação cruzada

Experimento	Conjunto de Treinamento	Conjunto de Validação	Conjunto de Teste
1	[5, 6, 7, 8, 9, 10]	[1, 2]	[3, 4]
2	[1, 6, 7, 8, 9, 10]	[2, 3]	[4, 5]
3	[1, 2, 7, 8, 9, 10]	[3, 4]	[5, 6]
4	[1, 2, 3, 8, 9, 10]	[4, 5]	[6, 7]
5	[1, 2, 3, 4, 9, 10]	[5, 6]	[7, 8]
6	[1, 2, 3, 4, 5, 10]	[6, 7]	[8, 9]
7	[1, 2, 3, 4, 5, 6]	[7, 8]	[9, 10]
8	[2, 3, 4, 5, 6, 7]	[8, 9]	[10, 1]
9	[3, 4, 5, 6, 7, 8]	[9, 10]	[1, 2]
10	[4, 5, 6, 7, 8, 9]	[10, 1]	[2, 3]

Para cada experimento serão simuladas as três melhores arquiteturas encontradas anteriormente, Seção 4.3. E para cada uma dessas redes foi utilizada uma metodologia de 10 treinamentos com 10 inicializações distintas dos pesos para cada topologia sugerida, visando a garantia de que o desempenho de uma determinada arquitetura não é resultado de uma inicialização ótima de pesos.

Após os treinamentos e as simulações das 10 inicializações de cada treinamento, foi realizada a seleção do melhor experimento. Onde o melhor experimento é aquele

que obteve um menor erro de classificação médio das 10 inicializações dentre os 10 experimentos realizados.

#### 4.4.1 Melhores Experimentos - Centróide - Modelo 1

Os resultados encontrados para a seleção dos melhores experimentos, com a metodologia de seleção mostrada acima, é apresentado na Tabela 4.16, para a técnica do centróide com 1 neurônio na camada de saída.

Tabela 4.16: EC - Centróide - Modelo 1

Experimento	Média dos EC das 10 Inicializações		
	Rede 1	Rede 2	Rede 3
1	33,833	33,833	36,167
2	40,667	31,500	33,750
3	<b>27,417</b>	34,083	32,167
4	30,667	37,583	<b>29,667</b>
5	34,250	36,167	32,667
6	35,834	37,667	36,583
7	38,000	37,500	39,750
8	36,583	39,083	42,583
9	28,333	<b>29,333</b>	30,833
10	32,583	33,167	30,500

Na Tabela 4.16, podem ser observados os melhores experimentos para cada rede. Onde para a Rede 1 foi o experimento 3, para a Rede 2 o experimento 9 e para a Rede 3 o experimento 4.

#### 4.4.2 Melhores Experimentos - Centróide - Modelo 2

Os resultados encontrados para a seleção dos melhores experimentos é apresentado na Tabela 4.17, para a técnica do centróide com 6 neurônios na camada de saída.

Tabela 4.17: EC - Centróide - Modelo 2

Experimento	Média dos EC das 10 Inicializações		
	Rede 1	Rede 2	Rede 3
1	14,500	10,250	<b>6,833</b>
2	20,750	27,417	20,167
3	17,667	26,917	21,083
4	21,917	22,167	15,667
5	23,000	<b>7,583</b>	16,167
6	20,083	12,500	13,500
7	<b>8,417</b>	12,167	8,667
8	10,583	16,833	16,750
9	11,333	15,083	13,833
10	27,500	17,667	30,250

Na Tabela 4.17, podem ser observados os melhores experimentos para cada rede. Onde para a Rede 1 foi o experimentos 7, para a Rede 2 o experimento 5 e para a Rede 3 o experimento 1.

#### 4.4.3 Melhores Experimentos - Parâmetros Estatísticos - Modelo 1

Os resultados encontrados para a seleção dos melhores experimentos é apresentado na Tabela 4.18, para a técnica de parâmetros estatísticos com 1 neurônio na camada de saída.

Tabela 4.18: EC - Parâmetros Estatísticos - Modelo 1

Experimento	Média dos EC das 10 Inicializações		
	Rede 1	Rede 2	Rede 3
1	85,333	84,333	84,584
2	82,917	83,417	82,833
3	83,083	82,917	81,750
4	<b>80,750</b>	<b>81,083</b>	<b>81,583</b>
5	81,500	81,167	82,250
6	84,000	83,167	83,583
7	83,250	84,250	83,333
8	84,000	83,750	85,500
9	83,083	82,750	83,833
10	82,833	81,417	81,750

Na Tabela 4.18, podem ser observados os melhores experimentos para cada rede. Onde para todas as redes (Rede 1, Rede 2 e Rede 3) o melhor foi o experimento 4.

#### 4.4.4 Melhores Experimentos - Parâmetros Estatísticos - Modelo 2

Os resultados encontrados para a seleção dos melhores experimentos é apresentado na Tabela 4.19, para a técnica de parâmetros estatísticos com 6 neurônios na camada de saída.

Tabela 4.19: EC - Parâmetros Estatísticos - Modelo 2

Experimento	Média dos EC das 10 Inicializações		
	Rede 1	Rede 2	Rede 3
1	64,333	68,083	65,167
2	55,250	61,000	57,917
3	59,333	55,750	60,583
4	<b>52,667</b>	60,417	66,000
5	53,083	<b>49,667</b>	<b>54,167</b>
6	57,583	52,917	61,750
7	62,583	50,417	57,083
8	55,833	63,583	58,500
9	57,583	69,250	66,333
10	59,750	59,334	62,250

Na Tabela 4.19, podem ser observados os melhores experimentos para cada rede. Onde para a Rede 1 foi o experimento 4, para a Rede 2 o experimento 5 e para a Rede 3 o experimento 5.

#### 4.4.5 Melhores Experimentos - Domínio da Frequência - Modelo 1

Os resultados encontrados para a seleção dos melhores experimentos é apresentado na Tabela 4.20, para a técnica do domínio da frequência com 1 neurônio na camada de saída.

Tabela 4.20: EC - Domínio da Frequência - Modelo 1

Experimento	Média dos EC das 10 Inicializações		
	Rede 1	Rede 2	Rede 3
1	84,500	84,583	85,000
2	82,583	<b>82,583</b>	82,500
3	83,750	83,500	83,917
4	85,917	84,833	85,667
5	84,500	84,667	84,417
6	83,750	83,750	83,334
7	<b>81,833</b>	82,667	<b>82,083</b>
8	83,750	83,500	83,333
9	86,083	84,750	85,833
10	84,583	83,833	84,917

Na Tabela 4.20, podem ser observados os melhores experimentos para cada rede. Onde para a Rede 1 foi o experimento 7, para a Rede 2 o experimento 2 e para a Rede 3 o experimento 7.

#### 4.4.6 Melhores Experimentos - Domínio da Frequência - Modelo 2

Os resultados encontrados para a seleção dos melhores experimentos é apresentado na Tabela 4.21, para a técnica do domínio da frequência com 6 neurônios na camada de saída.

Tabela 4.21: EC - Domínio da Freqüência - Modelo 2

Experimento	Média dos EC das 10 Inicializações		
	Rede 1	Rede 2	Rede 3
1	73,750	65,667	67,583
2	75,333	69,583	77,250
3	69,167	67,000	63,750
4	70,750	66,417	69,250
5	<b>64,000</b>	70,667	<b>61,250</b>
6	64,750	<b>64,333</b>	69,500
7	72,667	68,583	69,000
8	64,667	74,417	68,583
9	67,583	65,917	70,750
10	82,583	74,417	73,917

Na Tabela 4.21, podem ser observados os melhores experimentos para cada rede. Onde para a Rede 1 foi o experimento 5, para a Rede 2 o experimento 6 e para a Rede 3 o experimento 5.

## 4.5 Síntese do Capítulo

O principal objetivo deste Capítulo foi identificar a melhor representação dos padrões de saída, a melhor arquitetura e a melhor rede. Será apresentado neste tópico uma síntese dos resultados obtidos.

Serão mostradas abaixo as arquiteturas das melhores redes, para técnicas do Centróide, dos Parâmetros Estatísticos e do Domínio da Freqüência, assim como os melhores experimentos para cada rede de cada modelo de todas as técnicas de extração de atributos.

Tabela 4.22: Síntese dos Resultados da Modelagem das Redes

			Quantidade de Neurônios na Camada			Melhor Experimento
			Entrada	Oculto	Saída	
Centróide	Modelo 1	Rede 1	9	6	1	3
		Rede 2	9	5	1	9
		Rede 3	9	9	1	4
	Modelo 2	Rede 1	9	11	6	7
		Rede 2	9	12	6	5
		Rede 3	9	13	6	1
Parâmetros Estatísticos	Modelo 1	Rede 1	28	12	1	4
		Rede 2	28	10	1	4
		Rede 3	28	13	1	4
	Modelo 2	Rede 1	28	9	6	4
		Rede 2	28	6	6	5
		Rede 3	28	13	6	5
Domínio da Frequência	Modelo 1	Rede 1	21	12	1	7
		Rede 2	21	6	1	2
		Rede 3	21	5	1	7
	Modelo 2	Rede 1	21	14	6	5
		Rede 2	21	6	6	6
		Rede 3	21	15	6	5

Todas as arquiteturas apresentadas na Tabela 4.22 serão simuladas no Capítulo 5, visando encontrar a rede que melhor se adéqüe ao problema sob estudo.

# 5

## SIMULAÇÃO DAS REDES PERCEPTRON DE MÚLTIPLAS CAMADAS

Neste capítulo são apresentados os resultados obtidos, erros de classificação, das simulações do conjunto de teste das melhores redes modeladas no Capítulo 4, visando a seleção da rede com melhor performance para o problema proposto. Serão abordadas também as técnicas utilizadas para a obtenção da classificação realizada por um especialista, Modelo 1 e Modelo 2, e realizada por três especialistas, Modelo 3 e Modelo 4.

### 5.1 Classificação

Os resultados de saída dos modelos de rede serão as classificações dos isoladores sob análise. A Tabela 4.2 será utilizada como decodificação da classificação da rede com um neurônio na camada de saída, Modelo 1, e da rede com seis neurônios na camada de saída, Modelo 2. Onde para a classificação será levado em consideração o código da classe que mais se aproximar do valor de saída da rede.

#### 5.1.1 Classificador com Um Especialista

A classificação do grau de poluição dos isoladores para redes com um especialista será a próxima saída da rede, apenas codificada através da Tabela 4.2. Pode-se ilustrar esse classificador na Figura 5.1.

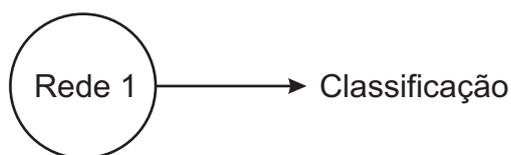


Figura 5.1: Classificação com um especialista

### 5.1.2 Classificador com Três Especialistas

A classificação do isolador sob análise é oriunda da média das saídas das três melhores redes, selecionadas na Seção 4.3, como mostrado na Figura 5.2. Porém as combinações das redes devem ser realizadas utilizando redes com o mesmo experimento (*k-fold*), pois o conjunto de teste deverá ser igual, possibilitando a comparação entre a saída da rede e a saída desejada.

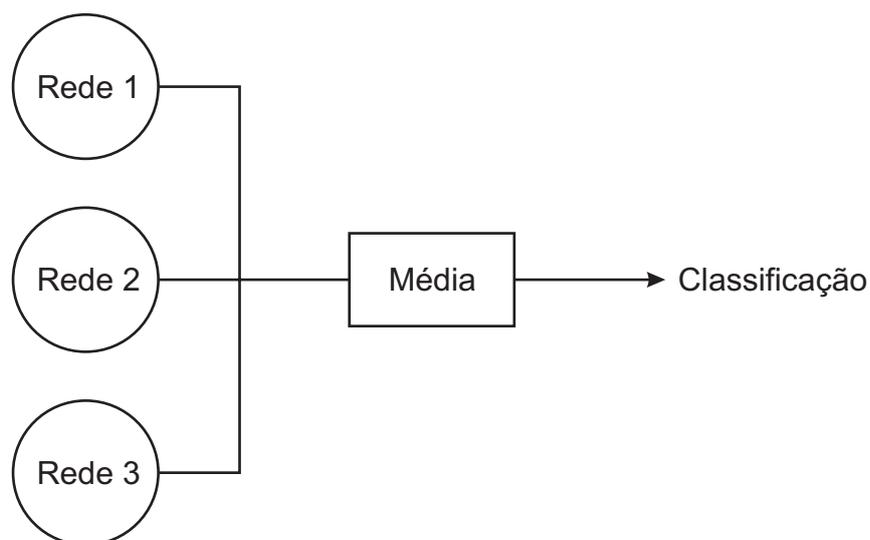


Figura 5.2: Classificação com três especialistas

## 5.2 Seleção das Melhores Inicializações

Nesta seção serão simuladas as três redes para os melhores experimentos apresentados na Seção 4.4.

### 5.2.1 Melhores Inicializações - Centróide - Modelo 1

Na modelagem da rede que tem como sinal de entrada atributos extraídos, do sinal de ultra-som, pela técnica do centróide os melhores experimentos foram:

- Rede 1 → Experimento 3;
- Rede 2 → Experimento 9;
- Rede 3 → Experimento 4;

Logo, as três redes foram simuladas utilizando como separação dos conjuntos de entrada (conjunto de validação, treinamento e teste) o experimento 3, o experimento 4 e o experimento 9.

Os resultados dos erros de classificação das referidas simulações que utilizam conjunto de entradas definidos pelo experimento 3 são apresentados na Tabela 5.1.

Tabela 5.1: EC - Centróide - Modelo 1 - Experimento 3

Inicialização	Erro de Classificação		
	Rede 1	Rede 2	Rede 3
1	34,167	31,667	<b>20,000</b>
2	29,167	38,333	23,333
3	27,500	27,500	30,000
4	22,500	40,833	50,833
5	38,333	35,000	36,667
6	<b>19,167</b>	45,000	29,167
7	22,500	29,167	38,333
8	25,000	25,000	39,167
9	27,500	45,000	23,333
10	28,333	<b>23,333</b>	30,833

Os resultados dos erros de classificação das referidas simulações que utilizam conjunto de entradas definidos pelo experimento 4 são apresentados na Tabela 5.2.

Tabela 5.2: EC - Centróide - Modelo 1 - Experimento 4

Inicialização	Erro de Classificação		
	Rede 1	Rede 2	Rede 3
1	35,000	28,333	38,333
2	28,333	<b>25,000</b>	28,333
3	24,167	35,000	29,167
4	50,000	51,667	<b>25,000</b>
5	35,000	<b>25,000</b>	33,333
6	<b>19,167</b>	55,833	30,833
7	25,000	42,500	25,833
8	23,333	34,167	32,500
9	40,833	30,833	26,667
10	25,833	47,500	26,667

Os resultados dos erros de classificação das referidas simulações que utilizam conjunto de entradas definidos pelo experimento 9 são apresentados na Tabela 5.3.

Tabela 5.3: EC - Centróide - Modelo 1 - Experimento 9

Inicialização	Erro de Classificação		
	Rede 1	Rede 2	Rede 3
1	25,000	26,667	53,333
2	40,000	25,000	34,167
3	25,833	25,000	26,667
4	30,000	30,000	24,167
5	35,833	42,500	30,833
6	29,167	19,167	31,667
7	23,333	32,500	20,833
8	<b>22,500</b>	36,667	<b>20,000</b>
9	<b>22,500</b>	40,000	39,167
10	29,167	<b>15,833</b>	27,500

### 5.2.2 Melhores Inicializações - Centróide - Modelo 2

Na modelagem da rede que tem como sinal de entrada atributos extraídos, do sinal de ultra-som, pela técnica do centróide os melhores experimentos foram:

- Rede 1 → Experimento 7;
- Rede 2 → Experimento 5;
- Rede 3 → Experimento 1;

Logo, as três redes foram simuladas utilizando como separação dos conjuntos de entrada (conjunto de validação, treinamento e teste) o experimento 1, o experimento 5 e o experimento 7.

Os resultados dos erros de classificação das referidas simulações que utilizam conjunto de entradas definidos pelo experimento 1 são apresentados na Tabela 5.4.

Tabela 5.4: EC - Centróide - Modelo 2 - Experimento 1

Inicialização	Erro de Classificação		
	Rede 1	Rede 2	Rede 3
1	22,500	42,500	9,167
2	7,500	3,333	5,833
3	<b>3,333</b>	4,167	4,167
4	20,000	2,500	5,833
5	50,833	5,000	<b>3,333</b>
6	<b>3,333</b>	3,333	18,333
7	5,833	33,333	5,833
8	6,667	3,333	6,667
9	4,167	3,333	5,833
10	20,833	<b>1,667</b>	<b>3,333</b>

Os resultados dos erros de classificação das referidas simulações que utilizam conjunto de entradas definidos pelo experimento 5 são apresentados na Tabela 5.5.

Tabela 5.5: EC - Centróide - Modelo 2 - Experimento 5

Inicialização	Erro de Classificação		
	Rede 1	Rede 2	Rede 3
1	34,167	3,333	3,333
2	55,833	5,833	<b>2,500</b>
3	20,833	4,167	5,000
4	7,500	5,000	20,833
5	4,167	5,833	22,500
6	86,667	19,167	3,333
7	7,500	5,000	20,000
8	6,667	<b>2,500</b>	5,833
9	<b>2,500</b>	21,667	74,167
10	4,167	3,333	4,167

Os resultados dos erros de classificação das referidas simulações que utilizam conjunto de entradas definidos pelo experimento 7 são apresentados na Tabela 5.6.

Tabela 5.6: EC - Centróide - Modelo 2 - Experimento 7

Inicialização	Erro de Classificação		
	Rede 1	Rede 2	Rede 3
1	4,167	40,000	4,167
2	7,500	16,667	18,333
3	4,167	<b>4,167</b>	<b>2,500</b>
4	5,000	7,500	5,000
5	4,167	9,167	8,333
6	26,667	5,833	6,667
7	6,667	5,000	5,833
8	19,167	5,000	23,333
9	<b>2,500</b>	5,833	<b>2,500</b>
10	4,167	22,500	10,000

### 5.2.3 Melhores Inicializações - Parâmetros Estatísticos - Modelo 1

Na modelagem da rede que tem como sinal de entrada atributos extraídos, do sinal de ultra-som, pela técnica de parâmetros estatísticos os melhores experimentos foram:

- Rede 1 → Experimento 4;
- Rede 2 → Experimento 4;
- Rede 3 → Experimento 4;

Logo, as três redes foram simuladas utilizando como separação dos conjuntos de entrada o experimento 4.

Os resultados dos erros de classificação das referidas simulações que utilizam conjunto de entradas definidos pelo experimento 4 são apresentados na Tabela 5.7.

Tabela 5.7: EC - Parâmetros Estatísticos - Modelo 1 - Experimento 4

Inicialização	Erro de Classificação		
	Rede 1	Rede 2	Rede 3
1	80,833	85,833	80,833
2	81,667	80,000	80,833
3	80,833	80,000	80,833
4	83,333	<b>79,167</b>	83,333
5	79,167	80,833	80,000
6	80,000	80,833	85,000
7	78,333	80,833	82,500
8	84,167	82,500	81,667
9	<b>77,500</b>	<b>79,167</b>	82,500
10	81,667	81,667	<b>78,333</b>

### 5.2.4 Melhores Inicializações - Parâmetros Estatísticos - Modelo 2

Na modelagem da rede que tem como sinal de entrada atributos extraídos, do sinal de ultra-som, pela técnica de parâmetros estatísticos os melhores experimentos foram:

- Rede 1 → Experimento 4;

- Rede 2 → Experimento 5;
- Rede 3 → Experimento 5;

Logo, as três redes foram simuladas utilizando como separação dos conjuntos de entrada o experimento 4 e o experimento 5.

Os resultados dos erros de classificação das referidas simulações que utilizam conjunto de entradas definidos pelo experimento 4 são apresentados na Tabela 5.8.

Tabela 5.8: EC - Parâmetros Estatísticos - Modelo 2 - Experimento 4

Experimento	Inicialização		
	Rede 1	Rede 2	Rede 3
1	46,667	51,667	67,500
2	<b>43,333</b>	69,167	61,667
3	47,500	65,833	<b>45,833</b>
4	46,667	62,500	57,500
5	48,333	48,333	83,333
6	53,333	80,000	52,500
7	46,667	53,333	83,333
8	83,333	<b>44,167</b>	54,167
9	44,167	70,000	78,333
10	66,667	59,167	75,833

Os resultados dos erros de classificação das referidas simulações que utilizam conjunto de entradas definidos pelo experimento 5 são apresentados na Tabela 5.9.

Tabela 5.9: EC - Parâmetros Estatísticos - Modelo 2 - Experimento 5

Inicialização	Erro de Classificação		
	Rede 1	Rede 2	Rede 3
1	52,500	43,333	48,333
2	66,667	57,500	40,833
3	83,333	52,500	<b>40,000</b>
4	40,000	66,667	43,333
5	37,500	43,333	<b>40,000</b>
6	56,667	45,833	46,667
7	65,833	42,500	65,000
8	49,167	59,167	68,333
9	42,500	44,167	70,833
10	<b>36,667</b>	<b>41,667</b>	78,333

### 5.2.5 Melhores Inicializações - Domínio da Frequência - Modelo 1

Na modelagem da rede que tem como sinal de entrada atributos extraídos, do sinal de ultra-som, pela técnica do Domínio da Frequência os melhores experimentos foram:

- Rede 1 → Experimento 7;
- Rede 2 → Experimento 2;
- Rede 3 → Experimento 7;

Logo, as três redes foram simuladas utilizando como separação dos conjuntos de entrada o experimento 2 e o experimento 7.

Os resultados dos erros de classificação das referidas simulações que utilizam conjunto de entradas definidos pelo experimento 2 são apresentados na Tabela 5.10.

Tabela 5.10: EC - Domínio da Frequência - Modelo 1 - Experimento 2

Inicialização	Erro de Classificação		
	Rede 1	Rede 2	Rede 3
1	<b>79,167</b>	84,167	83,333
2	80,833	<b>80,000</b>	83,333
3	84,167	82,500	<b>80,000</b>
4	80,000	<b>80,000</b>	83,333
5	84,167	83,333	82,500
6	85,000	85,000	83,333
7	83,333	84,167	81,667
8	83,333	82,500	80,833
9	83,333	81,667	82,500
10	82,500	82,500	84,167

Os resultados dos erros de classificação das referidas simulações que utilizam conjunto de entradas definidos pelo experimento 7 são apresentados na Tabela 5.11.

Tabela 5.11: EC - Domínio da Frequência - Modelo 1 - Experimento 7

Inicialização	Erro de Classificação		
	Rede 1	Rede 2	Rede 3
1	80,833	81,667	85,000
2	80,833	<b>80,833</b>	80,000
3	83,333	84,167	84,167
4	85,000	83,333	85,000
5	85,833	82,500	80,000
6	<b>80,000</b>	82,500	84,167
7	<b>80,000</b>	82,500	83,333
8	82,500	82,500	80,000
9	<b>80,000</b>	83,333	80,833
10	<b>80,000</b>	83,333	<b>78,333</b>

### 5.2.6 Melhores Inicializações - Domínio da Frequência - Modelo 2

Na modelagem da rede que tem como sinal de entrada atributos extraídos, do sinal de ultra-som, pela técnica do Domínio da Frequência os melhores experimentos foram:

- Rede 1 → Experimento 5;
- Rede 2 → Experimento 6;
- Rede 3 → Experimento 5;

Logo, as três redes foram simuladas utilizando como separação dos conjuntos de entrada o experimento 5 e o experimento 6.

Os resultados dos erros de classificação das referidas simulações que utilizam conjunto de entradas definidos pelo experimento 5 são apresentados na Tabela 5.12.

Tabela 5.12: EC - Domínio da Frequência - Modelo 2 - Experimento 5

Inicialização	Erro de Classificação		
	Rede 1	Rede 2	Rede 3
1	71,667	<b>51,667</b>	54,167
2	54,167	83,333	54,167
3	77,500	69,167	60,833
4	60,000	83,333	57,500
5	52,500	87,500	84,167
6	83,333	53,333	<b>50,000</b>
7	72,500	64,167	65,000
8	<b>51,667</b>	62,500	57,500
9	55,000	68,333	65,000
10	61,667	83,333	64,167

Os resultados dos erros de classificação das referidas simulações que utilizam conjunto de entradas definidos pelo experimento 6 são apresentados na Tabela 5.13.

Tabela 5.13: EC - Domínio da Freqüência - Modelo 2 - Experimento 6

Inicialização	Erro de Classificação		
	Rede 1	Rede 2	Rede 3
1	60,000	56,667	83,333
2	<b>55,833</b>	55,833	<b>52,500</b>
3	<b>55,833</b>	70,833	62,500
4	75,000	58,333	83,333
5	60,000	57,500	83,333
6	57,500	55,833	67,500
7	57,500	<b>54,167</b>	60,833
8	83,333	84,167	56,667
9	83,333	67,500	61,667
10	59,167	82,500	83,333

### 5.3 Resultados das Simulações com Um Especialista

Uma vez definida a arquitetura, foram realizados os treinamentos com a técnica de validação cruzada, k-fold. Para o Modelo 1 e o Modelo 2, com apenas um neurônio na camada de saída, foram selecionados os menores valores de erro de classificação da Rede 1 para os melhores experimentos.

#### 5.3.1 Resultados das Simulações - Centróide - Modelo 1

Para a seleção da melhor rede foram analisados os erros de classificação obtidos nas simulações dos melhores experimentos. Foi selecionado o menor valor de erro de classificação encontrado para a Rede 1 da Tabela 5.1, Tabela 5.2 e da Tabela 5.3. O menor erro de classificação encontrado para a referida rede foi de 19,167 %, isto é, dentre os 120 exemplos que compõem o conjunto de teste, a referida rede errou 23 exemplos. Logo as características dessa rede selecionada são as seguintes:

Um Especialista	{	Neurônios na camada de entrada: 9
		Neurônios na camada oculta: 6
		Neurônios na camada de saída: 1
		Experimento: 3
		Inicialização de pesos: 6

### 5.3.2 Resultados das Simulações - Parâmetros Estatísticos - Modelo 1

Para a seleção da melhor rede foram analisados os erros de classificação obtidos nas simulações dos melhores experimentos. Foi selecionado o menor valor de erro de classificação encontrado para a Rede 1 da Tabela 5.7. O menor erro de classificação encontrado para a referida rede foi de 77,500 %, isto é, dentre os 120 exemplos que compõem o conjunto de teste, a referida rede errou 93 exemplos. Logo as características dessa rede selecionada são as seguintes:

Um Especialista	{	Neurônios na camada de entrada: 28
		Neurônios na camada oculta: 12
		Neurônios na camada de saída: 1
		Experimento: 4
		Inicialização de pesos: 9

### 5.3.3 Resultados das Simulações - Domínio da Freqüência - Modelo 1

Para a seleção da melhor rede foram analisados os erros de classificação obtidos nas simulações dos melhores experimentos. Foi selecionado o menor valor de erro de classificação encontrado para a Rede 1 da Tabela 5.10 e da Tabela 5.11. O menor erro de classificação encontrado para a referida rede foi de 79,167 %, isto é, dentre os 120 exemplos que compõem o conjunto de teste, a referida rede errou 95 exemplos. Logo as características dessa rede selecionada são as seguintes:

$$\text{Um Especialista} \left\{ \begin{array}{l} \text{Neurônios na camada de entrada: 21} \\ \text{Neurônios na camada oculta: 12} \\ \text{Neurônios na camada de saída: 1} \\ \text{Experimento: 2} \\ \text{Inicialização de pesos: 1} \end{array} \right.$$

#### 5.3.4 Resultados das Simulações - Centróide - Modelo 2

Para a seleção da melhor rede foram analisados os erros de classificação obtidos nas simulações dos melhores experimentos. Foi selecionado o menor valor de erro de classificação encontrado para a Rede 1 da Tabela 5.4, da Tabela 5.5 e da Tabela 5.6. O menor erro de classificação encontrado para a referida rede foi de 2,500 %, isto é, dentre os 120 exemplos que compõem o conjunto de teste, a referida rede errou 3 exemplos. Logo as características dessa rede selecionada são as seguintes:

$$\text{Um Especialista} \left\{ \begin{array}{l} \text{Neurônios na camada de entrada: 9} \\ \text{Neurônios na camada oculta: 11} \\ \text{Neurônios na camada de saída: 6} \\ \text{Experimento: 5} \\ \text{Inicialização de pesos: 9} \end{array} \right.$$

#### 5.3.5 Resultados das Simulações - Parâmetros Estatísticos - Modelo 2

Para a seleção da melhor rede foram analisados os erros de classificação obtidos nas simulações dos melhores experimentos. Foi selecionado o menor valor de erro de classificação encontrado para a Rede 1 da Tabela 5.8 e da Tabela 5.9. O menor erro de classificação encontrado para a referida rede foi de 36,667 %, isto é, dentre os 120 exemplos que compõem o conjunto de teste, a referida rede errou 44 exemplos. Logo as características dessa rede selecionada são as seguintes:

Um Especialista	{	Neurônios na camada de entrada: 28
		Neurônios na camada oculta: 9
		Neurônios na camada de saída: 6
		Experimento: 5
		Inicialização de pesos: 10

### 5.3.6 Resultados das Simulações - Domínio da Freqüência - Modelo 2

Para a seleção da melhor rede foram analisados os erros de classificação obtidos nas simulações dos melhores experimentos. Foi selecionado o menor valor de erro de classificação encontrado para a Rede 1 da Tabela 5.12 e da Tabela 5.13. O menor erro de classificação encontrado para a referida rede foi de 51,667 %, isto é, dentre os 120 exemplos que compõem o conjunto de teste, a referida rede errou 62 exemplos. Logo as características dessa rede selecionada são as seguintes:

Um Especialista	{	Neurônios na camada de entrada: 21
		Neurônios na camada oculta: 14
		Neurônios na camada de saída: 6
		Experimento: 5
		Inicialização de pesos: 8

## 5.4 Resultados das Simulações com Três Especialistas

Uma vez definida a arquitetura, foram realizados os treinamentos com a técnica de validação cruzada, k-fold. Para o Modelo 3 e o Modelo 4, com seis neurônios na camada de saída, foram selecionados os menores valores de erro de classificação da média das saídas da Rede 1, Rede 2 e Rede 3 para os melhores experimentos.

#### 5.4.1 Resultados das Simulações - Centróide - Modelo 3

Neste modelo, ao invés de classificar um dado padrão de entrada com base em apenas um especialista, a classificação foi realizada a partir de três especialistas diferentes desenvolvidos a partir do Modelo 1, com um neurônio na camada de saída. Serão combinadas todas as melhores inicializações, encontradas na Tabela 5.1, na Tabela 5.2 e na Tabela 5.3, para cada experimento das três melhores redes.

Os desempenhos das combinações dos três especialistas na formulação deste modelo é apresentado na Tabela 5.14.

Tabela 5.14: EC das redes combinadas - Centróide - Modelo 3

Experimento	Inicializações			EC
	Rede 1	Rede 2	Rede 3	
3	6	10	1	16,667
4	6	2	4	17,500
4	6	5	4	20,833
9	8	10	8	<b>15,833</b>
9	9	10	8	20,000

O menor erro de classificação encontrado a partir da simulação das combinações forneceu um resultado igual a 15,833 %, o que representa errar 19 exemplos numa população de 120 exemplos do conjunto de teste. Os especialistas desta melhor combinação apresentam o seguinte perfil:

Especialista 1	{	Neurônios na camada de entrada: 9
		Neurônios na camada oculta: 6
		Neurônios na camada de saída: 1
		Experimento: 9
		Inicialização de pesos: 8

Especialista 2	Neurônios na camada de entrada: 9
	Neurônios na camada oculta: 5
	Neurônios na camada de saída: 1
	Experimento: 9
	Inicialização de pesos: 10
Especialista 3	Neurônios na camada de entrada: 9
	Neurônios na camada oculta: 9
	Neurônios na camada de saída: 1
	Experimento: 9
	Inicialização de pesos: 8

#### 5.4.2 Resultados das Simulações - Parâmetros Estatísticos - Modelo 3

Neste modelo, ao invés de classificar um dado padrão de entrada com base em apenas um especialista, a classificação foi realizada a partir de três especialistas diferentes desenvolvido a partir do Modelo 1, com um neurônio na camada de saída. Serão combinadas todas as melhores inicializações, encontradas na Tabela 5.7, para cada experimento das três melhores redes.

Os desempenhos das combinações dos três especialistas na formulação deste modelo é apresentado na Tabela 5.15.

Tabela 5.15: EC das redes combinadas - Parâmetros Estatísticos - Modelo 3

Experimento	Inicializações			EC
	Rede 1	Rede 2	Rede 3	
4	9	4	10	79,167
4	9	9	10	<b>78,333</b>

O erro de classificação encontrado a partir da simulação das combinações forneceu um resultado igual a 78,333 %, o que representa errar 94 exemplos numa população

de 120 exemplos do conjunto de teste. Os especialistas desta melhor combinação apresentam o seguinte perfil:

Especialista 1	Neurônios na camada de entrada: 28
	Neurônios na camada oculta: 12
	Neurônios na camada de saída: 1
	Experimento: 4
	Inicialização de pesos: 9
Especialista 2	Neurônios na camada de entrada: 28
	Neurônios na camada oculta: 10
	Neurônios na camada de saída: 1
	Experimento: 4
	Inicialização de pesos: 9
Especialista 3	Neurônios na camada de entrada: 28
	Neurônios na camada oculta: 13
	Neurônios na camada de saída: 1
	Experimento: 4
	Inicialização de pesos: 10

#### 5.4.3 Resultados das Simulações - Domínio da Frequência - Modelo 3

Neste modelo, ao invés de classificar um dado padrão de entrada com base em apenas um especialista, a classificação foi realizada a partir de três especialistas diferentes desenvolvido a partir do Modelo 1, com um neurônio na camada de saída. Serão combinadas todas as melhores inicializações, encontradas na Tabela 5.10 e na Tabela 5.11, para cada experimento das três melhores redes.

Os desempenhos das combinações dos três especialistas na formulação deste modelo é apresentado na Tabela 5.16.

Tabela 5.16: EC das redes combinadas - Domínio da Frequência - Modelo 3

Experimento	Inicializações			EC
	Rede 1	Rede 2	Rede 3	
2	1	2	3	<b>78,333</b>
2	1	4	3	79,167
7	6	2	10	80,833
7	7	2	10	80,000
7	9	2	10	80,833
7	10	2	10	79,167

O menor erro de classificação encontrado a partir da simulação das combinações forneceu um resultado igual a 78,333 %, o que representa errar 94 exemplos numa população de 120 exemplos do conjunto de teste. Os especialistas desta melhor combinação apresentam o seguinte perfil:

Especialista 1	{	Neurônios na camada de entrada: 21
		Neurônios na camada oculta: 12
		Neurônios na camada de saída: 1
		Experimento: 2
		Inicialização de pesos: 1
Especialista 2	{	Neurônios na camada de entrada: 21
		Neurônios na camada oculta: 6
		Neurônios na camada de saída: 1
		Experimento: 2
		Inicialização de pesos: 2
Especialista 3	{	Neurônios na camada de entrada: 21
		Neurônios na camada oculta: 5
		Neurônios na camada de saída: 1
		Experimento: 2
		Inicialização de pesos: 3

#### 5.4.4 Resultados das Simulações - Centróide - Modelo 4

Neste modelo, ao invés de classificar um dado padrão de entrada com base em apenas um especialista, a classificação foi realizada a partir de três especialistas diferentes desenvolvidos a partir do Modelo 2, com seis neurônios na camada de saída. Serão combinadas todas as melhores inicializações, encontradas na Tabela 5.4, na Tabela 5.5 e na Tabela 5.6, para cada experimento das três melhores redes.

Os desempenhos das combinações dos três especialistas na formulação deste modelo é apresentado na Tabela 5.17.

Tabela 5.17: EC das redes combinadas - Centróide - Modelo 4

Experimento	Inicializações			EC
	Rede 1	Rede 2	Rede 3	
1	3	10	5	<b>0,833</b>
1	3	10	10	1,667
1	6	10	5	1,667
1	6	10	10	2,500
5	9	8	2	1,667
7	9	3	3	2,500
7	9	3	9	1,667

O menor erro de classificação encontrado a partir da simulação das combinações forneceu um resultado igual a 0,833 %, o que representa errar 1 exemplo numa população de 120 exemplos do conjunto de teste. Os especialistas desta melhor combinação apresentam o seguinte perfil:

Especialista 1	{	Neurônios na camada de entrada: 9
		Neurônios na camada oculta: 11
		Neurônios na camada de saída: 6
		Experimento: 1
		Inicialização de pesos: 3

Especialista 2	Neurônios na camada de entrada: 9
	Neurônios na camada oculta: 12
	Neurônios na camada de saída: 6
	Experimento: 1
	Inicialização de pesos: 10
Especialista 3	Neurônios na camada de entrada: 9
	Neurônios na camada oculta: 13
	Neurônios na camada de saída: 6
	Experimento: 1
	Inicialização de pesos: 5

#### 5.4.5 Resultados das Simulações - Parâmetros Estatísticos - Modelo 4

Neste modelo, ao invés de classificar um dado padrão de entrada com base em apenas um especialista, a classificação foi realizada a partir de três especialistas diferentes desenvolvidos a partir do Modelo 2, com seis neurônios na camada de saída. Serão combinadas todas as melhores inicializações, encontradas na Tabela 5.8 e na Tabela 5.9, para cada experimento das três melhores redes.

Os desempenhos das combinações dos três especialistas na formulação deste modelo é apresentado na Tabela 5.18.

Tabela 5.18: EC das redes combinadas - Parâmetros Estatísticos - Modelo 4

Experimento	Inicializações			EC
	Rede 1	Rede 2	Rede 3	
4	2	8	3	44,167
5	10	10	3	<b>38,300</b>
5	10	10	5	41,667

O menor erro de classificação encontrado a partir da simulação das combinações forneceu um resultado igual a 38,300 %, o que representa errar 46 exemplos numa

população de 120 exemplos do conjunto de teste. Os especialistas desta melhor combinação apresentam o seguinte perfil:

Especialista 1	Neurônios na camada de entrada: 28
	Neurônios na camada oculta: 9
	Neurônios na camada de saída: 6
	Experimento: 5
	Inicialização de pesos: 10
Especialista 2	Neurônios na camada de entrada: 28
	Neurônios na camada oculta: 6
	Neurônios na camada de saída: 6
	Experimento: 5
	Inicialização de pesos: 10
Especialista 3	Neurônios na camada de entrada: 28
	Neurônios na camada oculta: 13
	Neurônios na camada de saída: 6
	Experimento: 5
	Inicialização de pesos: 3

#### 5.4.6 Resultados das Simulações - Domínio da Frequência - Modelo 4

Neste modelo, ao invés de classificar um dado padrão de entrada com base em apenas um especialista, a classificação foi realizada a partir de três especialistas diferentes desenvolvidos a partir do Modelo 2, com seis neurônios na camada de saída. Serão combinadas todas as melhores inicializações, encontradas na Tabela 5.12 e na Tabela 5.13, para cada experimento das três melhores redes.

Os desempenhos das combinações dos três especialistas na formulação deste modelo é apresentado na Tabela 5.19.

Tabela 5.19: EC das redes combinadas - Domínio da Frequência - Modelo 4

Experimento	Inicializações			EC
	Rede 1	Rede 2	Rede 3	
5	8	1	6	<b>47,500</b>
6	2	7	2	58,333
6	3	7	2	59,167

O menor erro de classificação encontrado a partir da simulação das combinações forneceu um resultado igual a 47,500%, o que representa errar 57 exemplos numa população de 120 exemplos do conjunto de teste. Os especialistas desta melhor combinação apresentam o seguinte perfil:

Especialista 1 {  
 Neurônios na camada de entrada: 21  
 Neurônios na camada oculta: 14  
 Neurônios na camada de saída: 6  
 Experimento: 5  
 Inicialização de pesos: 8

Especialista 2 {  
 Neurônios na camada de entrada: 21  
 Neurônios na camada oculta: 6  
 Neurônios na camada de saída: 6  
 Experimento: 5  
 Inicialização de pesos: 1

Especialista 3 {  
 Neurônios na camada de entrada: 21  
 Neurônios na camada oculta: 15  
 Neurônios na camada de saída: 6  
 Experimento: 5  
 Inicialização de pesos: 6

## 5.4.7 Síntese do Capítulo

Neste Capítulo foram simuladas as melhores redes encontradas no Capítulo 4, Tabela 4.22. Visando selecionar a rede que ofereça a melhor performance para o problema proposto, garantindo que o desempenho da uma determinada arquitetura não foi resultado de uma inicialização ótima. Será apresentada na Tabela 5.20 uma Síntese dos resultados obtidos para as melhores inicializações para cada técnica de extração de atributos.

Tabela 5.20: Síntese dos Resultados da Modelagem das Redes

			Melhores Inicializações		
			Rede 1	Rede 2	Rede 3
Centróide	Modelo 1	Exp 3	6	10	1
		Exp 4	6	2;5	4
		Exp 9	8;9	10	8
	Modelo 2	Exp 1	3;6	10	5;10
		Exp 5	9	8	2
		Exp 7	9	3	3;9
Parâmetros Estatísticos	Modelo 1	Exp 4	9	4;9	10
	Modelo 2	Exp 4	2	8	3
		Exp 5	10	10	3;5
Domínio da Frequência	Modelo 1	Exp 2	1	2;4	3
		Exp 7	6;7;9;10	2	10
	Modelo 2	Exp 5	8	1	6
		Exp 6	2;3	7	2

# 6

## MODELAGEM E SIMULAÇÃO DAS REDES DE FUNÇÕES DE BASES RADIAIS

Como apresentado no Capítulo 5, a rede MLP que obteve o menor erro de classificação tem a entrada advinda da extração de atributos pela técnica do Centróide. Logo este presente capítulo tem como finalidade modelar e simular as redes de Função de Base Radial (RBF) para o melhor conjunto de entrada visando uma análise comparativa com a rede MLP.

### 6.1 Introdução Teórica

Neste capítulo, serão utilizadas redes RBF que são redes multicamadas em que a ativação do nodo é função da distância euclidiana entre seus vetores de entrada e de peso, como descrito na Seção 2.3. Para modelagem da referida rede no MATLAB é necessário o ajuste da variável *SPREAD* que será apresentada abaixo.

O *SPREAD* tem como principal objetivo ajustar a função de Base Radial, suavizando-a, ou não, dependendo do valor a ele fornecido. Cada bias  $b$  da primeira camada é ajustado para  $b = \frac{0,8326}{SPREAD}$ . Isto dará uma função de base radial que permite uma saída acima de 0,5 quando as entradas das funções de bases radiais, denominada *radbas* no MATLAB, tiverem valores de  $\pm SPREAD$ .

Pela Figura 6.1 temos que  $f(n) = a$ , logo se quisermos limitar a saída  $a$  entre valores

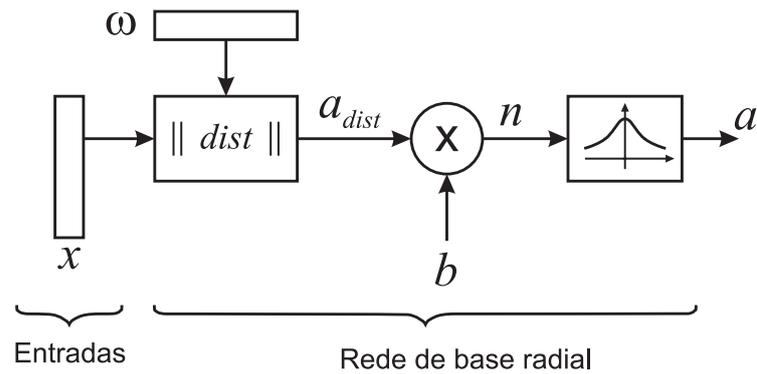


Figura 6.1: Diagrama de um classificador de padrões

superiores a 0,5 deveremos definir o SPREAD da seguinte forma:

$$f(n) = a,$$

onde,

$$a \geq 0,5$$

$$n = a_{dist} \cdot b.$$

Logo,

$$\begin{aligned} f(a_{dist} \cdot b) &\geq 0,5 \\ b &= \frac{0,8326}{SPREAD} \\ f\left(a_{dist} \cdot \frac{0,8326}{SPREAD}\right) &\geq 0,5 \end{aligned}$$

(6.1)

Pelo gráfico da função *radbas*, Figura 6.2, poderemos chegar a seguinte conclusão:

$$\begin{aligned} -1 &\leq \frac{a_{dist}}{SPREAD} \leq +1 \\ -SPREAD &\leq a_{dist} \leq +SPREAD \end{aligned}$$

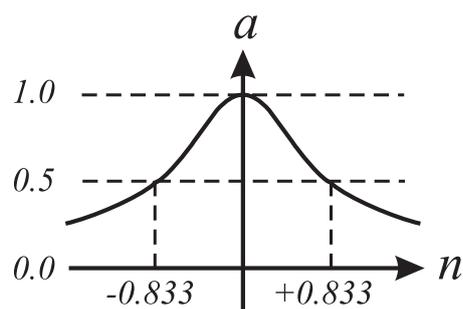


Figura 6.2: Função de base radial

## 6.2 Seleção dos Melhores Experimentos

Nesta etapa da modelagem serão selecionados os melhores experimentos das redes, fixando a priori o valor do *SPREAD* ( $SPREAD=1$ ) para todas as redes. A seleção do melhor experimento foi baseada no menor erro de classificação.

Quando são obtidos mais de um valor mínimo do erro de classificação será considerado apenas um, selecionado de forma aleatória qualquer um dos que obtiveram menor resultado, como sendo o melhor experimento que será utilizado na próxima etapa da seleção do melhor *SPREAD*.

### 6.2.1 Melhores Experimentos - RBF Exata - Modelo 1

Na Tabela 6.1 são apresentados os resultados obtidos das 10 simulações para os 10 experimentos distintos. Para redes com um neurônio na camada de saída os que obtiveram melhor performance foram o Experimento 2 e o Experimento 4.

Tabela 6.1: Seleção do melhor experimento - RBF Exata - Modelo 1

Experimento	Erro de Classificação
1	36,667
<b>2</b>	<b>34,167</b>
3	36,667
<b>4</b>	<b>34,167</b>
5	37,500
6	39,167
7	38,333
8	41,667
9	39,167
10	39,167

### 6.2.2 Melhores Experimentos - RBF Exata - Modelo 2

Na Tabela 6.2 são apresentados os resultados obtidos das 10 simulações para os 10 experimentos distintos. Para redes com seis neurônios na camada de saída os que obtiveram melhor performance foram o experimento 1.

Tabela 6.2: Seleção do melhor experimento - RBF Exata - Modelo 2

Experimento	Erro de Classificação
<b>1</b>	<b>10,833</b>
2	15,833
3	20,000
4	25,833
5	18,333
6	12,500
7	11,667
8	15,000
9	14,167
10	15,833

### 6.2.3 Melhores Experimentos - RBF Eficiente - Modelo 1

Na Tabela 6.3 são apresentados os resultados obtidos das 10 simulações para os 10 experimentos distintos. Para redes com um neurônio na camada de saída os que obtiveram melhor performance foram o experimento 5, o experimento 8 e o experimento 10.

Tabela 6.3: Seleção do melhor experimento - RBF Eficiente - Modelo 1

<b>Experimento</b>	<b>Erro de Classificação</b>
1	47,500
2	37,500
3	35,833
4	41,667
<b>5</b>	<b>35,000</b>
6	42,500
7	38,333
<b>8</b>	<b>35,000</b>
9	39,167
<b>10</b>	<b>35,000</b>

### 6.2.4 Melhores Experimentos - RBF Eficiente - Modelo 2

Na Tabela 6.4 são apresentados os resultados obtidos das 10 simulações para os 10 experimentos distintos. Para redes com seis neurônios na camada de saída os que obtiveram melhor performance foram o experimento 3 e o experimento 7.

Tabela 6.4: Seleção do melhor experimento - RBF Eficiente - Modelo 2

Experimento	Erro de Classificação
1	11,667
2	12,500
<b>3</b>	<b>10,000</b>
4	17,500
5	16,667
6	15,000
<b>7</b>	<b>10,000</b>
8	10,833
9	15,833
10	17,500

### 6.2.5 Melhores Experimentos - RBF Probabilística - Modelo 1

Na Tabela 6.3 são apresentados os resultados obtidos das 10 simulações para os 10 experimentos distintos. Para redes com um neurônio na camada de saída os que obtiveram melhor performance foram o experimento 3 e o experimento 10.

Tabela 6.5: Seleção do melhor experimento - RBF Probabilística - Modelo 1

Experimento	Erro de Classificação
1	33,333
2	33,333
<b>3</b>	<b>31,667</b>
4	34,167
5	35,000
6	35,000
7	36,667
8	33,333
9	35,833
<b>10</b>	<b>31,667</b>

### 6.3 Seleção da Melhor Constante do *SPREAD*

Nesta etapa da modelagem serão selecionadas as melhores constantes de *SPREAD* para melhor experimento selecionado na etapa anterior, Seção 6.2. Onde agora o valor do experimento foi fixado, o que obteve um melhor desempenho, e os valores da constante do *SPREAD* foi variada de 0,1 a 2, em intervalos de 0,1. A seleção da melhor constante de *SPREAD* também foi baseada no menor erro de classificação.

#### 6.3.1 Melhor *SPREAD* - RBF Exata - Modelo 1

Na Tabela 6.6 são apresentados os resultados das 20 simulações, para os 20 *SPREAD* distintos, realizadas para o experimento 4.

Tabela 6.6: Seleção do SPREAD - RBF Exata - Modelo 1

Spread	Erro de Classificação	Spread	Erro de Classificação
0,1	59,167	1,1	34,167
0,2	50,833	1,2	35,833
0,3	40,000	1,3	35,833
0,4	38,333	1,4	37,500
0,5	39,167	1,5	30,000
0,6	36,667	<b>1,6</b>	<b>28,333</b>
0,7	36,667	1,7	36,667
0,8	35,000	1,8	40,000
0,9	35,000	1,9	38,333
1	34,167	2	31,667

#### 6.3.2 Melhor *SPREAD* - RBF Exata - Modelo 2

Na Tabela 6.7 são apresentados os resultados das 20 simulações, para os 20 *SPREAD* distintos, realizadas para o experimento 1.

Tabela 6.7: Seleção do SPREAD - RBF Exata - Modelo 2

Spread	Erro de Classificação	Spread	Erro de Classificação
0,1	54,167	1,1	10,833
0,2	23,333	1,2	10,833
0,3	15,000	<b>1,3</b>	<b>7,500</b>
0,4	15,833	1,4	10,833
0,5	16,667	1,5	10,833
0,6	20,833	1,6	11,667
0,7	14,167	1,7	13,333
0,8	14,167	1,8	12,500
0,9	12,500	1,9	14,167
1	10,833	2	14,167

### 6.3.3 Melhor *SPREAD* - RBF Eficiente - Modelo 1

Na Tabela 6.8 são apresentados os resultados das 20 simulações, para os 20 *SPREAD* distintos, realizadas para o experimento 5.

Tabela 6.8: Seleção do SPREAD - RBF Eficiente - Modelo 1

Spread	Erro de Classificação	Spread	Erro de Classificação
0,1	61,667	1,1	33,333
0,2	45,833	1,2	35,000
0,3	40,000	1,3	35,833
0,4	45,000	1,4	37,500
0,5	35,833	1,5	32,500
0,6	35,833	<b>1,6</b>	<b>31,667</b>
0,7	32,500	1,7	32,500
0,8	37,500	1,8	35,000
0,9	32,500	1,9	35,833
1	35,000	2	33,333

6.3.4 Melhor *SPREAD* - RBF Eficiente - Modelo 2

Na Tabela 6.9 são apresentados os resultados das 20 simulações, para os 20 *SPREAD* distintos, realizadas para o experimento 3.

Tabela 6.9: Seleção do SPREAD - RBF Eficiente - Modelo 2

Spread	Erro de Classificação	Spread	Erro de Classificação
0,1	54,167	1,1	10,000
0,2	22,500	1,2	10,833
0,3	15,000	1,3	10,000
0,4	12,500	1,4	10,000
<b>0,5</b>	<b>8,333</b>	1,5	9,167
0,6	9,167	1,6	10,833
0,7	10,000	1,7	9,167
0,8	9,167	1,8	10,000
0,9	9,167	1,9	9,167
1	10,000	2	10,833

6.3.5 Melhor *SPREAD* - RBF Probabilística - Modelo 1

Na Tabela 6.10 são apresentados os resultados das 20 simulações, para os 20 *SPREAD* distintos, realizadas para o experimento 3.

Tabela 6.10: Seleção do SPREAD - RBF Probabilística - Modelo 1

Spread	Erro de Classificação	Spread	Erro de Classificação
<b>0,1</b>	<b>31,667</b>	1,1	67,500
0,2	32,500	1,2	70,000
0,3	41,667	1,3	72,500
0,4	46,667	1,4	81,667
0,5	47,500	1,5	85,000
0,6	49,167	1,6	86,667
0,7	52,500	1,7	86,667
0,8	57,500	1,8	86,667
0,9	63,333	1,9	86,667
1	64,167	2	86,667

# 7

## CONCLUSÕES E TRABALHOS FUTUROS

Neste Capítulo serão apresentados os melhores resultados das Redes MLP, obtidos das simulações dos Modelo 1, Modelo 2, Modelo 3 e Modelo 4 para as três técnicas de extração de atributos apresentadas no presente trabalho: Centróide, Parâmetros Estatísticos e Domínio da Frequência. Serão também apresentados os melhores desempenhos das redes RBF, visando realizar apenas uma análise comparativa com os resultados obtidos pela rede MLP para a técnica do Centróide. Em seguida foram realizadas análises e justificativas dos resultados obtidos. E finalmente apresentadas sugestões de trabalhos futuros.

### 7.1 Conclusões

Os resultados aqui encontrados, Tabela 7.1, mostram a superioridade da codificação da rede utilizada com seis neurônios na saída, apresentada na Tabela 4.2, Modelo 2, comparando-se à codificação simples, que utiliza apenas uma saída para codificar as classes, Modelo 1. Outro resultado importante obtido neste trabalho foi à observação do comportamento dos modelos formados por combinação de especialistas. Além disso, quando se compara os resultados encontrados neste trabalho com outras pesquisas desenvolvidas neste âmbito [5], demonstra a potencialidade do uso de RNA. Para todas as redes treinadas e simuladas neste trabalho foi utilizado o critério de parada

antecipada (*early stop*) assegurando a generalização da rede.

Tabela 7.1: Resumo dos resultados das simulações das redes MLP

<b>Erro de Classificação</b>			
	Centróide	Parâmetros Estatísticos	Domínio da Freqüência
Modelo 1	19,167	77,500	79,167
Modelo 2	2,500	36,667	51,667
Modelo 3	15,833	78,333	78,333
Modelo 4	0,833	38,300	47,500

Logo pode-se verificar que a rede que apresentou a melhor performance foi a rede com a técnica de extração de atributos do Centróide. Serão destacadas abaixo as principais características do melhor modelo:

Erro de Classificação: 0,83%;

Tipo de Rede: MLP;

Técnica de extração de atributos: Centróide;

Redes combinadas (três especialistas);

Algoritmo de treinamento: Levenberg-Marquardt;

Experimento da validação cruzada: Experimento 1;

Quantidade de neurônios na camada de entrada: 9 neurônios;

Quantidade de neurônios ocultos da Rede 1: 11 neurônios;

Quantidade de neurônios ocultos da Rede 2: 12 neurônios;

Quantidade de neurônios ocultos da Rede 3: 13 neurônios;

Quantidade de neurônios na camada de saída: 6 neurônios;

Através do resumo das melhores redes para cada tipo de extração de atributos e para cada modelo da rede, apresentado na Tabela 7.1, chegamos à conclusão que nem sempre a combinação de três especialistas obterá uma melhor performance, como é o caso das redes treinadas e simuladas com Parâmetros Estatísticos. Pode-se justificar um pior desempenho da rede com três especialistas, cuja classificação é a média da saída da rede de cada especialista individual, através de um estudo probabilístico, onde a rede

com três especialistas só classificará corretamente um exemplo se pelo menos dois dos especialistas individuais classificarem corretamente. Logo, através do referido estudo obteve-se o seguinte gráfico, Figura 7.1, que representa a faixa de valores possíveis do erro de classificação para as redes combinadas a partir dos erros de classificação das redes isoladas.

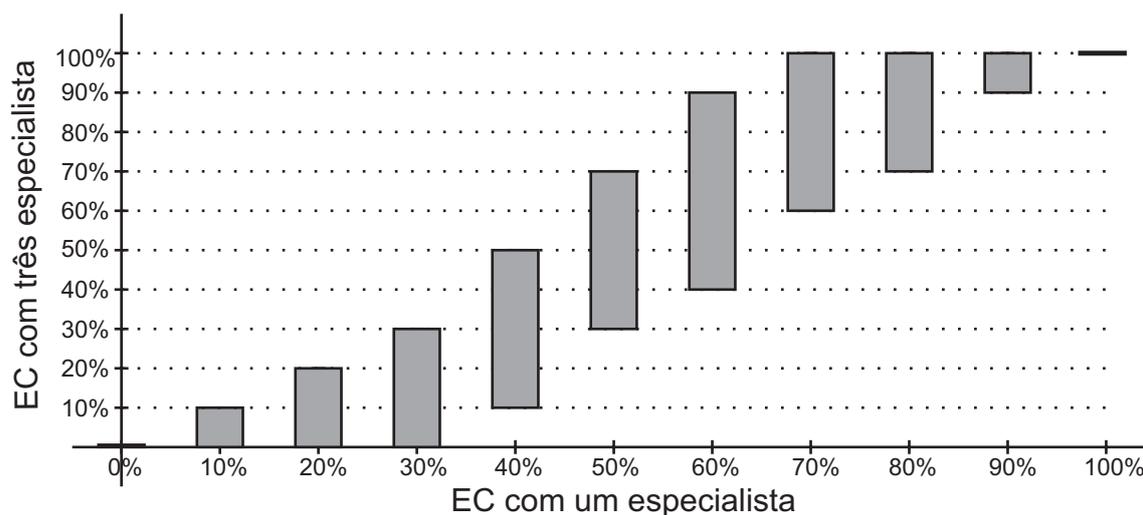


Figura 7.1: Faixa de variação do EC da rede combinada

Logo, só podemos assegurar que o EC da rede com três especialistas será obrigatoriamente menor ou igual ao EC da rede com apenas um especialista se o valor do referido EC for inferior, ou igual, a 30%. Quanto menor o EC da rede com um especialista, maior a probabilidade de ser menor o EC da rede combinada, como ilustra o gráfico apresentado na Figura 7.1.

Também da Tabela 7.1 pode-se observar um melhor desempenho das redes com seis neurônios na camada de saída do que das redes com apenas um neurônio. A rede neural com um neurônio da camada de saída apresenta uma maior sensibilidade no ajuste dos pesos, haja vista que sua classificação requer a saída da rede em um pequeno intervalo de valores. O que não ocorre com as redes com seis neurônios na camada oculta, que utiliza a técnica *winner takes all*, onde o neurônio que apresentar o maior valor é o que classificará o padrão.

Uma outra análise que pode ser feita a partir dos dados apresentados na Tabela

7.1 é que as redes que receberam como sinais de entrada os atributos extraídos pelas técnicas do centróide obtiveram melhor performance do que as redes que tiveram como entrada os Parâmetros Estatísticos e os no Domínio da Frequência. Esta divergência de resultados é devida à relação [17] entre a quantidade de parâmetros livres da rede ( $W$ ) e o tamanho do conjunto de treinamento ( $N$ ),

$$N = O\left(\frac{W}{\epsilon}\right) \quad (7.1)$$

Onde:

$O(\cdot) \rightsquigarrow$  Representa a ordem da quantidade;

$\epsilon \rightsquigarrow$  Erro de classificação do conjunto de teste.

Os atributos extraídos pela técnica do Centróide é consideravelmente menor que as demais aqui analisadas, implicando em uma quantidade inferior de parâmetros livres para um conjunto de treinamento de mesma dimensão, ou seja, um menor erro de classificação.

Como foi verificado um melhor desempenho das redes MLP cujas entradas foram originárias da extração de atributos pela técnica do Centróide. Foram realizados treinamentos e simulações para este melhor extrator de atributos, Centróide, para redes RBF, visando apenas uma análise comparativa entre os resultados obtidos pelas redes RBF e MLP. Os melhores resultados obtidos das simulações das redes RBF são apresentados na Tabela 7.2.

Tabela 7.2: Resumo dos resultados das simulações das redes RBF

<b>Erro de Classificação</b>			
	Exata	Eficiente	Probabilística
Modelo 1	28,333	31,667	31,667
Modelo 2	7,500	8,333	

## 7.2 Trabalhos Futuros

Como sugestão de trabalhos futuros seria a generalização de detecção de falhas em isoladores, não apenas ocasionadas pela poluição, como o caso do presente trabalho, mas também falhas ocasionadas por pequenos espaços vazios na região interna do isolador. Analisando se a metodologia de seleção da melhor rede neural, assim como as técnicas de extração de atributos, aqui abordadas são satisfatórias também no caso de descargas internas.

Visando reduzir o erro de classificação das redes que utilizaram como entrada sinais obtidos através de Dados Estatístico e do Domínio da Freqüência é aconselhável que haja uma redução da dimensão do vetor de entrada da rede ou um aumento na quantidade de exemplos do conjunto de treinamento, Equação 7.1.

Seria interessante também fazer uma análise comparativa com outras técnicas de extração de atributos, tais como Wavelet, Componentes Principais, Lógica Fuzzy, para o mesmo problema proposto neste trabalho. Pode-se também utilizar outra metodologia para extração dos Parâmetros Estatísticos e no Domínio da Freqüência, visando extrair informações que melhor caracterizem cada classe.

Outra sugestão, e não menos importante, seria realizar o estudo do desempenho do classificador no sistema elétrico real, com o objetivo de analisar influências externas.

Há também como sugestão a implantação em hardware da melhor rede de classificação de nível de poluição nos isoladores, deixando de ser imprescindível a utilização do computador para análise dos resultados. Sendo possível obter um resultado *in loco*, solucionando, seja com limpeza ou substituição do isolador, a iminente falha no sistema no exato momento da detecção do nível elevado de poluição.

## REFERÊNCIAS BIBLIOGRÁFICAS

- [1] K. Naito, Y. Mizuno, and W. Naganawa, “A study on probabilistic assessment os contamination flashover of high voltage insulator,” *IEEE Transaction on Power Delivery*, vol. 10, no. 3, pp. 1378–1384, July 1995.
- [2] A. I. Nigri, “Desempenho de linhas de transmissão - ponto de vista de manutenção,” *XV Seminário Nacional de Produção e Transmissão de Energia Elétrica*, 1999.
- [3] R. R. B. de Aquino, J. M. de Barros Bezerra, G. S. M. e Santos, O. N. Neto, M. M. da Silva Lira, A. A. Ferreira, and J. B. de Oliveira, “Aplicação de redes neurais para diagnóstico de poluição em isoladores de alta tensão,” *XIX Seminário Nacional de Produção e Transmissão de Energia Elétrica*, 2007.
- [4] J. M. B. Bezerra, “Caracterização de atributos de sinais para utilização de técnicas de reconhecimento de padrões na avaliação do isolamento de instalações e de equipamentos elétricos,” Tese de Doutorado, UFCG, 2004.
- [5] J. M. B. Bezerra, A. M. N. Lima, G. S. Deep, and E. G. Costa, “Estudo da distribuição de tensão em cadeias de isoladores aéreos de alta tensão em ambientes poluídos,” *XVII Seminário Nacional de Produção e Transmissão de Energia Elétrica*, 2003.
- [6] F. A. M. Rizk, “Mathematical models for pollution flashover,” *Electra*, vol. 78, pp. 71–103, 1981.
- [7] R. Sundararajan and R. S. Gorur, “Dynamic arc modeling of pollution flashover of insulators under dc voltage,” *IEEE Transactions on Electrical Insulation*, vol. 28, no. 2, pp. 209–218, April 1993.

- [8] I. R. Vazquez and J. L. F. Chavez, “Criteria for the diagnostic of polluted ceramic insulators based on the leakage current monitoring technique,” *IEEE Conference on Electrical Insulation and Dielectric Phenomena*, pp. 715–718, 1999.
- [9] CHESF, “Relatório de desempenho de linhas de transmissão,” 2008, documento interno.
- [10] D. R. de Melo, R. T. D. C. Jr, J. T. Batista, and M. E. Melo, “Avaliação do grau de poluição em instalações de transmissão, subestação e distribuição,” *Cepel*, 2006.
- [11] M. G. Danikas, N. Gao, and M. Aro, “Partial discharge recognition using neural networks: a review,” *Electrical Engineering. Springer-Verlag*, pp. 87–93, 2003.
- [12] R. Schalkoff, *Pattern Recognition Statistical, Structural and Neural Approaches*. Canada: John Wiley and Sons Inc., 1992.
- [13] J. M. B. Bezerra, “Aplicação de técnicas de reconhecimento de padrões no diagnóstico de defeitos em linhas de transmissão,” Dissertação de Mestrado, UFPE, 1995.
- [14] Wikipédia, “Coeficiente de variação — wikipédia, a enciclopédia livre,” 2008, [Online; accessed 5-Abril-2008]. [Online]. Available: [http://pt.wikipedia.org/w/index.php?title=Coeficiente\\_de\\_varia%C3%A7%C3%A3o&oldid=9968986](http://pt.wikipedia.org/w/index.php?title=Coeficiente_de_varia%C3%A7%C3%A3o&oldid=9968986)
- [15] —, “Valor eficaz — wikipédia, a enciclopédia livre,” 2008, [Online; accessed 5-Abril-2008]. [Online]. Available: [http://pt.wikipedia.org/w/index.php?title=Valor\\_eficaz&oldid=9421149](http://pt.wikipedia.org/w/index.php?title=Valor_eficaz&oldid=9421149)
- [16] A. A. Ferreira, “Comparação de arquiteturas de redes neurais para sistemas de reconhecimento de padrões em narizes artificiais,” Ph.D. dissertation, UFPE, 2004.
- [17] S. Haykin, *Redes Neurais: Princípios e Prática*. Editora bookman s.a., 2001.
- [18] A. P. Braga, T. B. Ludemir, and A. C. P. L. F. Carvalho, *Redes Neurais Artificiais: Teoria e Aplicações*. Editora LTC, 2000.

- [19] M. M. S. Lira, R. R. B. Aquino, A. A. Ferreira, M. A. C. Jr, and C. A. B. O. Lira, "Improving disturbance classification by combining multiple artificial neural networks," *IEEE World Congress on Computational Intelligence*, 2006.
- [20] H. Demuth, M. Beale, and M. Hagan, *Neural Network Toolbox 6 - User Guide*, The MathWorks, Inc., 2008.
- [21] W. Ullmann, "Descrição de processos e termos utilizados com ferramentas de medição fft," *dBW Ltda and Active Media Ltda*, 2007.
- [22] L. Prechelt, "A set of neural network benchmark problems and benchmarking rules," *Proben*, September 1994.
- [23] P. D. Wasserman, *Neural Computing: Theory and Practice*. Van Nostrand Reinhold, 1993.
- [24] S. O. Rezende, *Sistemas Inteligentes: Fundamentos e Aplicações*. Manole, 2005.
- [25] A. Yamazaki, "Uma metodologia para otimização de arquiteturas e pesos de redes neurais," Dissertação de Mestrado, UFPE, 2003.
- [26] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons, Inc., 1960.
- [27] R. Beale and T. Jackson, *Neural Computing*. Adam Hilger, 1991.
- [28] J. M. B. Bezerra and A. F. D. Silva, "Use of pattern recognition techniques in transmission lines defects monitoring," *IEEE/KTH Stockholm Power Tech Conference*, 1995.
- [29] J. M. B. Bezerra, S. S. Deep, and A. M. N. Lima, "Application of risk criteria on transmission system insulation upgrade," *IEEE Conference on Electrical Insulation and Dielectric Phenomena*, 2002.
- [30] M. A. R. M. Fernando and S. M. Gubanski, "Leakage current patterns on artificially polluted composite insulators," *IEEE Conference on Electrical Insulation and Dielectric Phenomena*, 1996.

- [31] P. G. Hoel, *Estatística Matemática*. Editora Guanabara Dois S.A., 1980.
- [32] G. B. da Silva, “Sistema híbrido de previsão de carga elétrica em curto prazo utilizando redes neurais artificiais e lógica fuzzy,” Master’s thesis, UFPE, 2006.
- [33] O. N. Neto, “Aplicação de redes neurais artificiais e algoritmos genéticos na previsão de carga elétrica em médio prazo,” Master’s thesis, UFPE, 2006.
- [34] J. Arrillaga and N. R. Watson, *Power System Harmonics*, second edition ed. John Wiley and Sons, Ltd, 2003.
- [35] A. V. Oppenheim and A. S. Willsky, *Signals & Systems*. Prentice Hall, 1997.