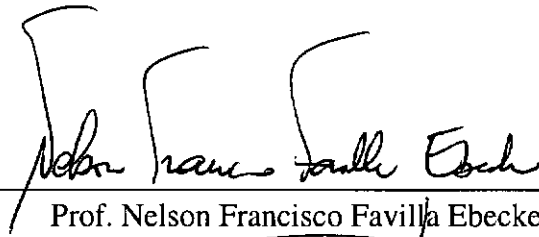


ESTUDO SOBRE REDES NEURAIIS COM FUNÇÃO DE BASE RADIAL

Olimpio Alves dos Santos

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA CIVIL.

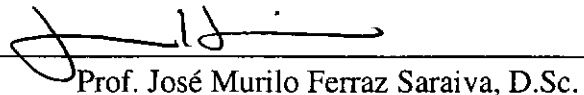
Aprovada por:



Prof. Nelson Francisco Favilla Ebecken, D.Sc.



Prof. Tiago Alberto Piedras Lopes, D.Sc.



Prof. José Murilo Ferraz Saraiiva, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

MARÇO DE 2000

DOS SANTOS, OLIMPIO ALVES

Um Estudo sobre Redes Neurais com Função
de Base Radial [Rio de Janeiro] 2000

VII, 81 p. 29,7 cm (COPPE/UFRJ, M.Sc.,
Engenharia Civil, 2000)

Tese - Universidade Federal do Rio de
Janeiro, COPPE

1. Função de Base Radial
2. Redes Neurais Artificiais

I. COPPE/UFRJ II. Título (série)

DEDICATÓRIA

Ao meu querido filho e amigo
Ramon Navarro de Oliveira Santos.

AGRADECIMENTOS

Ao professor Nelson Ebecken pela orientação, amizade e pela oportunidade dada em desenvolver este trabalho.

A minha esposa Virgínia pelo incentivo e apoio nas tarefas cotidianas.

Aos meus pais por terem me propiciado a oportunidade de estudar.

As minhas irmãs pelo incentivo.

A amiga Maria Inês pelo convívio durante todo este período, pelo incentivo durante a elaboração deste trabalho, pela digitação, críticas e sugestões.

A amiga Valéria Roitman pelo convívio e incentivo durante este período.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

UM ESTUDO SOBRE REDES NEURAIIS COM FUNÇÃO DE BASE RADIAL

Olimpio Alves dos Santos

Março/2000

Orientador: Nelson Francisco Favilla Ebecken

Programa: Engenharia Civil

O objetivo deste trabalho é apresentar uma introdução a um tipo de função de aproximação, denominada Redes Neurais com Função de Base Radial (RBFN's). Expõe-se a interpretação do modelo no campo das Redes Neurais Artificiais, no da Teoria da Regularização, no dos Sistemas de Inferência Fuzzy e no dos Núcleos de Regressão. Faz-se, também, uma discussão sobre alguns tipos algoritmos e os resultados alcançados com sua aplicação prática

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

A STUDY ABOUT THE PREDICTION OF THE QUALITY OF THE AAIR USING
NEURAL NETWORKS

Olimpio Alves dos Santos

March/2000

Advisor: Nelson Francisco Favilla Ebecken

Department: Civil Engineering

The objective of this work is to present an introduction to a type function of approach, called Radial Basis Functions Networks (RBFN's). the model interpretation is exposed in the field of Artificial Neural Networks, in the Theory of the Regularization, in the Fuzzy Inference Systems and in the one of the Kernel Regression. The work also present discussion about some algorithms and their results.

ÍNDICE

DEDICATÓRIA.....	III
AGRADECIMENTOS.....	IV
RESUMO	V
ABSTRACT	VI
ÍNDICE	VII
CAPÍTULO 1 - INTRODUÇÃO.....	1
CAPÍTULO 2 - REDES NEURAIS COM FUNÇÃO DE BASE RADIAL.....	5
2.1 - HISTÓRICO.....	5
2.2 - FUNÇÕES DE BASE RADIAL.....	6
2.3 - METODOLOGIAS UTILIZADAS EM REDES COM FUNÇÃO DE BASE RADIAL.....	8
2.3.1 - Metodologia segundo Moody e Darken.....	9
2.3.2 - Metodologia Conforme a Teoria de Regularização.....	12
2.3.2.1 - Construção da Funcional de Tikhonov.....	12
2.3.2.2 - Solução do Problema da Regularização.....	15
2.3.2.3 - Determinação dos Coeficientes de Expansão.....	15
2.3.2.4 - Considerações sobre as Funções de Green.....	18
2.3.3 - Redes Regularizadas.....	19
2.3.4 - Redes Generalizadas.....	20
2.3.3 - Metodologia Segundo os Estimadores de Regressão de Núcleo.....	24
2.3.4 - Equivalência de Sistemas Fuzzy e Redes Neurais com Função de Base Radial.....	25
2.4 - ALGORITMOS E ESTRATÉGIAS DE APRENDIZADO PARA RBFN.....	25
2.4.1 - Estratégias de Aprendizado.....	26
2.4.2 - Algoritmos.....	30
2.4.2.1 - Algoritmo Estático com a Pseudo-inversa matriz [23].....	30
2.4.2.2 - Algoritmo Estático com regra Delta [15].....	33
2.4.2.3 - Algoritmo Dinâmico.....	36
CAPÍTULO 3 - CARACTERIZAÇÃO DO PROBLEMA.....	41
3.1 - BASES DE DADOS UTILIZADAS.....	41
3.1.1 - Base de Dados "Plantas Iris".....	42

3.1.2 - Base de Dados "Aprovação de Crédito"	42
3.1.3 - Base de Dados "Predição de Sinais de Vibração"	44
CAPÍTULO 4 - ESTUDO DE CASO	45
4.1 - SOLUÇÃO DOS PROBLEMAS DE CLASSIFICAÇÃO E PREDIÇÃO	45
4.2 - PARÂMETROS UTILIZADOS PELOS SOFTWARES	46
4.2.1 - Parâmetros para o CASO 1	46
4.2.2 - Parâmetros para o CASO 2	48
4.3 - PARÂMETROS DE CONFIGURAÇÃO DAS REDES	49
4.3.1 - Redes para a Base de Dados Iris	49
4.3.2 - Redes para Base de Dados "Aprovação de Crédito"	50
4.3.3 - Rede para a Base de Dados "Predição de Sinais de Vibração"	51
CAPÍTULO 5 - ANÁLISE DOS RESULTADOS	52
5.1 - DESEMPENHO DAS RBFN'S	52
5.2 - AVALIAÇÃO DA PRECISÃO DOS RESULTADOS	53
5.2.1 - Classificação Percentual dos Padrões	53
5.2.2 - Raiz do Erro Médio Quadrático - RMSE	54
5.2.3 - Fator de Correlação	54
5.3 - APRESENTAÇÃO DOS RESULTADOS DA BASE DE DADOS "IRIS"	55
5.3.1 - Tabela com os Resultados	55
5.3.2 - Ilustração da Classificação	55
5.4 - APRESENTAÇÃO DOS RESULTADOS PARA A BASE DE DADOS "APROVAÇÃO DE CRÉDITO"	56
5.4.1 - Tabela com Resultados	56
5.4.2 - Ilustração da Classificação	56
5.5 - APRESENTAÇÃO DOS RESULTADOS BASE DE DADOS "PREDIÇÃO DE SINAIS DE VIBRAÇÃO"	58
5.5.1 - Tabela com Resultados	58
5.5.2 - Ilustração da Predição	58
5.6 - ANÁLISE DOS RESULTADOS	64
5.7 - RESULTADOS NUMÉRICOS OBTIDOS	65
5.7.1 - Base de Dados "Iris"	66
5.7.2 - Base de Dados "Aprovação de Crédito"	68
5.7.3 - Base de Dados "Predição de Sinais de Vibração"	70
6 - CONCLUSÃO	78
REFERÊNCIAS BIBLIOGRÁFICAS	80

CAPÍTULO 1 - INTRODUÇÃO

A motivação deste trabalho é a apresentação de um modelo de rede, denominado "Redes com Função de Base Radial". Capaz de desenvolver treinamentos bem rápidos e alcançar excelentes resultados.

Expõe-se neste estudo as interpretações desta rede no campo das Redes Neurais Artificiais, no da Teoria da Regularização, no dos Sistemas de Inferência Fuzzy e no dos Núcleos de Regressão. Mostra-se também alguns modelos de algoritmos de aprendizado e a performance que estes podem atingir, quando aplicados à casos práticos.

Neste estudo, a Rede com Função de Base Radial é vista, primordialmente, como um caso particular de redes Neurais Artificiais.

As Redes Neurais Artificiais são, basicamente, modelos matemáticos utilizados como função de aproximação ou na identificação de sistemas dinâmicos, os quais foram intensamente usados como modelos neurofisiológicos e de cognição, desde seu aparecimento. Como exemplo, pode-se citar o primeiro modelo matemático publicado, que tinha a intenção de ser uma formalização da atividade neural do cérebro [1], e os perceptrons que eram inspirados nas características do sistema visual [2]. Cita-se também o subtítulo dos dois livros mais representativos da época "*Explorations in the Microstructure of Cognition*", para mostrar como a atenção dos autores estava fortemente focalizada nas implicações cognitivas de seus modelos.

As Redes Artificiais resolveram uma classe de problemas que pode ser denominada como: Aprendizado por Indução. O aprendizado é uma atividade complexa desenvolvida por sistemas que podem adaptar-se, dependendo, para isto, de sua própria historia e experiência, e é um fenômeno natural. Pode-se considerar todos os animais como sistemas de aprendizado e, de uma forma mais geral, todos os seres vivos.

O ramo da Inteligência Artificial, denominado *Machine Learning*, tem o objetivo de construir sistemas artificiais que podem adaptar-se para resolver diferentes tarefas, ou pelo menos, um exemplo desta tarefa.

A idéia básica que envolve o aprendizado é a indução. A habilidade de inferir regras gerais a partir de observações particulares. De um ponto de vista puramente abstrato, pode-se dizer que aprendizado induzido significa construir um modelo geral de um fenômeno, a partir de um dado conjunto de exemplos específicos deste.

Um sistema de aprendizado adapta-se usando os dados disponíveis com o objetivo de aperfeiçoar seu desempenho em uma tarefa específica. A fim de alcançar este objetivo, o sistema constrói uma estrutura de conhecimento que permite a generalização para outros dados.

Distingue-se diferentes tipos de aprendizado, a primeira distinção é entre aprendizado supervisionado e não supervisionado. No aprendizado supervisionado, as informações sobre a função objetivo estão disponíveis. No caso mais simples, um conjunto de dados de entrada e saída da função são conhecidos, e a tarefa consiste em reconstruir a função a partir destes dados. Se a função é booleana, a tarefa corresponde a um problema de classificação, se a função é contínua, a tarefa é posta como um problema de regressão. No aprendizado não supervisionado, a função objetivo não está disponível, então o problema consiste em organizar os dados ou descobrir relações entre estes.

No caso particular de Redes Neurais com Função de Base Radial usa-se, para os algoritmos estáticos, o aprendizado não supervisionado e o supervisionado em duas etapas distintas. No caso de algoritmos dinâmicos, os dois tipos de aprendizado são realizados simultaneamente.

As Redes Neurais com Função de Base Radial são caracterizadas por ter nas unidades da camada escondida, funções de transferência com simetria radial. Na grande maioria dos casos, são utilizadas funções Gaussianas, desta maneira, a ativação da unidade tem um máximo no centro e um valor bem próximo de zero, longe deste. Esta particularidade propicia um caráter local às unidades da camada escondida, possibilitando sua modificação sem afetar a rede como um todo, tornando-as adequadas de serem implementadas com estratégias de aprendizado incremental.

O modelo tem sido aplicado na solução de problemas de classificação, reconhecimento de padrões, regressão, predição de séries temporais, reconhecimento de imagens, modelagem de fenômenos cognitivos, etc. Pode-se citar, como ilustração, dois exemplos. O primeiro é a aplicação na modelagem do fenômeno da comunicação que vem sendo desenvolvido no centro de Ciência da Cognição da Universidade de Turin. O segundo é um projeto da União Européia de Investigação das Rotas de Gelo no Oceano Ártico com vistas a automação das rotas dos navios.

O trabalho recente, mais importante, no campo das Redes Neurais com Função de Base Radial, foi o desenvolvimento do algoritmo para modelagem de bases com distribuição não estacionária de dados que podem ser encontradas em vários tipos de processos técnicos, biológicos ou econômicos.

A distribuição dos temas abordados, no corpo deste trabalho, está organizada como apresentado a seguir.

No Capítulo 1 faz-se uma breve introdução ao assunto das Redes com Função de Base Radial e das Redes Neurais Artificiais. Apresenta-se as principais aplicações e os trabalhos que estão sendo elaborados.

No Capítulo 2 expõe-se a arquitetura e as diferentes interpretações da Redes com Função de Base Radial. Estuda-se também três tipos de algoritmos aplicáveis a este modelo de rede.

No Capítulo 3 caracteriza-se o problema à resolver e descreve-se as bases de dados utilizadas.

No Capítulo 4 apresenta-se os parâmetros necessários a configuração dos algoritmos empregados, bem como a configuração das redes para cada caso.

No Capítulo 5 realiza-se análise dos resultados obtidos pelos algoritmos para cada uma das bases de dados estudadas.

No Capítulo 6 descreve-se as conclusões alcançadas no presente estudo e faz-se algumas sugestões para seu prosseguimento.

CAPÍTULO 2 - REDES NEURAIIS COM FUNÇÃO DE BASE RADIAL

2.1 - Histórico

A aplicação de funções de base radial para a construção de redes foi inicialmente introduzida com fundamentos da teoria da interpolação pelo trabalho de Broomhead e Lowe [3]. Praticamente ao mesmo tempo foram também apresentadas por Moody e Darken [4] através da teoria de Redes Neurais Artificiais como sendo uma alternativa aos perceptrons multicamadas. Posteriormente foram reintroduzidas por Poggio e Girosi [5] como um caso particular de Regularização de redes.

Conforme Roger Jang e C. T. Jun [6], a estrutura de rede proposta por Moody e Darken, baseia-se nos campos de sintonia localizada e de recepção superposta, que são estruturas já bem conhecidas e foram estudadas na região do córtex cerebral, do córtex visual, etc.

A figura 2.1 ilustra a posição desta estrutura no cérebro.

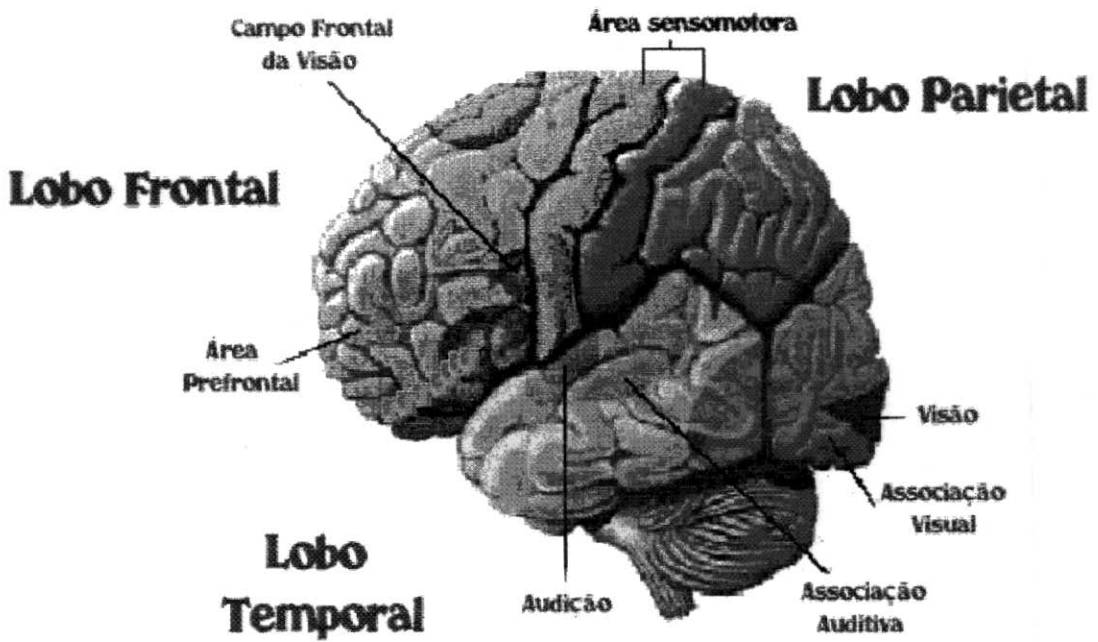


Figura 2.1 – Córtex Cerebral

Fonte : FURTADO⁷

2.2 - Funções de Base Radial

As funções de base radiais são uma classe especial de função. Sua denominação como radial deve-se ao fato de sua dependência com relação a variável de entrada x dar-se apenas em relação ao centro w_c .

A principal característica das funções de base radial é que sua resposta em relação a x cresce ou decresce monotonamente conforme a distancia de x em relação ao centro.

Existem diversos tipos de função de base radial, entre os quais pode-se destacar a Gaussiana, Multiquadrada, Multiquadrada Inversa e a de Cauchy, apresentadas nas equações de 2.1 a 2.4.

$$h(x) = e^{-\left[\frac{1}{2} \frac{(x-w_c)^2}{\sigma^2} \right]} \quad \text{Gaussiana} \quad (2.1)$$

$$h(x) = \frac{1}{\sqrt{(x-w_c)^2 + \sigma^2}} \quad \text{Multiquadrada} \quad (2.2)$$

$$h(x) = \frac{1}{\sqrt{(x - w_c)^2 + \sigma^2}} \quad \text{Multiquadrada Inversa} \quad (2.3)$$

$$h(x) = \frac{1}{(x - w_c)^2 + \sigma^2} \quad \text{Cauchy} \quad (2.4)$$

onde:

$(x - w_c)$, para valores escalares, é a distância da variável de entrada x ao centro w_c ;

σ é a amplitude da função.

As equações das funções de base radial citadas estão apresentadas na figura 2.2, onde o centro foi situado em zero e a amplitude σ é igual a unidade.

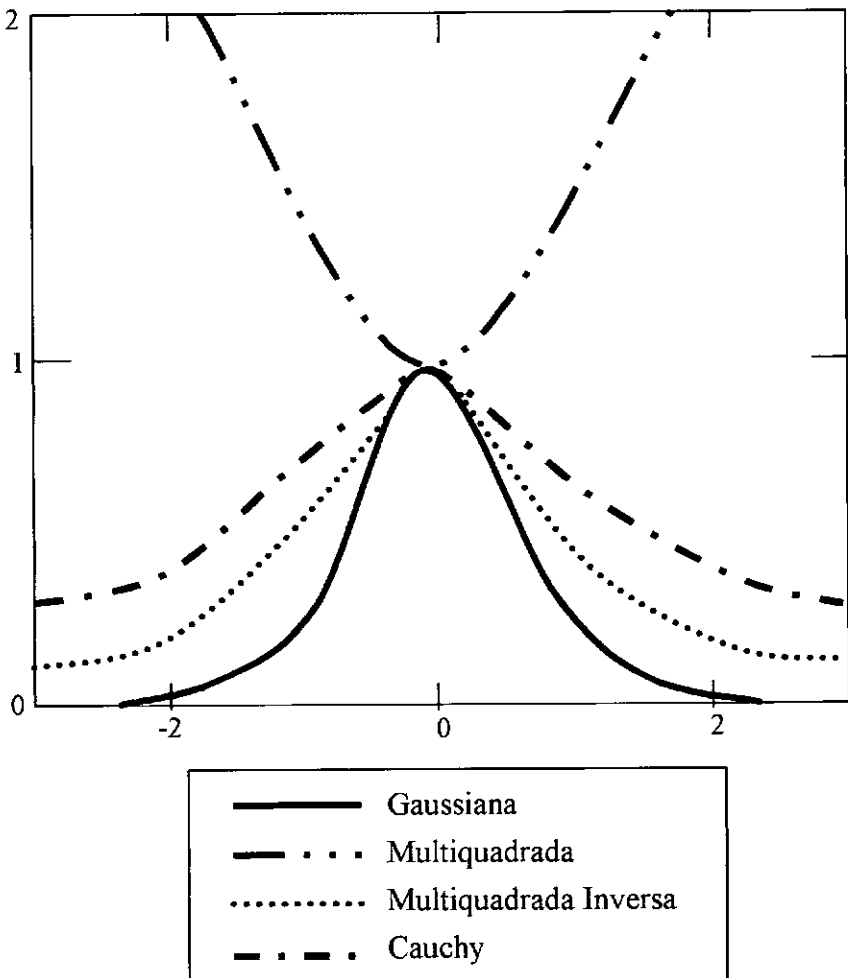


Figura 2.2 – RBFNs de raio unitário e centradas na origem.

Fonte : ORR⁸

Pode-se observar que as funções Gaussianas, Multiquadradas Inversas e de Cauchy possuem resposta local, isto é, tendem para o valor máximo quanto mais a entrada se aproxima do centro, e para zero quando x se afasta do centro.

Formalmente, pode-se escrever as equações 2.5 e 2.6:

$$\lim_{x \rightarrow 0} h(x) = 1 \quad (2.5)$$

$$\lim_{x \rightarrow \infty} h(x) = 0 \quad (2.6)$$

A função Multiquadrada possui resposta global, seu valor mínimo dá-se quando x se aproxima do centro e aumenta quando dele se afasta, como apresentado nas equações 2.7 e 2.8.

$$\lim_{x \rightarrow 0} h(x) = 1 \quad (2.7)$$

$$\lim_{x \rightarrow \infty} h(x) = \infty \quad (2.8)$$

Na estrutura das Redes com Função de Base Radial - RBFN's - as funções de base adquirem diferentes nomenclaturas na literatura existente. Nomes como "Campo de Recepção Localizada", "Campo de Recepção", "Unidade", "Célula" e "Nó" são empregados para designar as funções de base radial.

No presente estudo emprega-se indistintamente todas as designações existentes referindo-se às funções de base radial numa rede.

2.3 - Metodologias Utilizadas em Redes com Função de Base Radial

As arquiteturas das Redes com Função de Base Radial foram definidas independentemente de sua representação matemática e, freqüentemente, são apresentadas com nomes diferentes.

Como visto no item 2.1, foram inicialmente apresentadas com a metodologia das Redes Neurais Artificiais por Moody e Darken [4], como uma alternativa aos perceptrons multicamadas, pois apresentavam um treinamento mais rápido. Logo após,

reintroduzidas por Poggio e Girosi [5] como um caso particular de regularização de redes. Em trabalhos a parte, no campo da Lógica Fuzzy demonstrou-se, que sob certas condições, existe uma equivalência entre Redes com Função de Base Radial e Sistemas de Inferências Fuzzy.[6]

No campo da estatística aplicada, os problemas de regressão, classificação e, de uma forma geral, os de densidade de estimação foram desenvolvidos com base no método denominado "núcleo de regressão" que apresenta uma grande similaridade com a arquitetura das Redes Neurais com Função de Base Radial.

Finalmente, as RBFN's podem também ser inseridas nos modelos de aprendizado que são bem adequados para se trabalhar de forma incremental.

2.3.1 - Metodologia segundo Moody e Darken

A estrutura da rede proposta no trabalho "*Learning with Localized Receptive Fields*" [4] consiste de uma camada de entrada, uma escondida H com funções de ativação Gaussianas e uma de saída com unidades lineares.

As unidades de ativação da camada escondida são conectadas pelos pesos às unidades lineares da camada de saída.

Cada unidade de ativação c , pertencente à camada escondida, H, possui associado a ela uma amplitude σ e um vetor \vec{w}_c , pertencente ao espaço vetorial de entrada e que define a posição da unidade de ativação neste espaço. O vetor \vec{w}_c é também o centro da função Gaussiana e liga a camada escondida à camada de entrada.

A ativação das unidades Gaussianas é obtida da seguinte forma:

- Supõe-se um vetor de entrada \vec{x} pertencente ao espaço de entrada R^n ;
- Um vetor centro \vec{w}_c pertencente também ao espaço R^n .

A distância D_c entre o vetor de entrada \vec{x} e o centro \vec{w}_c é dada pela equação

$$D_c = \|\vec{x} - \vec{w}_c\| = \sqrt{\sum_{i=1}^n (x_i - w_{c_i})^2} \quad (2.9)$$

onde:

$\|\cdot\|$ indica a distância Euclidiana.

Desta forma, a ativação v_c da unidade Gaussiana para um vetor de entrada \vec{x} é dada pela equação 2.10.

$$v_c = \exp\left(-\frac{\|\vec{x} - \vec{w}_c\|^2}{2\sigma^2}\right) = \exp\left(-0,5 \frac{D_c^2}{\sigma^2}\right) \quad (2.10)$$

Arquitetura da Unidade de Ativação

A arquitetura da unidade de recepção está representada na figura 2.3.

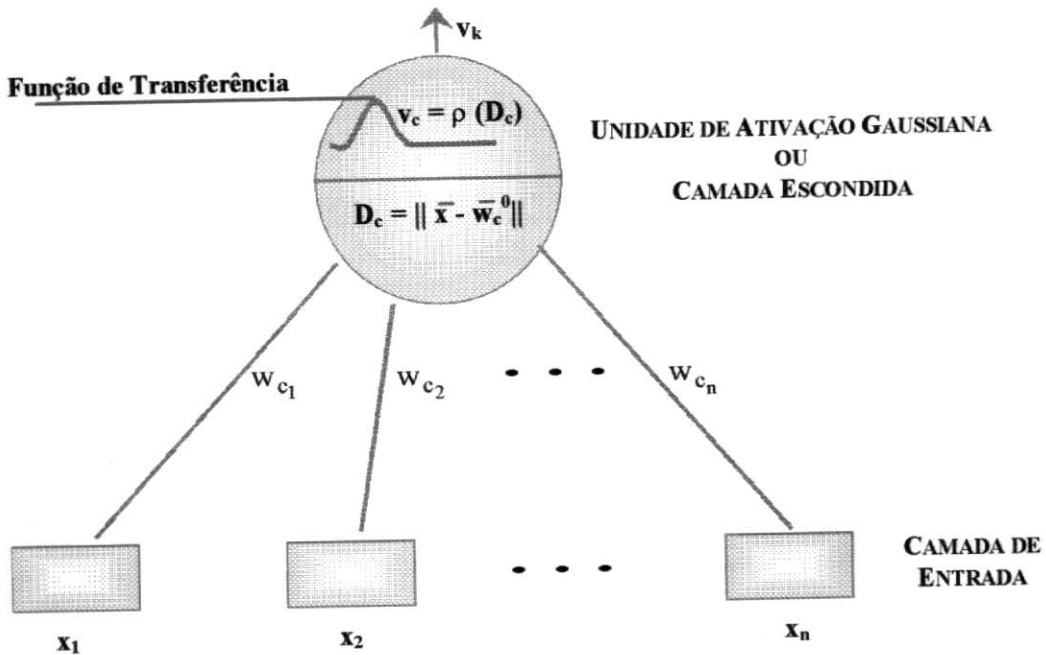


Figura 2.3 - Arquitetura da Unidade de Ativação

Nesta figura,

x_1, x_2, \dots, x_n são as componentes do vetor de entrada \vec{x} ;

$w_{c1}, w_{c2}, \dots, w_{cn}$ são as componentes do vetor centro \vec{w}_c .

O aprendizado na unidade de recepção consiste nos seguintes passos:

- i) Através de um algoritmo de clusterização, encontrar o vetor centro \bar{w}_c ;
- ii) Utilizar heurística (*Nearest Neighbor*) para determinar o valor de σ .

As unidades lineares da camada de saída executam apenas a soma dos seus pesos de entrada, que são determinados pela Regra Delta.

A saída y_i de uma rede com função de base radial pode ser computada de duas formas, considerando-a como única saída.

a) Não Normalizada

Considerando-se os pesos w , a saída não normalizada será dada pela equação 2.11.

$$y_i = \sum_{i=1}^{m_0} v_{c_i} \cdot \bar{w}_i \quad (2.11)$$

b) Normalizada

$$y_i = \frac{\sum_{i=1}^{m_0} v_{c_i} \cdot \bar{w}_i}{\sum_{i=1}^{m_0} v_{c_i}} \quad (2.12)$$

Na figura 2.4 apresenta-se a arquitetura de uma Rede Neural com função de base radial, conforme proposto por Moody e Darken. [9]

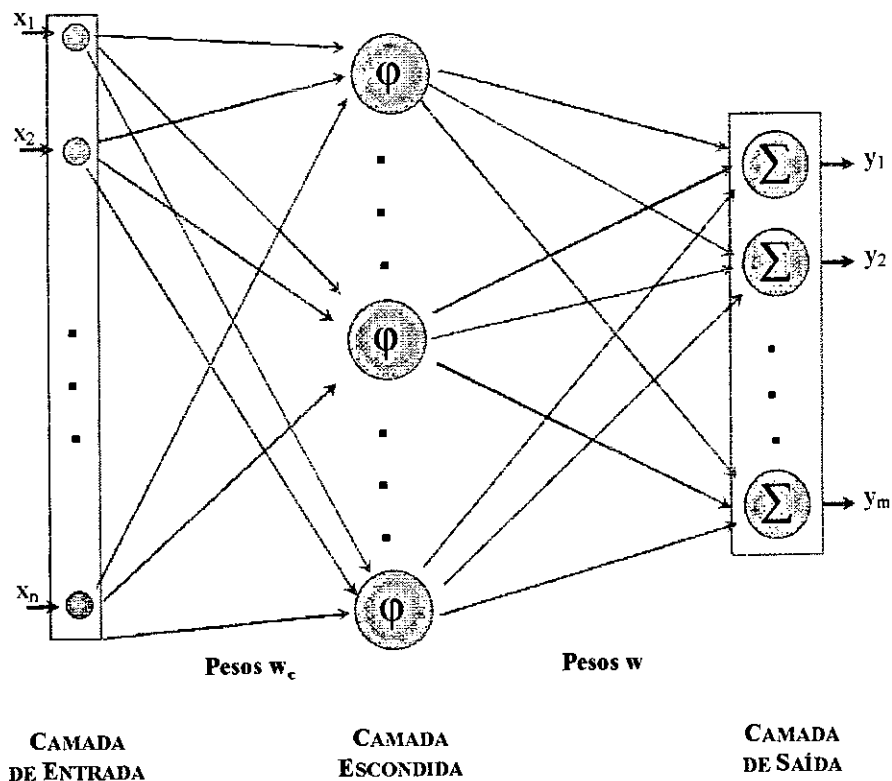


Figura 2.4 - Rede Neural com Função de Base Radial

2.3.2 - Metodologia Conforme a Teoria de Regularização

A introdução por Poggio e Girosi das Redes Neurais com Função de Base Radial com um caso particular de regularização de redes, foi elaborada com base na Teoria da regularização de Tikhonov. [10]

A idéia básica da teoria, apresentada resumidamente neste item, no contexto das redes neurais com Função de Base Radial, consiste em reduzir o problema da aproximação à minimização da Funcional de Tikhonov. [10]

2.3.2.1 - Construção da Funcional de Tikhonov

Supondo o conjunto de dados de entrada e saída, isto é, o conjunto de treinamento disponível para a aproximação sendo definido por $g = \{ (\bar{x}_i, d_i) \}$, onde:

\bar{x}_i é o vetor de entrada e pertence ao espaço R^m , isto é, tem dimensão m ;

d_i é a saída desejada, pertencente ao espaço R^1 , o que significa dizer que é unidimensional;

$i = 1, 2, \dots, N$, é a dimensão do conjunto de treinamento.

Ao considerar a saída como sendo unidimensional não limita a aplicabilidade geral da teoria da Regularização aqui desenvolvida.

A função de aproximação é denotada por $F(\bar{x})$, onde, por conveniência de apresentação, omite-se o vetor peso \bar{w} como argumento da função. De acordo com a teoria da Regularização de Tikhonov [10], a função F é obtida minimizando-se a funcional de custo $\xi(F)$, composta de dois termos:

a) O Termo de Erro

Este termo denotado com $\xi_s(F)$, mede o erro entre o valor desejado d_i e o valor y_i obtido pela rede, e pode ser definido na equação 2.13.

$$\xi_s(F) = \frac{1}{2} \sum_{i=1}^N (d_i - y_i)^2$$

ou

(2.13)

$$\xi_s(F) = \frac{1}{2} \sum_{i=1}^N [d_i - F(\bar{x}_i)]^2$$

b) O Termo de Regularização

Este termo denotado com $\xi_c(F)$, depende das propriedades geométricas da função de aproximação $F(\bar{x})$ e pode ser definido na equação 2.14.

$$\xi_c(F) = \frac{1}{2} \|D F\|^2 \quad (2.14)$$

onde:

D é um operador diferencial linear.

As informações a priori sobre a solução da Função $F(\bar{x})$ estão contidas no operador D , o que, naturalmente, faz da escolha de D um problema dependente.

D também pode ser identificado como um estabilizador, visto que ele estabiliza a solução do problema de Regularização, fazendo-a uniforme e, deste modo, satisfazendo a propriedade de continuidade. Embora, é importante que se ressalte, que uniformidade implique em continuidade, mas a recíproca não é verdadeira.

Desta forma a função de custo a ser minimizada pode ser apresentada como a equação 2.15.

$$\xi(F) = \xi_s(F) + \lambda \xi_c(F)$$

ou

(2.15)

$$\xi(F) = \frac{1}{2} \sum_{i=1}^N [d_i - F(\bar{x}_i)]^2 + \frac{1}{2} \lambda \|D F\|^2$$

onde:

λ é um número real e positivo, denominado parâmetro de Regularização;

$\xi(F)$ é a funcional de Tikhonov.[10]

Mínimizar a função de Tikhonov $\xi(F)$ significa encontrar a solução do problema de Regularização. O mínimo de $\xi(F)$ é denotado por $F \lambda(\bar{x})$. [10]

Pode-se ver o parâmetro de Regularização λ como o indicador de suficiência de um dado conjunto de dados. No caso limite em que $\lambda \rightarrow 0$ diz-se que o problema não possui nenhuma condição, desta forma, a solução de $F \lambda(\bar{x})$ é completamente determinada pelo conjunto de dados. No outro caso extremo, quando $\lambda \rightarrow \infty$ as condições de uniformidade impostas ao operador diferencial D são, por si só, suficientes para especificar a solução $F \lambda(\bar{x})$, o que, em outras palavras, é o mesmo que dizer que o conjunto de dados é irreal. Nos exemplos de aplicação prática, o valor de λ situa-se entre esse dois extremos, desta forma, tanto os dados como as considerações sobre a função, contribuem para a solução de $F \lambda(\bar{x})$.

2.3.2.2 - Solução do Problema da Regularização

Estabelece-se o princípio da Regularização com o seguinte:

↳ Achar a função $F_\lambda(\bar{x})$, que minimiza a funcional $\xi(F)$, definida pela equação 2.16

$$\xi(F) = \xi_s(F) + \lambda \xi_c(F) \quad (2.16)$$

onde:

$\xi_s(F)$ é o termo de erro;

ξ_c é o termo de Regularização;

λ parâmetro de Regularização.

O desenvolvimento matemático aplicado para minimizar o funcional de Tikhonov $\xi(F)$ [10] e, conseqüentemente, achar $F_\lambda(x)$, não faz parte do escopo deste estudo. Encontra-se maiores detalhes na referência [11].

Desta forma, minimizando-se $\xi(F)$ obtém-se a equação 2.17.

$$F_\lambda(x) = \frac{1}{\lambda} \sum_{i=1}^N [d_i - F(\bar{x}_i)] G(\bar{x}, \bar{x}_i) \quad (2.17)$$

onde:

$G(x, x_i)$ é a função de Green com centro em \bar{x}_i .

Pode-se observar pela equação 2.17 que a solução $F_\lambda(\bar{x})$ do problema de Regularização é uma superposição de N funções de Green, onde o vetor \bar{x}_i representa o centro de expansão e os pesos $[d_i - F(\bar{x}_i)] / \lambda$, representam os coeficientes de expansão.

Pode-se verificar também, que os coeficientes de expansão da equação $F_\lambda(x)$, isto é, os pesos $w [d_i - F(\bar{x}_i)] / \lambda$, são lineares na estimativa do erro definido pela diferença entre o valor desejado d_i e o correspondente valor saída da função $F(\bar{x}_i)$, e inversamente proporcional ao parâmetro de regularização λ .

2.3.2.3 - Determinação dos Coeficientes de Expansão

O próximo problema a resolver é determinar os coeficientes de expansão.

Como os termos $[d_i - F(\bar{x}_i)] / \lambda$ representam os coeficientes de expansão, ou seja, os pesos w , então pode-se escrever a equação 2.18.

$$w_i = \frac{1}{\lambda} [d_i - F(\bar{x}_i)], \quad i = 1, 2, \dots, N \quad (2.18)$$

Então, da equação 2.17 tem-se a equação 2.19.

$$F \lambda(\bar{x}) = \sum_{i=1}^N w_i G(\bar{x}, \bar{x}_i) \quad (2.19)$$

Calculando a equação 2.19 para x_j , tem-se a equação 2.20.

$$F \lambda(\bar{x}_j) = \sum_{i=1}^N w_i G(\bar{x}_j, \bar{x}_i) \quad (2.20)$$

onde:

$\bar{x}_j \in \mathbb{R}^m$ é o vetor de entrada;

$\bar{x}_i \in \mathbb{R}^m$ é o vetor centro;

$j = 1, 2, \dots, N$.

Escrevendo $F\lambda$, d , e G na forma matricial, tem-se as matrizes 2.21, 2.22 e 2.23.

$$F \lambda = [F \lambda(x_1), F \lambda(x_2), \dots, F \lambda(x_N)]^T \quad (2.21)$$

$$d = [d_1, d_2, \dots, d_N]^T \quad (2.22)$$

$$G = \begin{bmatrix} G(x_1, x_1) & G(x_1, x_2) & \dots & G(x_1, x_N) \\ G(x_2, x_1) & G(x_2, x_2) & \dots & G(x_2, x_N) \\ \vdots & \vdots & \ddots & \vdots \\ G(x_N, x_1) & G(x_N, x_2) & \dots & G(x_N, x_N) \end{bmatrix} \quad (2.23)$$

$$w = [w_1, w_2, \dots, w_N] \quad (2.24)$$

Escrevendo as equações 2.18 e 2.20 na forma matricial tem-se as equações 2.25 e 2.26.

$$w = \frac{1}{\lambda} (d - F\lambda) \quad (2.25)$$

e

$$F\lambda = G W \quad (2.26)$$

Rearrmando a equação 2.25, tem-se a equação 2.27.

$$d = (G + \lambda I)W \quad (2.27)$$

onde:

I é uma matriz identidade de mesma dimensão de G , isto é, $N \times N$.

A matriz G é positivamente definida para algumas classes de função, entre as quais está a Gaussiana e Multiquadrada Inversa. Pode-se também, na prática, fazer λ suficientemente grande, de tal forma que $(G + \lambda I)$ seja sempre positivamente definida, e, portanto, exista a sua inversa. [11]

Escreve-se então a equação 2.28.

$$W = (G + \lambda I)^{-1} d \quad (2.28)$$

O que significa que o sistema linear de equações dado em 2.23 tem apenas uma única solução.

Então, tendo solucionado o operador diferencial D e, conseqüentemente, identificada a função de Green (\bar{x}_j, \bar{x}_j) a ele associada, pode-se então usar a equação 2.28 para obter o vetor \bar{w} (dos pesos) para um específico vetor d e um apropriado parâmetro de regularização λ .

Estabelece-se que a solução do problema de Regularização é dada pela equação 2.29.

$$F \lambda (x) = \sum_{i=1}^N w_i G(x_1, x_i) \quad (2.29)$$

onde:

$G(x, x_i)$ são as funções de Green do operador diferencial linear D ;

w_i é o i ésimo elemento do vetor de peso \bar{w} .

A equação 2.29 estabelece o seguinte:

- \Rightarrow O método da Regularização é equivalente a expansão da solução em termos de um conjunto de Funções de Green. A caracterização destas funções dependem apenas da forma adotada para o estabilizador D e das condições de contorno a ele associadas;
- \Rightarrow O número de funções de Green usadas na expansão, é igual ao número de vetores utilizados no conjunto de treinamento.

2.3.2.4 - Considerações sobre as Funções de Green

A caracterização das Funções de Green, para um centro específico x_i , depende somente da forma do operador diferencial (estabilizador) D , isto é, das considerações feitas a priori com respeito a função que será construída para obter os dados de saída a partir dos de entrada.

Se o estabilizador for invariável à translação e a rotação, a função de Green $G(\bar{x}, \bar{x}_i)$, depende apenas da distância geométrica entre os vetores $\bar{x} - \bar{x}_i$, então tem-se a equação 2.30.

$$G(\bar{x}, \bar{x}_i) = G(\|\bar{x} - \bar{x}_i\|) \quad (2.30)$$

onde:

$\|\cdot\|$ denota a distância geométrica, que em geral é considerada Euclidiana.

Nestas condições, as funções de Green têm, obrigatoriamente, que ser funções de base radial e, neste caso, a solução da equação 2.30 toma a forma da equação 2.31.

$$F(x) = \sum_{i=1}^N w_i G(\|\bar{x} - \bar{x}_i\|) \quad (2.31)$$

A função de Green, cujo operador diferencial D é invariável à rotação e a translação, e que satisfaça a equação 2.31, é de grande interesse. Um exemplo de função de Green com estas propriedades é a função Gaussiana, definida pela equação 2.32.

$$G(\mathbf{x}, \mathbf{x}_i) = \exp \left[-\frac{1}{2\sigma_i^2} (\|\bar{\mathbf{x}} - \bar{\mathbf{x}}_i\|)^2 \right] \quad (2.32)$$

onde:

$\bar{\mathbf{x}}_i$ denota o centro da função;

σ_i a amplitude.

Com a função de Green definida conforme exposto anteriormente, a solução do problema da regularização terá a forma a seguir.

Substituindo a equação 2.32 na equação 2.31 tem-se a equação 2.33.

$$F\lambda(\mathbf{x}) = \sum_{i=1}^N w_i \exp \left[-\frac{1}{2\sigma_i^2} (\|\bar{\mathbf{x}} - \bar{\mathbf{x}}_i\|)^2 \right] \quad (2.33)$$

2.3.3 - Redes Regularizadas

Derivada da Teoria da Regularização pode-se verificar que a equação 2.29 sugere uma rede que, por razões óbvias, foram denominadas por Poggio e Girosi [5], de "Redes Regularizadas".

A expansão da função $F\lambda(\mathbf{x})$ da equação 2.29 em termos das funções de Green com centro em $\bar{\mathbf{x}}_i$ sugere uma rede com uma estrutura conforme mostrado na figura 2.4.

A rede constitui-se de 3 camadas:

- ⇒ Primeira Camada: Composta pelos nós de entrada, cujo número é igual a dimensão m do vetor de entrada $\bar{\mathbf{x}}_j$.
- ⇒ Segunda Camada: É a camada escondida. Composta por unidades não lineares, funções de Green, que estão diretamente conectadas a todos os nós da camada de entrada.

Verifica-se que há uma unidade escondida para cada vetor $\bar{\mathbf{x}}_i$, $i = 1, 2, \dots, N$, onde o N é a dimensão do conjunto de treinamento.

⇒ Camada de Saída: Consiste de uma única unidade linear, que se apresenta conectada à todas as unidades da camada escondida.

Ser uma unidade linear significa que a saída da rede é a soma dos pesos que saem da camada escondida.

Os pesos da camada escondida são os coeficientes de expansão definidos em termos das funções de Green e do parâmetro de regularização λ , conforme a equação 2.28.

Conforme premissa inicial, a figura 2.5 apresenta a arquitetura da rede com apenas uma unidade de saída. Conforme Haykin [11], evidentemente esta arquitetura pode ser estendida de forma a acomodar qualquer número de nós de saída desejados.

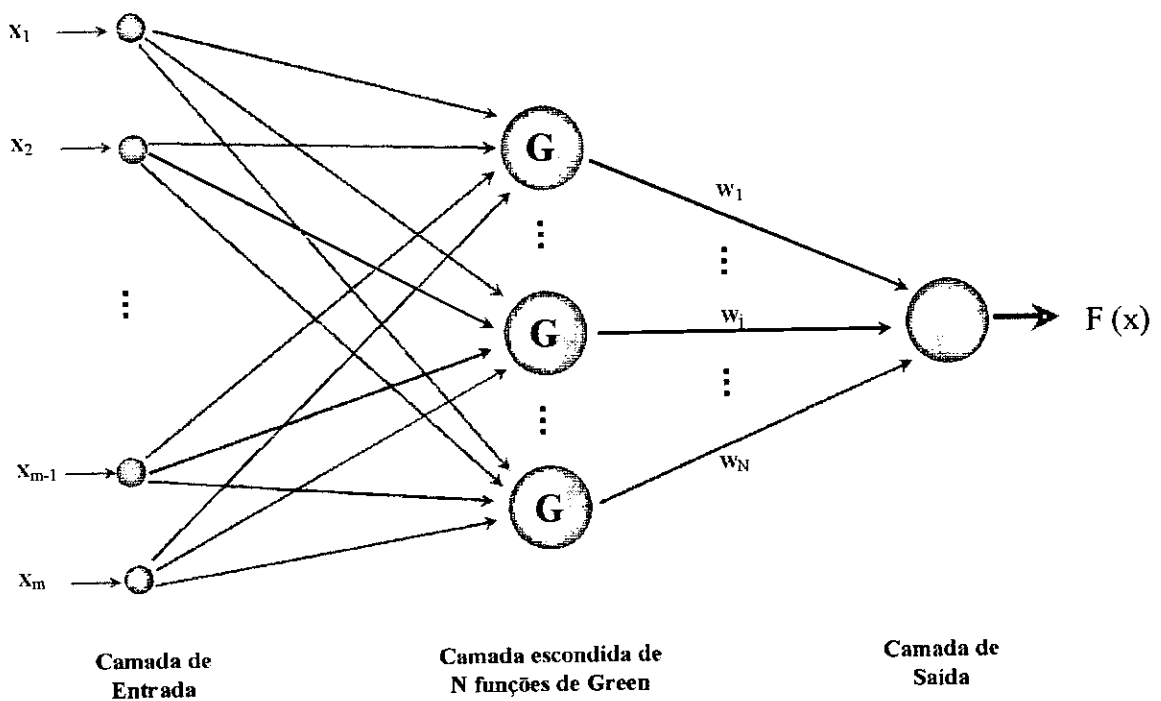


Figura 2.5 – Rede de Regularização

2.3.4 - Redes Generalizadas

Observa-se pela arquitetura das redes regularizadas, que na camada escondida há uma função para cada vetor \bar{x}_i do conjunto de treinamento. Entretanto não é necessário expandir as funções por todo conjunto de treinamento, pode-se considerar o caso mais geral, em que o número de nós M , da camada escondida é menor do que N , isto é, $M < N$. Neste caso tem-se redes menores e, computacionalmente, menos dispendiosas.

O método adotado envolve a busca de uma solução sub-ótima, que se aproxima da solução Regularizada dada pela equação 2.34, em um espaço de menor dimensão.⁹

$$F \lambda (\bar{x}) = \sum_{i=1}^N w_i G(\bar{x}, \bar{x}_i) \quad (2.34)$$

A solução aproximada é denominada de $F^*(\bar{x})$ e será expandida no espaço m_0 , tem-se então a equação 2.35.

$$F^*(\bar{x}) = \sum_{i=1}^{m_0} w_i \varphi_i(\bar{x}_i) \quad (2.35)$$

onde:

$\{\varphi_i(\bar{x}), i = 1, 2, \dots, m_0\}$ é o novo conjunto de funções de base radial;

w_i é o novo conjunto de pesos.

Considerando-se que as funções continuam sendo de base radial, pode-se escrever a equação 2.36.

$$\varphi_i(\bar{x}) = G(\|\bar{x} - \bar{t}_i\|), i = 1, 2, \dots, m_0 \quad (2.36)$$

onde:

$\{\bar{t}_i\}_{i=1}^{m_0}$ é o novo conjunto de centro a ser determinado.

Deve-se observar que a escolha de $\{\bar{t}_i\}_{i=1}^{m_0}$ e de m_0 são críticas, pois devem recuperar a solução para:

$$\bar{t}_i = \bar{x}_i \quad e \quad m_0 = N$$

Pode-se escrever a equação 2.35 na forma das equações 2.37 e 2.38.

$$F^*(\mathbf{x}) = \sum_{i=1}^{m_0} w_i G(\bar{\mathbf{x}}, \bar{t}_i) \quad (2.37)$$

ou

$$F^*(\mathbf{x}) = \sum_{i=1}^{m_0} w_i G(\|\bar{\mathbf{x}} - \bar{t}_i\|) \quad (2.38)$$

A nova função de custo será dada então pela equação 2.39.

$$\xi(F^*) = \sum_{i=1}^N \left(d_i - \sum_{j=1}^M w_j G(\|\bar{\mathbf{x}}_i - \bar{t}_j\|) \right)^2 + \lambda \|DF^*\|^2 \quad (2.39)$$

O problema agora é encontrar o novo conjunto de pesos $\{w_i\}_{i=1}^{m_0}$ que minimiza a equação 2.39.

O primeiro termo a direita, na equação 2.39, pode ser expresso na equação 2.40.

$$\|\mathbf{d} - G \mathbf{w}\|^2 \quad (2.40)$$

onde:

$$\mathbf{d} = [d_1, d_2, \dots, d_N]^T$$

$$G = \begin{bmatrix} G(\mathbf{x}_1, t_1) & G(\mathbf{x}_1, t_2) & \dots & G(\mathbf{x}_1, t_{m_0}) \\ G(\mathbf{x}_2, t_1) & G(\mathbf{x}_2, t_2) & \dots & G(\mathbf{x}_2, t_{m_0}) \\ \vdots & \vdots & \vdots & \vdots \\ G(\mathbf{x}_N, t_1) & G(\mathbf{x}_N, t_2) & \dots & G(\mathbf{x}_N, t_{m_0}) \end{bmatrix}$$

$$\mathbf{w} = [w_1, w_2, \dots, w_{m_0}]$$

Deve-se observar que o vetor coluna $\bar{\mathbf{d}}$ tem a dimensão anterior.

Entretanto, a matriz G e o vetor coluna $\bar{\mathbf{w}}$ têm agora dimensões diferentes. A matriz G passa a ter dimensão $m_0 \times N$ e o vetor $\bar{\mathbf{w}}$, dimensão m_0 .

Minimizando a Função de Custo $\xi (F^*)$ em relação ao vetor de pesos \bar{w} , tem-se então a equação 2.41.

$$(G^T G + \lambda G_0)W = G^T d$$

$$W = (G^T G + \lambda G_0)^{-1} G^T d \tag{2.41}$$

onde:

$$G_0 = \begin{bmatrix} G(t_1, t_1) & G(t_1, t_2) & \dots & G(t_1, t_M) \\ G(t_2, t_1) & G(t_2, t_2) & \dots & G(t_2, t_M) \\ \vdots & \vdots & \ddots & \vdots \\ G(t_M, t_1) & G(t_M, t_2) & \dots & G(t_M, t_M) \end{bmatrix}$$

Se o parâmetro λ aproximar-se de zero, tem-se a equação 2.42.

$$W = G^+ d \tag{2.42}$$

onde

$$G^+ = \left[(G^T G)^{-1} G^T \right] \text{ é a pseudo-inversa matriz de } G.$$

Esquemáticamente apresenta-se a rede na figura 2.6.

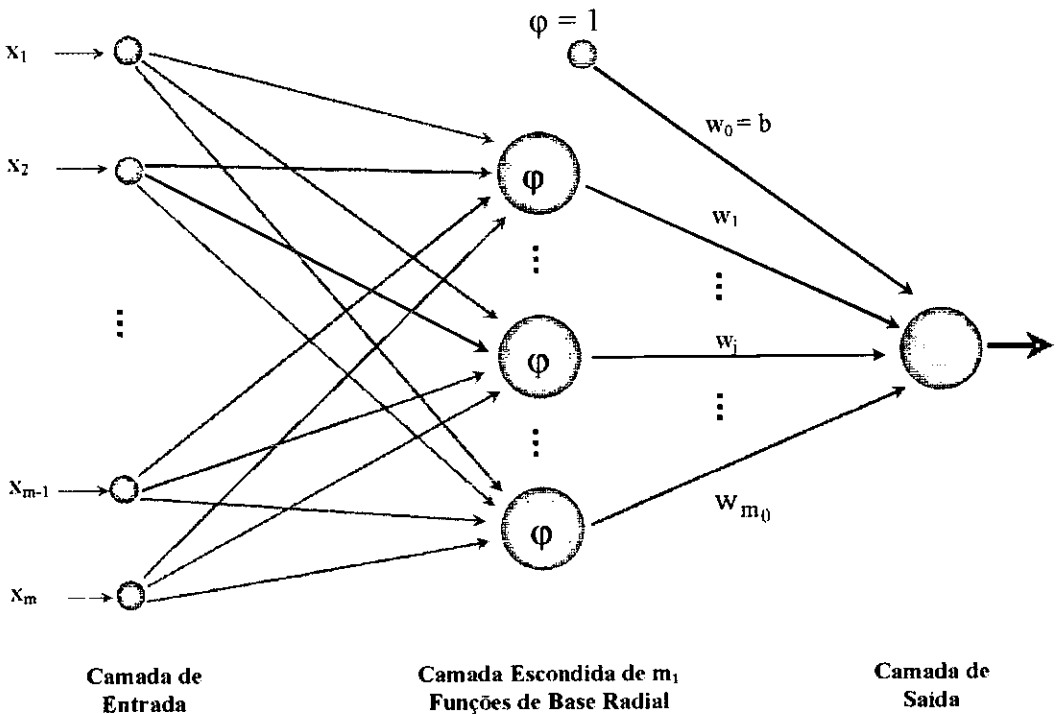


Figura 2.6 – Rede de Função de Base Radial

2.3.3 - Metodologia Segundo os Estimadores de Regressão de Núcleo

A arquitetura de Redes com Função de Base Radial apresenta uma forte similaridade com as técnicas de regressão, baseada na estimativa não paramétrica de uma função de densidade desconhecida. [11, 12]

A idéia básica dos Estimadores de Regressão de Núcleo é poder construir uma função aleatória desconhecida $f(\bar{x}) = \bar{y}$ estimando-se a função densidade de probabilidade $g(\bar{x}, \bar{y})$.

Conforme Blanzieri, a aproximação de f é dada pela equação 2.43.

$$\hat{f}(\bar{x}) = \frac{\sum_{i=1}^N y_i k_n [H_x^{-1}(\bar{x} - \bar{x}_i)]}{\sum_{i=1}^N k_n [H_x^{-1}(\bar{x} - \bar{x}_i)]} \quad (2.43)$$

Para o caso de uma variável, é denominado como Estimador de Nuradaya-Watson, e é dado pela equação 2.44.

$$N(x) = \frac{\sum_{i=1}^N w_i e_i [H_x^{-1}(\|\bar{x} - \bar{c}_i\|)]}{\sum_{i=1}^N e_i \|\bar{x} - \bar{c}_i\|} \quad (2.44)$$

Este pode ser facilmente comparado às RBFN's normalizadas, conforme inicialmente proposto por Moody e Darken. [9]

Visto que o objetivo neste item é apenas apresentar a convergência entre as duas metodologias, deixou-se de definir os parâmetros e variáveis das equações 2.43 e 2.44.

[13]

2.3.4 - Equivalência de Sistemas Fuzzy e Redes Neurais com Função de Base

Radial

Conforme mostrado por Jang [6], existe sob certas condições uma equivalência entre Sistemas Fuzzy e Redes Neurais com Funções de Base Radial. Para que a equivalência possa ser válida, as seguintes condições necessitam ser alcançadas:

- O número de unidades de ativação deve ser igual ao número de regras Fuzzy;
 - A saída, isto é, o conseqüente, de cada regra Fuzzy deve ser constante;
 - As funções associadas a cada regra devem ser funções Gaussianas com a mesma amplitude;
 - O operador usado para computar a ativação de cada regra é a multiplicação;
 - A rede com função de base radial e o sistema fuzzy, sob consideração, utilizam o mesmo método para obter sua saída, isto é, normalizada ou não normalizada.
- [6, 14]

2.4 - Algoritmos e Estratégias de Aprendizado para RBFN

As propriedades universais de aproximação, já apresentadas anteriormente, expressam a existência de uma solução ótima para o problema da aproximação. Achar os parâmetros ótimos da rede que minimizam a função de custo é o objetivo final do algoritmo de Aprendizado.

Para a definição completa de uma rede com Função de Base Radial é necessária a escolha dos seguintes parâmetros:

- ⇒ Número de funções de Base Radial;
- ⇒ A posição dos centros no espaço vetorial de entrada;
- ⇒ A amplitude das funções de Base Radial;
- ⇒ Os pesos que ligam a camada escondida, isto é, a camada de saída.

A escolha do número de funções Radiais é um ponto crítico na montagem de uma RBFN, dependendo do tipo de estratégia adotada, pode ser determinado a priori ou forma incremental.

2.4.1 - Estratégias de Aprendizado

Antes de descrever alguns algoritmos, destaca-se as diferentes estratégias de aprendizado que podem ser adotadas na construção de uma RBFN. Dependendo da forma de como os centros das funções de base radial da rede são especificadas, apresenta-se três estratégias de aprendizado.

As estratégias são:

a) Seleção Aleatória de Centros Fixos

Os centros são escolhidos aleatoriamente do conjunto de dados de treinamento. Conforme observado por Lowe [11], este método é considerado sensível e os dados devem estar uniformemente distribuídos no conjunto de treinamento. Considera-se a amplitude ou desvio padrão das funções de base radial como sendo a mesma para todas as unidades localizadas, e seu valor obtido pela equação 2.39.

$$\sigma = \frac{d_{\max}}{\sqrt{2 m_1}} \quad (2.45)$$

onde:

m_1 é o número de centros;

d_{\max} é a distância máxima entre os centros escolhidos.

Esta fórmula garante que as funções radiais não tenham um pico nem muito elevado nem muito baixo. Como uma alternativa à equação 2.45 pode-se também usar amplitudes maiores em áreas onde os dados possuem baixa densidade, o que evidentemente requererá experimentações com os dados do conjunto de treinamento.

Desta forma, o único parâmetro que necessita ser apurado neste método são os pesos da camada de saída da rede, aplicando, por exemplo, o procedimento da pseudo-inversa matriz, apresentado na equação 2.46.

$$W = G^+ d \quad (2.46)$$

onde:

d é o vetor resposta desejado;

G^+ é a pseudo-inversa matriz da matriz $G = \{g_{ij}\}$.

Sendo g_{ij} apresentado na equação 2.47.

$$g_{ij} = \exp\left[-\frac{m_1}{d^2} \|\bar{x}_j - \bar{t}_i\|^2\right], \text{ para } \{j\}_1^N \text{ e } \{i\}_1^{m_1} \quad (2.47)$$

onde:

\bar{x}_j é o j -ésimo vetor do conjunto de treinamento;

\bar{t}_i é o centro da i -ésima função Radial.

Este método necessita de um conjunto de treinamento muito grande para atingir um performance adequada.

b) Seleção dos Centros por Processo Não-Supervisionado

Um forma de superar as limitações da estratégia anterior de escolha aleatória dos centros é usar um processo de aprendizado híbrido, que consiste de duas etapas:

- i) Uma primeira etapa de aprendizado não supervisionado: o objetivo desta fase é estimar a localização dos centros e a amplitude das funções de base radial da camada escondida.
- ii) Uma segunda etapa de aprendizado supervisionado: este estágio conclui o projeto da rede estimando os pesos da camada de saída.

Para o estágio de aprendizado não supervisionado aplica-se um algoritmo de clusterização que irá particionar o conjunto de dados de treinamento em subconjuntos. Sendo que, cada subconjunto deverá ser o mais homogêneo possível. O algoritmo adequado a desenvolver tal função é o *Standard K-Means*, ou uma de suas versões. O *Adaptative K-Means* ou *Dynamic K-Means* com inicialização dinâmica, que situam os centros das funções radiais de base somente nas regiões do espaço de dados de entrada, onde haja uma significativa presença de dados.[11, 15]

Após ter identificado os centros das funções de base radial e suas respectivas amplitudes, a segunda etapa do processo híbrido de aprendizado é a estimativa dos pesos da camada de saída. Um método que pode ser usado nesta estimativa é o do gradiente descendente.

c) Seleção Supervisionada dos Centros

Neste método os centros das funções de base radial, bem como os demais parâmetros, passam por um processo de aprendizado supervisionado.

O primeiro passo no desenvolvimento deste processo de aprendizado é definir o valor instantâneo da função de custo da equação 2.48.

$$\xi = \frac{1}{2} \sum_{j=1}^N e_j^2 \quad (2.48)$$

onde:

N é o tamanho do conjunto de treinamento;

e_j é o indicativo de erro definido pela equação 2.49.

$$e_j = d_j - F^*(\bar{x}_j)$$

$$e_j = d_j - \sum_{i=1}^M W_i G(\|\bar{x}_j - \bar{t}_i\|) \quad (2.49)$$

O objetivo é achar os centros, os pesos e as amplitudes que minimizem a função de custo ξ .

O resultado das derivadas parciais que minimizam os parâmetros da rede são apresentados a seguir.

↪ Posição dos Centros na Camada Escondida (equações 2.50 e 2.51)

$$\frac{\partial \xi(\mathbf{x})}{\partial \mathbf{t}_1(\mathbf{x})} = 2\mathbf{w}_i \sum_{j=2}^N t_j(\mathbf{n}) G'(\|\bar{\mathbf{x}}_j - \bar{\mathbf{t}}_i(\mathbf{n})\| \mathbf{c}_i) \sum_i^{-1} [\mathbf{x}_j - \mathbf{t}_i(\mathbf{n})] \quad (2.50)$$

$$\mathbf{t}_i(\mathbf{n}+1) = \mathbf{t}_i(\mathbf{n}) - \eta_2 \frac{\partial \xi(\mathbf{x})}{\partial \mathbf{t}_1(\mathbf{x})} \quad \text{para } i = 1, 2, \dots, n_1 \quad (2.51)$$

↪ Pesos da Camada de Saída (equações 2.52 e 2.53)

$$\frac{\partial \xi(\mathbf{n})}{\partial \mathbf{w}_1(\mathbf{n})} = \sum_{j=1}^N e_j(\mathbf{n}) G(\|\bar{\mathbf{x}}_j - \bar{\mathbf{t}}_i\| \mathbf{c}_i) \quad (2.52)$$

$$\mathbf{w}_i(\mathbf{n}+1) = \mathbf{w}_i(\mathbf{n}) - \eta_1 \frac{\partial \xi(\mathbf{n})}{\partial \mathbf{w}_1(\mathbf{n})} \quad \text{para } i = 1, 2, \dots, n_1 \quad (2.53)$$

↪ Amplitude dos Centros da Camada Escondida (equações 2.54 e 2.55)

$$\frac{\partial \xi(\mathbf{n})}{\partial \sum_i^{-1}(\mathbf{n})} = -\mathbf{w}_i(\mathbf{n}) \sum_{j=1}^N e_j(\mathbf{n}) G'(\|\bar{\mathbf{x}}_j - \bar{\mathbf{t}}_i(\mathbf{n})\| \mathbf{c}_i) Q_{ji}(\mathbf{n}) \quad (2.54)$$

onde
$$Q_{ji}(\mathbf{n}) = [\bar{\mathbf{x}}_j - \bar{\mathbf{t}}_i(\mathbf{n})][\bar{\mathbf{x}}_j - \bar{\mathbf{t}}_i(\mathbf{n})]^T$$

$$\sum_i^{-1}(\mathbf{n}+1) = \sum_i^{-1}(\mathbf{n}) - \eta_3 \frac{\partial \xi(\mathbf{n})}{\partial \sum_i^{-1}(\mathbf{n})} \quad (2.55)$$

O termo $e_j(n)$ é o erro da unidade j no tempo n ; G' é a derivada primeira da função de Green $G(\cdot)$ em relação a seu argumento.[11]

2.4.2 - Algoritmos

Conforme Blanzieri [13], os algoritmos de aprendizado para Redes de Função de Base Radial podem ser classificados em Estáticos e Dinâmicos.

Algoritmos estáticos são aqueles que têm um número fixo de funções de base radial definida a priori e são, em geral, baseados em dois estágios de aprendizado, podendo executar cada um destes estágios de diferentes maneiras.

Os algoritmos dinâmicos permitem, durante o processo computacional, modificações no número de funções de base radial da rede, provocando, desta forma, mudanças em sua estrutura, possibilitando também, de forma incremental, a integração da ação de localização dos centros com a de definição dos pesos da camada de saída. O primeiro trabalho nesta direção foi *Resource Allocation Network* proposto por Platt. [16] Outros trabalhos que introduziram o aprendizado com mudanças estruturais na rede foram posteriormente apresentados por Fritzke e Blanzieri. [17, 18, 19, 20, 21, 22]

Apresentar-se-ão dois algoritmos estáticos . O primeiro utiliza o Algoritmo *Standard K-Means* para determinação dos centros e a pseudo-inversa matriz para determinação dos pesos. O segundo utiliza o algoritmo *Dynamic K-Means* para determinação dos centros e a Regra Delta para estimar os pesos, além da base teórica de um algoritmo dinâmico.

2.4.2.1 - Algoritmo Estático com a Pseudo-inversa matriz [23]

O algoritmo usa funções de base radial Gaussiana, e necessita que seja informado o número de funções de base radial, a forma como será computada a amplitude: através do *K-Means* ou do *Nearest Cluster* e, também, o fator de superposição.

Os parâmetros calculados pelo algoritmo são: a localização dos centros, a amplitude de cada função de base radial e os pesos. O treinamento é feito em dois estágios:

- Determinação da localização dos centros e da amplitude das funções utilizando o algoritmo *Standard K-Mean*;
- Ajuste dos pesos de cada unidade de ativação para a saída, usando a pseudo-inversão de matrizes.

Os passos do algoritmo são os seguintes:

Passo 1

Calcular a localização do centro e a amplitude das funções de base radial, utilizando o *K-Means*. Inicializar a amplitude de cada unidade de ativação usando o desvio padrão calculado pelo algoritmo *K-Means*, ou a distância entre cada cluster do algoritmo *Nearest Cluster*.

Passo 2

Resolver pelo método da Regularização, o problema linear dos mínimos quadrados, calculando os pesos das funções de base radial para cada saída.

Para uma dada saída de erro a ser minimizado, calcula-se a equação 2.56.

$$E = \sum_{j=1}^N (\bar{d}_j - \bar{y}_j)^2 \quad (2.56)$$

onde:

\bar{d}_j é a saída desejada, para o j -ésimo vetor de entrada;

\bar{y}_j é a saída gerada pela rede para o j -ésimo vetor de entrada.

A solução para o conjunto ótimo de pesos, desta saída, é dada pela equação 2.57.

$$W = (G^T G)^{-1} (G^T D) \quad (2.57)$$

onde:

W é um vetor coluna contendo os pesos, incluindo o bias, como apresentando em 2.58.

$$G = \begin{bmatrix} 1 & G_1 \\ 1 & G_2 \\ \vdots & \vdots \\ 1 & G_N \end{bmatrix} \quad D = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} \quad (2.58)$$

onde:

G é um vetor linha de saída das funções Gaussianas para o vetor de entrada \bar{x}_i ;

$$G_i = [g_1(x_i), g_2(x_i), \dots, g_N(x_i)];$$

d_i é a saída desejada para a entrada i .

⇒ Algoritmo *K-Means Clustering*

O Algoritmo *K-Means Clustering*, usado para determinar os centros, apresenta os seguintes passos:

- 1- Inicializar o número de clusters definido pelo usuário, selecionando-os aleatoriamente do conjunto de treinamento;
- 2- Classificação do conjunto de treinamento: Para cada vetor \bar{x}_i do conjunto de treinamento ache o centro c^* do cluster mais próximo e classifique \bar{x}_i como um membro de c^* ;
- 3- Para cada cluster, recomputar o seu centro achando a média do cluster, utilizando a equação 2.59;

$$M_k = \frac{1}{N_k} \sum_{j=1}^{N_k} \bar{x}_{jk} \quad (2.59)$$

onde:

M_k é a nova média;

N_k é o número de vetores no cluster k ;

\bar{x}_{jk} é o j -ésimo vetor pertencente ao cluster k .

4- Se o número de centros de cluster é menor do que o número especificado, dividir cada cluster em dois, achando o vetor com o maior desvio padrão, através da equação 2.60.

$$\sigma_i = \sum_{j=1}^{N_k} (\bar{x}_{ij} - \bar{M}_{ij})^2 \quad (2.60)$$

onde:

\bar{x}_{ij} é a i -ésima dimensão do j -ésimo vetor no cluster k ;

\bar{M}_{ij} é a i -ésima dimensão do centro do cluster;

N_k é o número de vetores no cluster k .

Se a divisão de todos os clusters exceder o número destes desejado, escolhe-se, aleatoriamente, um subconjunto de clusters para dividir.

Voltar ao passo 2 e manter o loop até que a mudança na média do cluster seja menor do que a especificada pelo usuário.

Senão, armazenar os centros do cluster.

2.4.2.2 - Algoritmo Estático com regra Delta [15]

Da mesma forma que o algoritmo apresentado anteriormente, este também é executado em duas etapas.

- Uma fase não supervisionada para determinação dos centros utilizando o *Dynamic K-Means Algorithm* e a *Nearest Neighbor Heuristic* para encontrar a amplitude.
- Uma fase supervisionada, utilizando a Regra Delta, para determinar os pesos da camada de saída.

a) ETAPA NÃO SUPERVISIONADA

O desenvolvimento do *Adaptative K-Means* e *Dynamic K-Means Algorithm* é apresentado na seqüência.

Adaptative K-Means

Passo 1 - Inicialização

Iniciar com um número aleatório de centros $\bar{w}_{c_1}, \bar{w}_{c_2}, \dots, \bar{w}_{c_k}$.

Passo 2 - Interação

- Ler o próximo vetor de entrada \bar{x}_i ;
- Modificar somente o centro do cluster mais próximo de $x^{(i)}$ conforme a equação 2.61:

$$\bar{w}_{c_k}^{(\text{novo})} = \bar{w}_{c_k}^{(\text{antigo})} + \alpha [\bar{x}_i - \bar{w}_{c_k}^{(\text{antigo})}] \quad (2.61)$$

onde:

α é a taxa de aprendizado que decresce conforme i cresce.

- Terminar após um número fixo de interações ou quando a taxa de aprendizado cair a zero.

A inicialização dos centros do *Adaptative K-Means* pode ser feita dinamicamente, de forma a impedir que alguns centros importantes deixem de ser utilizados. O algoritmo passaria a ter os seguintes passos:

Passo 1 - Inicialização

Desabilitar todos os centros de cluster $\bar{w}_{c_1}, \bar{w}_{c_2}, \dots, \bar{w}_{c_k}$.

Passo 2 - Interação

- Ler o próximo vetor de entrada \bar{x}_i ;
- Se o centro do cluster \bar{w}_{c_k} mais próximo estiver a uma distância R de \bar{x}_i , ou se todos os centros de cluster já estiverem habilitados, modificar \bar{w}_{c_k} conforme proposto pelo *Adaptive K-Means*, utilizando a fórmula 2.61;
- Senão, habilite um novo centro de cluster e o inicialize com o vetor \bar{x}_i ;
- Terminar após um número fixo de interações ou quando a taxa de aprendizado estar em zero.

Após a definição da localização dos centros do cluster, o passo seguinte é determinar a amplitude das funções de base radial.

O valor de σ_c é obtido através do uso do *Nearest Neighbor Heuristic* entre os P centros cluster mais próximos.

Dado os centros de cluster $\{\bar{w}_{c_i}\}_{i=1}^P$, a amplitude é dada pela equação 2.62.

$$\sigma_i = \sqrt{\frac{1}{P} \sum_{p=1}^P \left(\|\bar{w}_{c_k} - \bar{w}_{c_i}\| \right)^2} \quad (2.62)$$

Este é a raiz quadrada do valor médio quadrático das distâncias entre um dado centro c_i e os seus P vizinhos mais próximos.

b) ETAPA SUPERVISIONADA

Após a conclusão da fase de clusterização e determinação da amplitude dos campos de recepção, a camada de saída que, adaptativamente, minimiza o erro médio quadrático na equação 2.63.

$$\varepsilon = \frac{1}{2} \sum (\bar{d}_i - \bar{y}_i)^2 \quad (2.63)$$

onde:

\bar{d}_i é a saída desejada para o vetor de entrada \bar{x}_i ;

\bar{y}_i é o valor obtido pela rede.

2.4.2.3 - Algoritmo Dinâmico

Na estratégia de treinamento dos algoritmos dinâmicos a fase supervisionada e a não supervisionada são executadas paralelamente, o oposto dos algoritmos estáticos, em que as duas fases são executadas em duas etapas distintas.

O erro entre o valor desejado e o obtido pela rede é utilizado para definir onde deve-se inserir um nova célula.

A alocação incremental de unidades localizadas foi inicialmente proposta por Platt, que utilizou este modelo em problemas de regressão e em predição de séries temporais.

O método proposto insere uma nova unidade quando um vetor de entrada situar-se distante dos centros já existentes, ou causar um erro maior que a média.

Fritzke observou que este modelo diante de uma base de dados com ruído fica suscetível a *over-fitting*, pois para cada entrada situada longe dos centros, uma nova célula é criada.

No trabalho "*Fast Learning with Incremental RBF Networks*" [17] apresentou-se um modelo similar ao *Resourcing Allocating Network* proposto por Platt [16], mas que busca superar os problemas mencionados.

Apresenta-se a seguir a base teórica do algoritmo Growing Neural GAS. [17, 19]

a) Descrição Geral do Modelo

A rede constitui-se, inicialmente, de duas unidades localizadas posicionadas aleatoriamente no espaço vetorial de entrada.

As unidades são ligadas entre si por uma conexão denominada "conexão de vizinhança" ou "aresta", que apenas indica que as unidades são próximas. A cada unidade localizada é associada uma função Gaussiana.

A amplitude de cada função Gaussiana é obtida pela média do comprimento de todas as arestas que partem dela.

O treinamento é realizado apresentando pares dos vetores de entrada e de saída. Cada vetor de entrada ativa todas as unidades Gaussianas com certo grau de intensidade. As ativações são propagadas através dos pesos para as unidades de saída, que somam todos os sinais de entrada. A comparação entre o valor obtido pela rede e o valor de saída desejado expressa o erro para aquele vetor.

b) Adaptação dos Centros

A cada vetor de entrada que é apresentado a unidade localizada mais próxima - *best matching unit* (bmu) - é movida em sua direção por uma fração constante ϵ_0 da distancia entre eles. As unidades localizadas que tem aresta com a "bmu" também são movidas por uma pequena fração ϵ_1 , na direção do vetor de entrada. O objetivo desta adaptação é dar às unidades a tendência a seguir aos vetores de entrada.

c) Treinamento dos Pesos

Os pesos que conectam a camada de unidades localizada à camada de saída são treinados pela Regra Delta. Para cada par de vetor de entrada-vetor de saída, que é apresentado, executa-se um passo de adaptação pela regra delta.

d) Topologia da Vizinhaça

Considerando-se que a amplitude das unidades localizadas é definida com base nas arestas que partem dela e a adaptação dos centros causam um deslocamento das unidades, as informações sobre a proximidade das unidades localizadas necessitam ser constantemente atualizadas.

A metodologia aplicada, para tanto, é basicamente a proposta pelo Algoritmo Neural GAS. [24]

Para cada sinal de entrada conecta-se a *best matching unit* e a segunda *best matching unit* por uma aresta. Caso a conexão já exista, coloca-se a idade desta variável como sendo igual a zero. Uma variável idade é associada a cada aresta e determina-se que ela seja igual a zero quando ela é criada. A cada passo de adaptação, logo após a

possível inserção de uma nova aresta, a idade de todas as arestas que saem da "bmu" é aumentada. Arestas cuja idade excedem a determinado valor a_{max} são removidas e, desta forma, surgem nos que não têm mais arestas partindo dele.

O procedimento descrito permite que haja uma continua atualização sobre as informações de vizinhança na rede.

e) Acumulação do Erro

A cada passo de adaptação o erro quadrático na unidade de saída é atribuído a "bmu" e adicionado a uma variável de erro local. O objetivo deste procedimento é identificar unidades em regiões do espaço de entrada onde estejam ocorrendo muitas falhas de classificação pode-se assumir que estas células têm padrões de treinamento de diferentes classes em sua vizinhança, pois de outra forma a regra delta ajustaria os pesos adequadamente, impedindo que o erro acumulado crescesse e adquirisse valores elevados. Consequentemente, faz sentido usar a informação sobre o erro acumulado para decidir a localização de uma nova unidade.

f) Inserção de Novas Unidades

A inserção de uma nova unidade poderia se dar de forma compassada, como por exemplo, a cada λ sinais de entrada.

Um outro método que pode ser mais interessante, é observar o erro médio por padrão ou erro num conjunto de validação independente. Se para uma dada rede, o erro não mais decresce, é porque a adaptação dos centros e a regra delta são incapazes de, por si só, melhorar o desempenho da rede. Então, determina-se a célula f que tenha o maior erro acumulado.

Insere-se uma nova unidade entre f e um dos seus vizinhos, como, por exemplo, a unidade q . a posição da nova unidade no espaço vetorial de entrada é a média aritmética das posições f e q .

A aresta original entre f e q é substituída por duas arestas que saem da nova unidade, uma para f e outra para q . após a inserção da nova unidade, o erro nas variáveis de f e q é reduzido em cinquenta por cento e o erro na nova célula tem um valor inicial igual a média da redução dos erros de f e q .

O valor da redução, bem como o de inicialização, são fixados heurísticamente, dado que não existe uma justificativa teórica para eles. O propósito da redução é impedir que um erro elevado em uma dada unidade acarrete noutra inserção no mesmo local onde já havia sido realizada.

Para enfatizar a importância dos erros ocorridos recentemente, todas as variáveis de erro têm ligeiro decréscimo, aplicando-se a estas um fator α ($0 < \alpha < 1$) a cada passo de adaptação.

g) Critério de Parada

O processo de crescimento descrito necessita ser interrompido para impedir a continua inserção de unidades.

Uma possibilidade é definir, a priori, o número máximo de unidades, como no caso de redes não incrementais, visto ser uma das variáveis do problema o número de unidades.

Outra possibilidade, muito adequada à redes incrementais, é definir um erro máximo para um conjunto de validação ou de treinamento. Treinamento e inserção prosseguem até alcançar um dado critério escolhido.

Para um melhor acompanhamento da evolução do algoritmo mostra-se a figura 2.7.

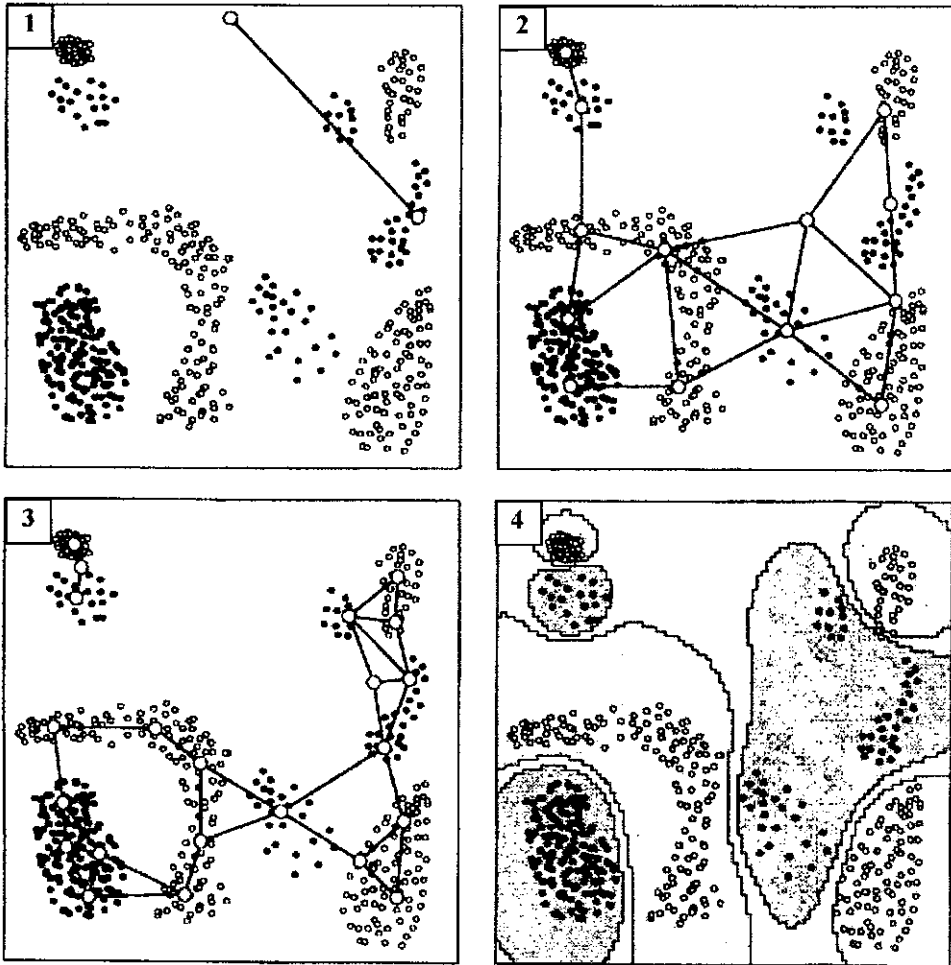


Figura 2.7 – Solução do problema de classificação para o Algoritmo Growing Neural GAS [17]

Na figura 2.7 o conjunto de treinamento é constituído de 591 vetores, e possui duas classes: a dos círculos preenchidos e a dos vazados.

O quadro 1 mostra a rede inicial com apenas duas unidades. O quadro 2 apresenta a rede após 4.800 passos de adaptação, o quadro 3 após 8.100 passos. O ultimo quadro, mostra a generalização da rede, onde cada ponto foi classificado como pertencente a uma das classes.

Os parâmetros adotados nesta simulação foram: $\epsilon_0 = 0,2$; $\epsilon_1 = 0,006$; $\alpha = 0,999$ e $a_{\max} = 88$.

CAPÍTULO 3 - CARACTERIZAÇÃO DO PROBLEMA

Neste capítulo. Redes Neurais com Função de Base Radial são aplicadas a três bases de dados distintas, afim de avaliar a performance deste método em tarefas de classificação e de predição de dados.

3.1 - Bases de Dados Utilizadas

No estudo de classificação são utilizadas 2 bases de dados:

- i) Plantas Iris - criada por R.A. Fisher e citada pela primeira vez em 1936, no artigo "*The use of Multiple Measurements in Taxonomic Problems*" [25];
- ii) Aprovação de Crédito - esta base está citada no artigo "*Simpling Decision Trees*" [26].

No estudo de predição de uma função contínua utilizar-se-á a base fornecida para o problema intitulado "Monitoramento de Condições baseado na Predição de Sinais de Vibração" [27] proposto pelo ERUDIT - *European Network in Uncertainly Techniques Developments for use in Information Technology* -. A base de dados foi obtida através de medições das condições de operação e dos sinais de vibração de uma turbina.

3.1.1 - Base de Dados "Plantas Iris "

A base de dados Iris é, certamente, a mais conhecida e mais citada na literatura de reconhecimento de padrões.

O conjunto possui 3 classes e, cada uma destas, possui 50 instâncias. Uma classe é linearmente separável das outras duas, estas porém, não são linearmente separáveis das demais.

A base de dados possui 4 atributos de entrada:

- Comprimento da sépala em cm;
- Largura da sépala em cm;
- Comprimento da pétala em cm;
- Largura da pétala em cm.

O atributo de saída é apenas 1, representando o tipo de flor, que pode ser:

- Iris Setosa;
- Iris Versicolor;
- Iris Virgínica.

3.1.2 - Base de Dados "Aprovação de Crédito "

Esta base de dados está relacionada ao uso de cartões de crédito. O nome dos atributos e os valores destes foram criptografados, de forma a ocultar os dados confidenciais.

A base possui 8 atributos e 6 atributos nominais. Os nominais são de dois tipos diferentes, os que possuem poucas categorias e os que possuem várias categorias.

O conjunto constitui-se de 690 instâncias e 15 atributos, destes, 14 são de entrada e 1 de saída.

A base de dados possui os seguintes atributos de entrada:

- A1: 0, 1 nominal
- A2: contínuo
- A3: contínuo
- A4: $\left\{ \begin{array}{l} 1, 2, 3 \quad (\text{p, g, gg}) \quad \text{nominal} \\ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14 \quad \text{nominal} \end{array} \right.$
- A5: $\left\{ \begin{array}{l} (\text{tt, d, i, k, j, aa, m, c, w, e, q, n, cc, x} \\ 1, 2, 3, 4, 5, 6, 7, 8, 9 \quad \text{nominal} \end{array} \right.$
- A6: $(\text{tf, dd, j, bb, v, n, o, n, z})$
- A7: contínuo
- A8: 1, 0 (t, f) nominal
- A9: 1, 0 (t, f) nominal
- A10: contínuo
- A11: 1, 0 (t, f) nominal
- A12: 1, 2, 3 (s, p, g)
- A13: contínuo
- A14: contínuo

O atributo de saída é apenas 1:

- A15: (0, 1) (+, -)

3.1.3 - Base de Dados "Predição de Sinais de Vibração "

A base de dados do Problema de Monitoração de Condições Baseada na Predição de Sinais de Vibração foi obtida medindo-se 2.900 dados durante um período de tempo, que cobriu todos os modos de operação relevantes de uma turbina.

Cada vetor contém 23 atributos (variáveis). Os 11 atributos iniciais são parâmetros de operação e, provavelmente, são relevantes ao propósito de monitoração das vibrações. Os 12 atributos subsequentes são parâmetros de vibração e foram medidos em diferentes pontos da turbina.

As variáveis de operação e os pontos de medição de vibração na turbina não foram identificados. Não há também informação da existência ou não, de valores omitidos nas variáveis de operação e nos pontos de medição de vibração.

CAPÍTULO 4 - ESTUDO DE CASO

4.1 - Solução dos Problemas de Classificação e Predição

O objetivo a ser atingido com as bases de dados apresentadas no capítulo 4 é:

- Para a base de dados "Iris", a classificação da planta como Iris Setosa, Versicolor ou Virgínica;
- Para a base de dados "Aprovação de Crédito", a classificação dos créditos como aprovado ou não-aprovado;
- Para a base de dados "Monitoração das Condições de Operação" do ERUDIT, a predição de sinais de vibração em diferentes pontos de uma turbina.

Na solução destes 3 problemas, utilizar-se-á dois softwares diferentes de Rede Neural com Função de Base Radial. [15, 23]

O primeiro utiliza o algoritmo "*Standard K-Means*" para clusterizar os dados e a Pseudo-Inversa Matriz para calcular os pesos. O estudo das bases de dados com este software é designado, no presente trabalho, como CASO 1. [23]

O segundo utiliza o algoritmo "*Dynamic K-Means*" para clusterizar os dados e a Regra Delta para estimar os pesos. O estudo das bases de dados com este software é designado como CASO 2. [15]

4.2 - Parâmetros Utilizados pelos Softwares

Os softwares utilizados na solução dos problemas apresentados empregam algoritmos de clusterização da base de dados e de estimativa dos pesos, distintos um do outro. Portanto, os parâmetros para montagem das redes em cada software são consideravelmente diferentes.

4.2.1 - Parâmetros para o CASO 1

Os seguintes parâmetros são necessários para a configuração de uma Rede:

⇒ Número de Entrada

Define o número de variáveis da base de dados.

⇒ Número de Saídas

Define o número de variáveis de saída da base de dados.

⇒ Número de Funções de Base Radial

Este valor além de especificar o número de funções de base radial, define o número de centros cluster utilizado pelo algoritmo *K-Means* para representar o conjunto de treinamento.

⇒ Amplitude das Funções de Base Radial

Pode-se escolher entre dois métodos para determinar a amplitude inicial das funções de base radial.

- N Clusters mais próximos

Define-se o número de cluster mais próximo afim de obter as amplitudes iniciais das funções de base radial.

- Amplitude do Cluster

As amplitudes iniciais das funções de base são obtidas a partir das amplitudes dos clusters.

⇒ Fator de Superposição

Este valor controla o grau de superposição entre as funções de base radial. Valores muito elevados do fator de superposição causam uma pobre discriminação entre os padrões de entrada. Valores muito pequenos causam uma baixa capacidade de generalização da rede.

Os parâmetros a seguir são do algoritmo de clusterização.

⇒ Método de Clusterização

Define-se (nesta área) o número inicial de clusters e o critério de divisão destes. Um cluster pode ser dividido por:

⇒ Número Inicial de Clusters

Define o número de centros de cluster em que a base de dados será inicialmente dividida.

- Épocas

Define o número máximo de épocas para dividir um cluster.

- % de Mudança no Desvio Padrão

Especifica a quantidade mínima de mudança no desvio padrão de qualquer cluster, para prosseguir com o movimento de divisão destes. Se todos os desvios padrão de todos os cluster for menor do que esse valor, então o movimento de divisão é interrompido.

⇒ Condições de Interrupção

O algoritmo *K-Means* roda até que o número especificado de centros seja alcançado. As condições de interrupção especificam quando parar o movimento de divisão no conjunto final de centros de cluster e interromper o treinamento.

As condições de interrupção podem ser dadas por:

- Épocas

Define o número de épocas a rodar, antes de dividir os clusters existentes.

- % de Mudança no Desvio Padrão

Define a porcentagem mínima de mudanças do desvio padrão antes de dividir os clusters existentes.

4.2.2 - Parâmetros para o CASO 2

Os seguintes parâmetros são necessários para a configuração desta Rede com Função de Base Radial:

⇒ Número de Entradas

Define o número de variáveis de entrada da base de dados.

⇒ Número de Saídas

Define o número de variáveis de saída da base de dados. Nos casos de classificação, cada classe deve ser tratada como uma variável de saída.

⇒ Número de Funções de Base Radial

Este valor especifica o número de funções de base radial na camada escondida da rede.

⇒ Regra de Aprendizado

Define o tipo de regra de aprendizado a ser adotada para estimar os pesos da camada de saída.

⇒ Função de Transferência

Define a função de transferência para a camada de saída.

4.3 - Parâmetros de Configuração das Redes

A solução dos problemas de classificação e predição propostos foram alcançados construindo-se as Redes Neurais com Função de Base Radial a partir dos 2 casos estudados.

Apresenta-se a seguir os parâmetros adotados para a configuração destas redes.

4.3.1 - Redes para a Base de Dados Iris

A- CASO 1

- Número de entradas: 4;
- Número de Saídas: 1;
- Número de Funções de Base Radial: 32;
- Para obtenção da amplitude das funções de base radial escolheu-se, em todos os casos, o método do número de cluster mais próximo;
- Número de Cluster mais Próximo: 4;
- Fator de Superposição: 0,8;
- Método de Clusterização:
 - ⇒ Número Inicial de Clusters: 8;
 - ⇒ % de Mudança no Desvio Padrão: 0,001;
- Condições de Interrupção:
 - ⇒ % de Mudança no Desvio Padrão: 0,0005.

B- CASO 2

- Número de entradas: 4;
- Número de Saídas: 3;

- Número de Funções de Base Radial: 35;
- Regra de Aprendizado: Regra Delta Extendida [15];
- Função de Transferência da Camada de Saída: tangente hiperbólica.

4.3.2 - Redes para Base de Dados "Aprovação de Crédito"

A- CASO 1

- Número de entradas: 14;
- Número de Saídas: 1;
- Número de Funções de Base Radial: 30;
- Número de Cluster mais Próximo: 4;
- Fator de Superposição: 0,6;
- Método de Clusterização:
 - ⇒ Número Inicial de Clusters: 8;
 - ⇒ % de Mudança no Desvio Padrão: 0,0001;
- Condições de Interrupção:
 - ⇒ % de Mudança no Desvio Padrão: 5×10^{-5} .

B- CASO 2

- Número de entradas: 14;
- Número de Saídas: 2;
- Número de Funções de Base Radial: 50;
- Regra de Aprendizado: Regra Delta Extendida [15];
- Função de Transferência da Camada de Saída: sigmóide.

4.3.3 - Rede para a Base de Dados "Predição de Sinais de Vibração"

A- CASO 1

- Número de entradas: 11;
- Número de Saídas: 12;
- Número de Funções de Base Radial: 198;
- Número de Cluster mais Próximo: 4;
- Fator de Superposição: 0,5;
- Método de Clusterização:
 - ⇒ Número Inicial de Clusters: 30;
 - ⇒ % de Mudança no Desvio Padrão: 1×10^{-5} ;
- Condições de Interrupção:
 - ⇒ % de Mudança no Desvio Padrão: 5×10^{-6} ;

B- CASO 2

- Número de entradas: 11;
- Número de Saídas: 12;
- Número de Funções de Base Radial: 52;
- Regra de Aprendizado: Regra Delta Extendida [15];
- Função de Transferência da Camada de Saída: sigmóide.

CAPÍTULO 5 - ANÁLISE DOS RESULTADOS

5.1 - Desempenho das RBFN's

A habilidade das Redes Neurais com Função de Base Radial em resolver diferentes tipos de problemas, aliada a sua capacidade em alcançar excelentes resultados com pouco tempo de treinamento, é, certamente, uma das vantagens deste modelo de rede e, um dos motivos, do grande interesse que vêm despertando.

Os resultados apresentados nos itens 5.2 a 5.6 do presente trabalho, foram obtidos treinando-se as bases de dados "*IRIS*", "*Aprovação de Crédito*" e "*Predição de Sinais de Vibração*" com Redes de Função de Base Radial, que utilizam Algoritmos Estáticos. [25, 26, 28]

Considera-se que o modelo em conjunto com as configuração adotadas, para cada base, atingiu um bom desempenho.

A aplicação de algoritmos com alocação dinâmica das funções de base radial está fora do escopo deste trabalho. Sua utilização, entretanto, certamente permitiria atingir resultados mais precisos.

5.2 - Avaliação da Precisão dos Resultados

A avaliação da performance da rede é medida de duas formas distintas.

Para os problemas de classificação, a qualidade dos resultados, é balizada pela classificação percentual de padrões corretos ou incorretos, definido no item 5.2.1.

Para o problema de regressão de uma função contínua, a qualidade da rede é avaliada pela raiz quadrada do erro médio quadrático, definido no item 5.2.2.

Apresenta-se, também, no item 5.2.3 o fator de correlação.

5.2.1 – Classificação Percentual dos Padrões

A medida de padrões classificados corretamente será dado pela equação 5.1.

$$\% \text{ Pad_Corret} = \left(1 - \frac{\text{NPad_Incorr}}{\text{NTot_Pad}} \right) \times 100 \quad (5.1)$$

onde:

% Pad_Corret - porcentagem de padrões classificados corretamente;

NPad_Incorr - número de padrões classificados incorretamente;

NTot_Pad – número total de padrões do conjunto.

A medida de padrões classificados incorretamente é dada pela equação 5.2.

$$\% \text{ Pad_Incorr} = \left(1 - \frac{\text{NPad_Corret}}{\text{NTot_Pad}} \right) \times 100 \quad (5.2)$$

onde:

% Pad_Incorr - porcentagem de padrões classificados incorretamente;

NPad_Corret - número de padrões classificados incorretamente;

NTot_Pad – número total de padrões do conjunto.

5.2.2 – Raiz do Erro Médio Quadrático – RMSE

O RMSE é a raiz do erro médio quadrático entre o valor desejado e o valor predito, ou estimado, pela rede, dado pela equação 5.3

$$\text{RMSE} = \sqrt{\sum_{j=1}^n \sum_{i=1}^m \frac{(d_i - y_i)^2}{m \cdot n}} \quad (5.3)$$

onde:

m – número de saídas;

n – número de padrões no conjunto considerado;

d_i – valor de saída desejado;

y_i – valor de saída predito pela rede.

5.2.3 – Fator de Correlação

O fator de correlação mede o grau de ajuste dos valores estimados em relação aos valores desejados. O fator de correlação varia de 0 a 1, um ajuste perfeito tem correlação igual a 1.

O fator de correlação é dado pela equação 5.4.

$$\text{FCORR}^2 = \left[\frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \right]^2 \quad (5.4)$$

onde:

FCORR^2 – fator de correlação;

x – valor desejado;

y – valor estimado;

5.3 - Apresentação dos Resultados da Base de Dados "IRIS"

5.3.1 - Tabela com os Resultados

Apresenta-se na tabela 5.1 o Erro Médio Quadrático (RMSE), a percentagem de padrões classificados correta e incorretamente para o CASO 1 e CASO 2.

ALGORITMO	RMSE	% PAD-INCORR	% PAD-CORRET
CASO 1	0,006236	1,90	98,10
CASO 2	0,005928	0,00	100,00

Tabela 5.1 – Medidas de Precisão dos Resultados para a Base de Dados "IRIS"

5.3.2 - Ilustração da Classificação

Os gráficos 5.1 a 5.2 exibem a classificação dos padrões para a base de dados IRIS no CASO 1 e no CASO 2.

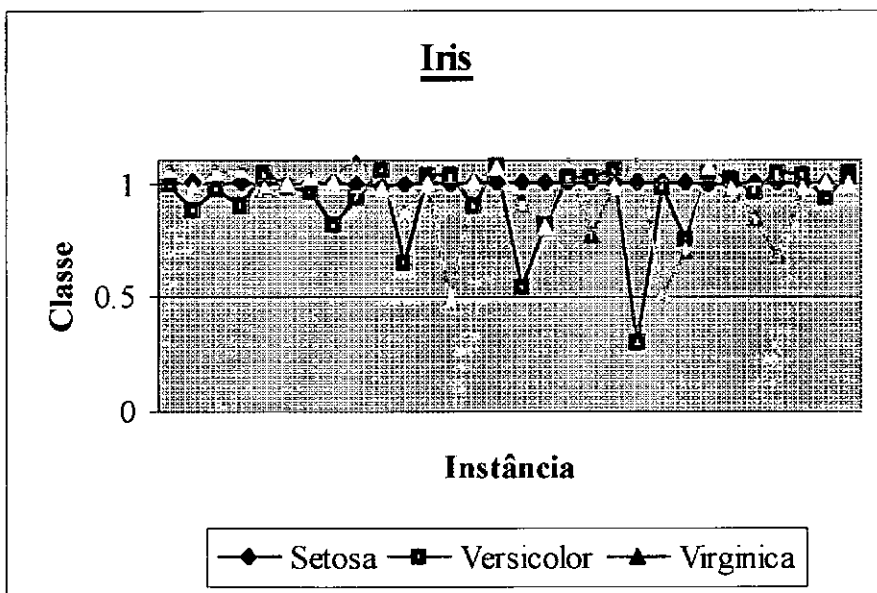


Figura 5.1 – Classificação da Base de Dados Iris pelo algoritmo da Pseudo-Inversa Matriz

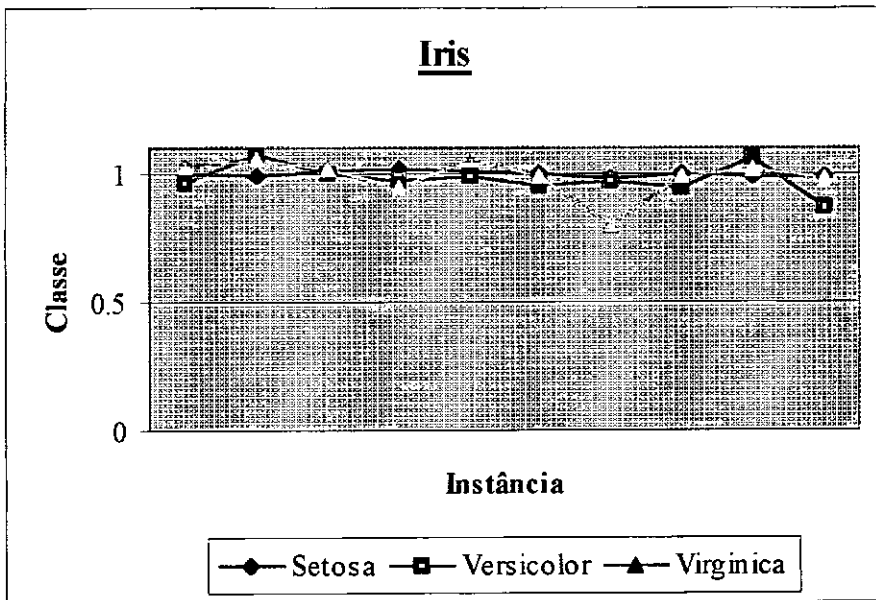


Figura 5.2 – Classificação da Base de Dados Iris pelo algoritmo da Regra Delta

5.4 - Apresentação dos Resultados para a Base de Dados "Aprovação de Crédito"

5.4.1 - Tabela com Resultados

A tabela 5.2 apresenta o Erro Médio Quadrático (RMSE) e a percentagem de padrões classificados correta e incorretamente para base *Aprovação de Crédito* no CASO 1 e no CASO 2.

ALGORITMO	RMSE	% PAD-INCORR	% PAD-CORRET
CASO 1	0,010023	10,97	89,03
CASO 2	0,016302	12,07	87,93

Tabela 5.2 – Medidas de Precisão dos Resultados para a Base de Dados "Aprovação de Crédito"

5.4.2 - Ilustração da Classificação

Os gráficos 5.3 a 5.4 mostram a classificação dos padrões para a base de dados *Aprovação de Crédito* no CASO 1 e no CASO 2.

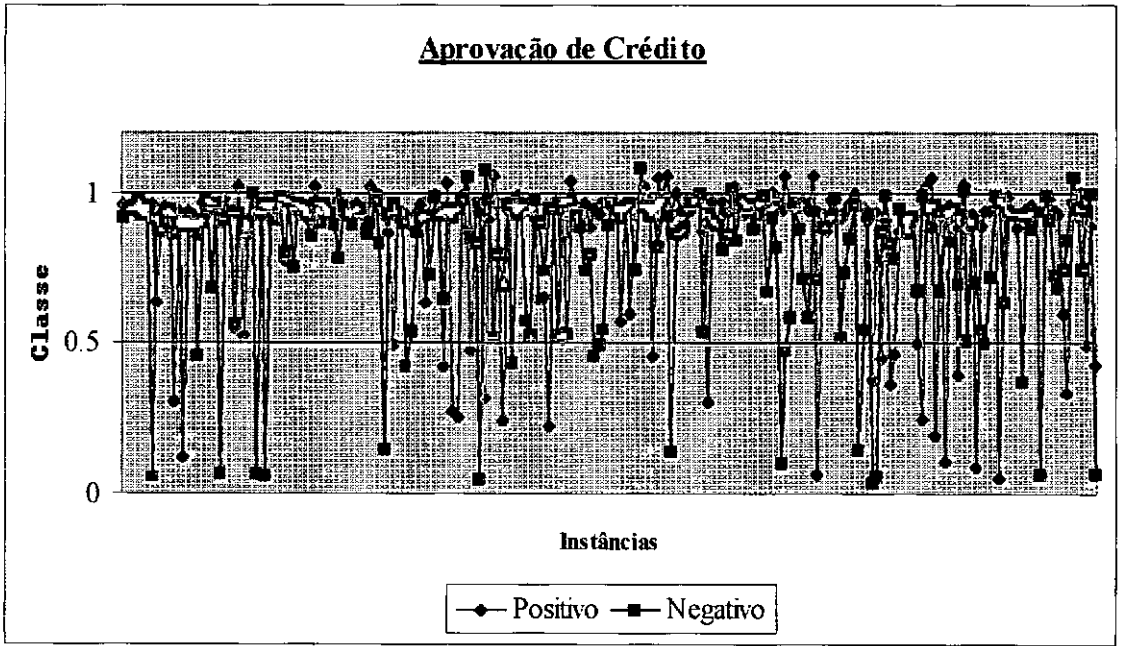


Figura 5.3 – Classificação da Base de Dados Aprovação de Crédito pelo algoritmo da Pseudo-Inversa Matriz

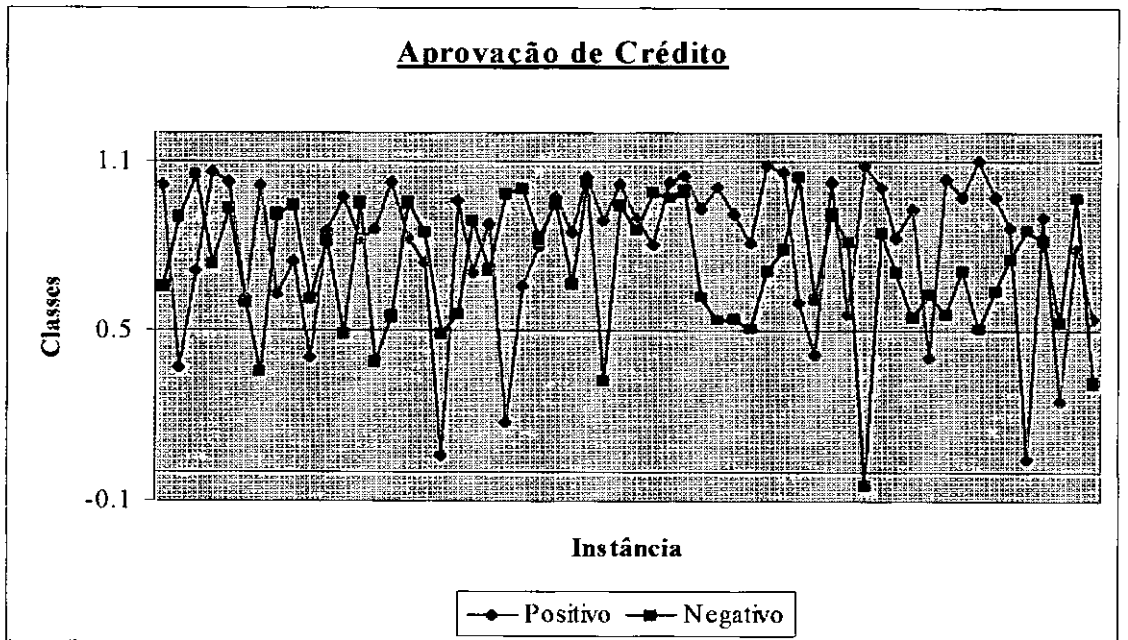


Figura 5.4 – Classificação da Base de Dados Aprovação de Crédito pelo algoritmo da Regra Delta

5.5 - Apresentação dos Resultados Base de Dados "Predição de Sinais de Vibração"

5.5.1 - Tabela com Resultados

A tabela 5.3 apresenta o Erro Médio Quadrático (RMSE) e a percentagem de padrões classificados correta e incorretamente para base *Predição de Sinais de Vibração* para o CASO 1 e para o CASO 2.

ALGORITMO	RMSE	FCORR
CASO 1	0,003047	0,7533
CASO 2	0,001849	0,7671

Tabela 5.3 – Medidas de Precisão dos Resultados para a Base de Dados "Predição de Sinais de Vibração"

5.5.2 - Ilustração da Predição

Os gráficos de 5.5 a 5.14 ilustram a precisão obtida para os a base de dados *Predição de Sinais de Vibração* para o CASO 1 e CASO 2.

Considera-se, para todos os gráficos deste item, que as letras maiúsculas representam os valores reais e as minúsculas, os valores desejados.

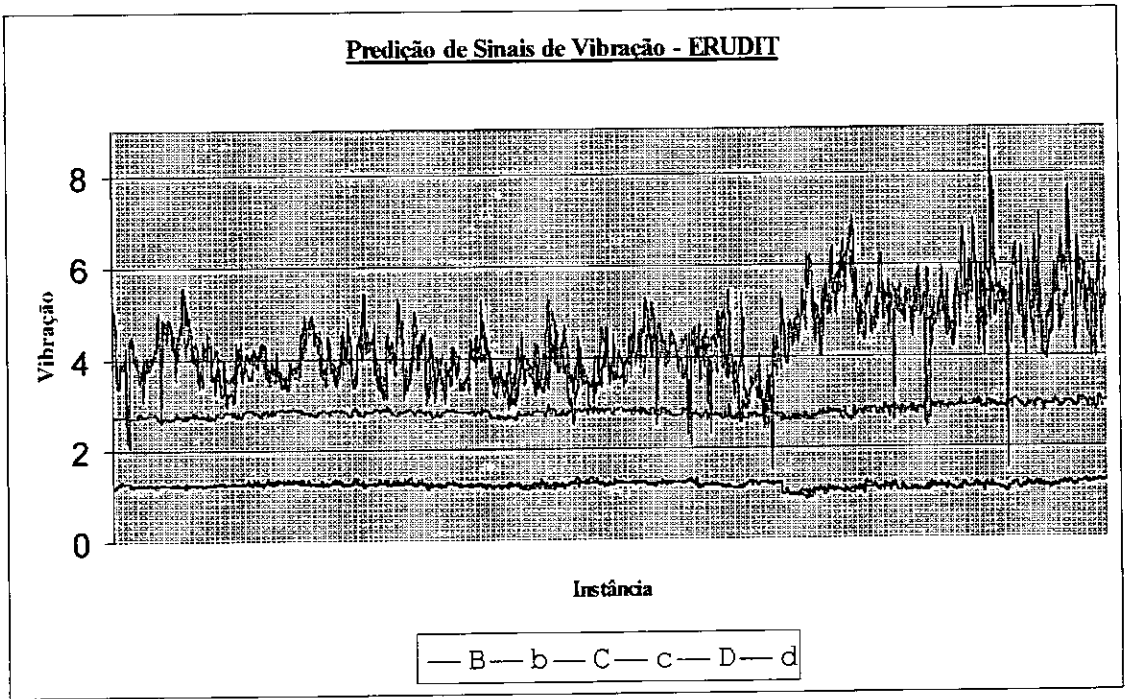


Figura 5.5 – Classificação da Base de Dados Predição de Sinais de Vibração pelo algoritmo da Regra Delta

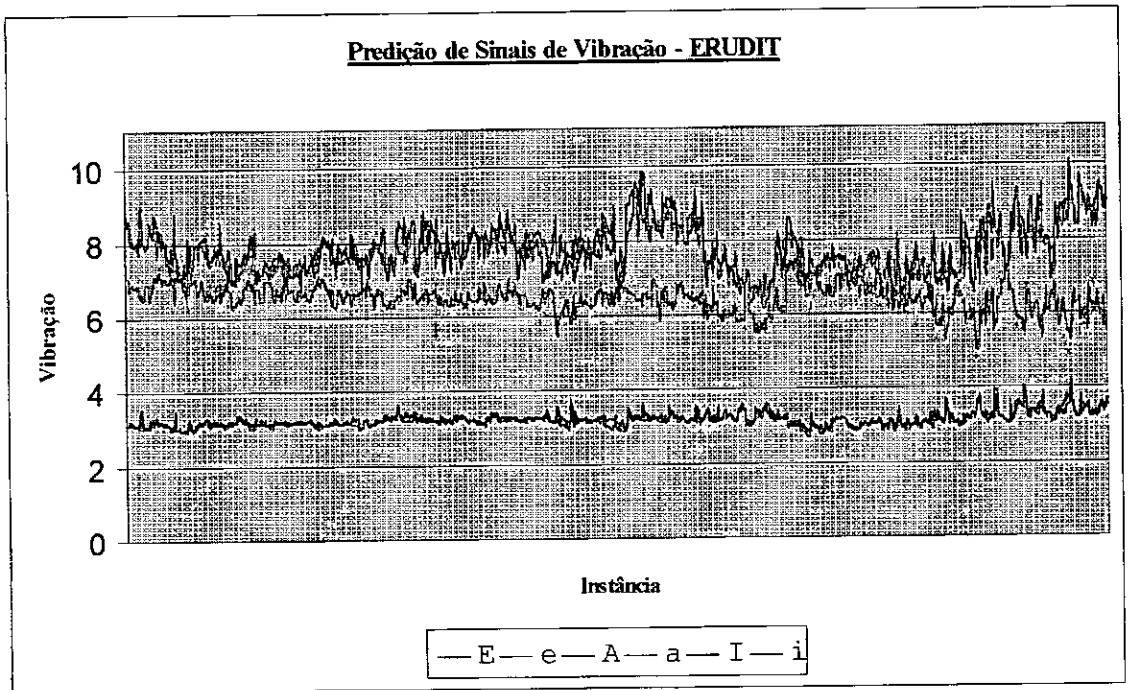


Figura 5.6 – Classificação da Base de Dados Predição de Sinais de Vibração pelo algoritmo da Regra Delta

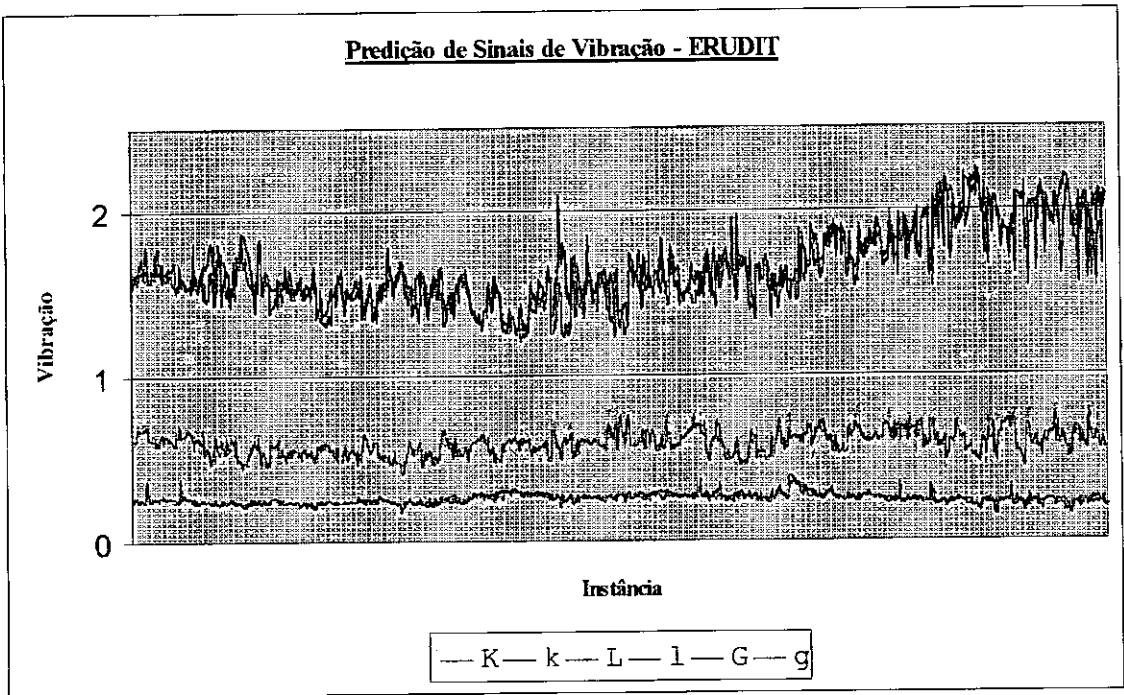


Figura 5.7 – Classificação da Base de Dados Predição de Sinais de Vibração pelo algoritmo da Regra Delta

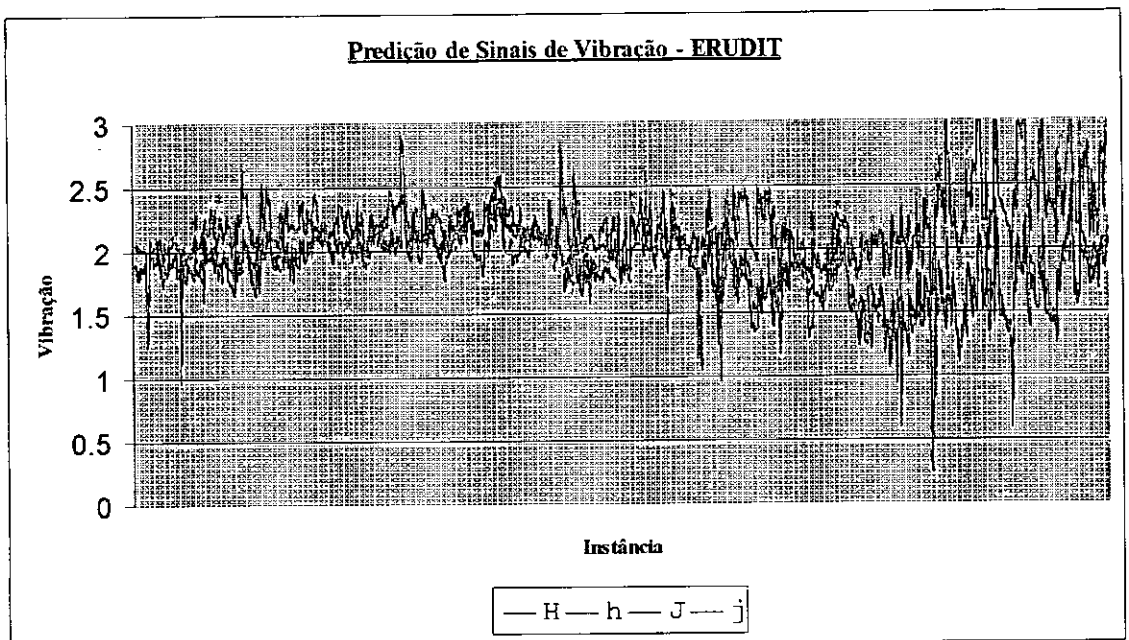


Figura 5.8 – Classificação da Base de Dados Predição de Sinais de Vibração pelo algoritmo da Regra Delta

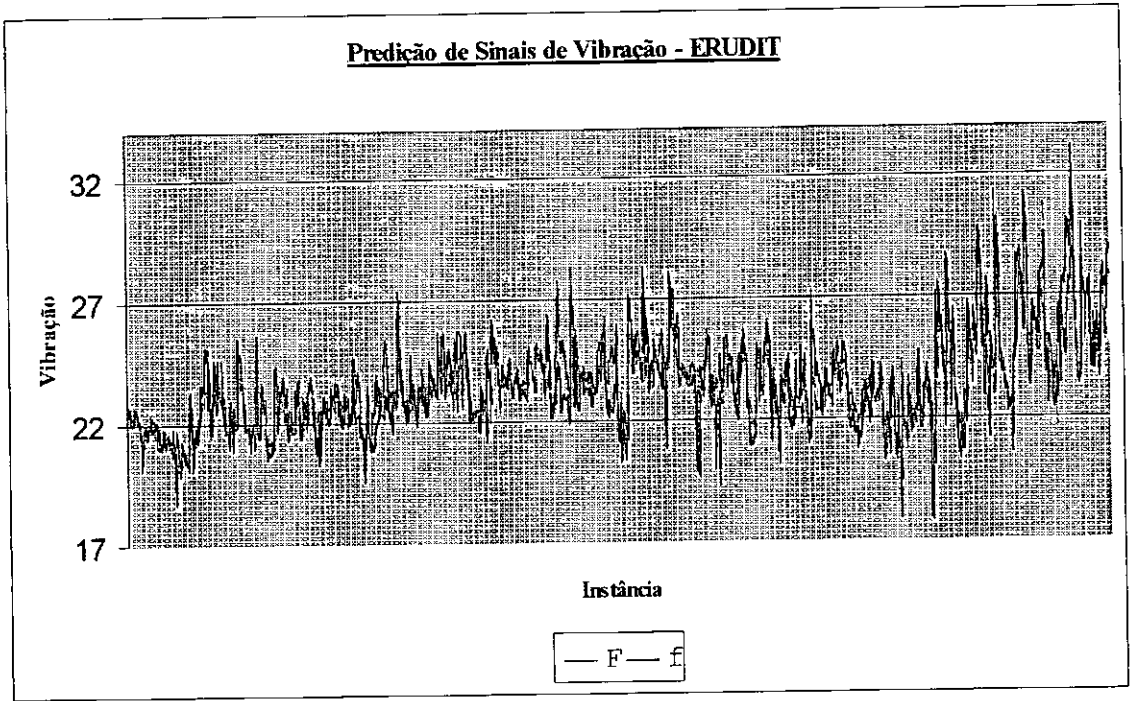


Figura 5.9 – Classificação da Base de Dados Predição de Sinais de Vibração pelo algoritmo da Regra Delta

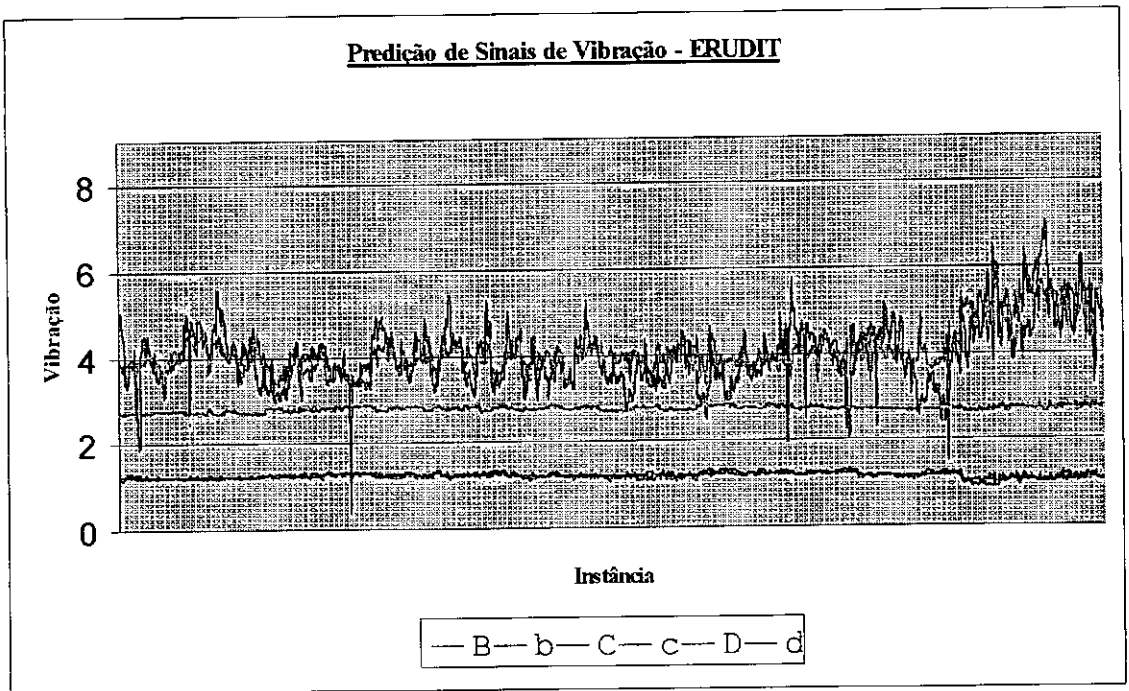


Figura 5.10 – Classificação da Base de Dados Predição de Sinais de Vibração pelo algoritmo da Pseudo-Inversa Matriz

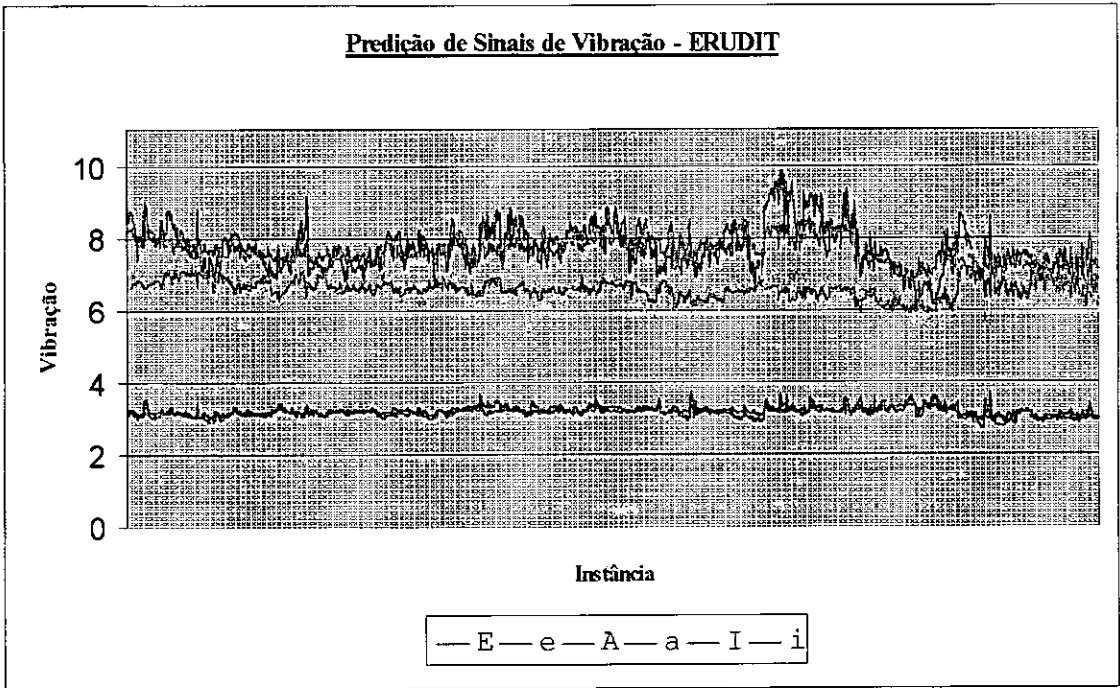


Figura 5.11 – Classificação da Base de Dados Predição de Sinais de Vibração pelo algoritmo da Pseudo-Inversa Matriz

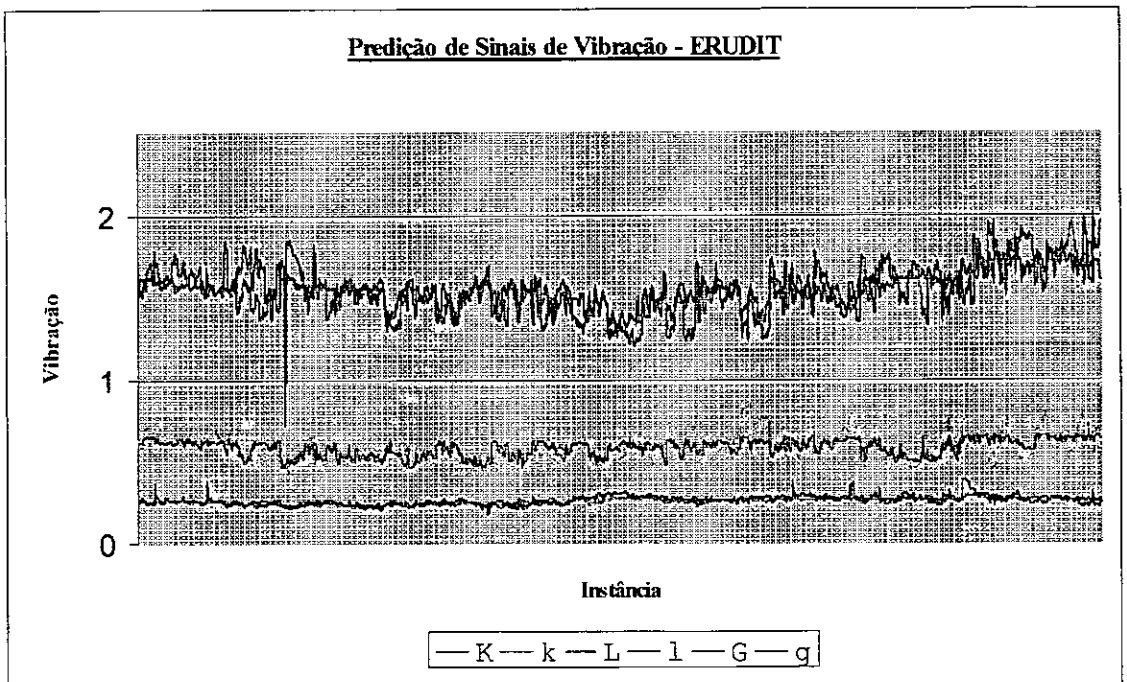


Figura 5.12 – Classificação da Base de Dados Predição de Sinais de Vibração pelo algoritmo da Pseudo-Inversa Matriz

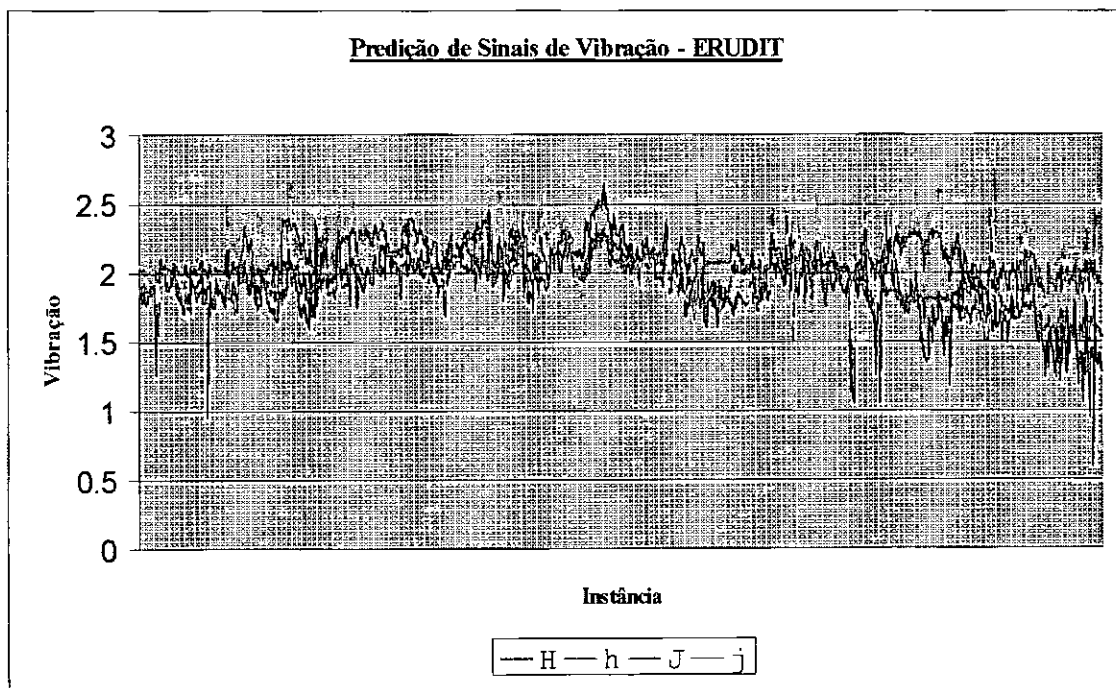


Figura 5.13 – Classificação da Base de Dados Predição de Sinais de Vibração pelo algoritmo da Pseudo-Inversa

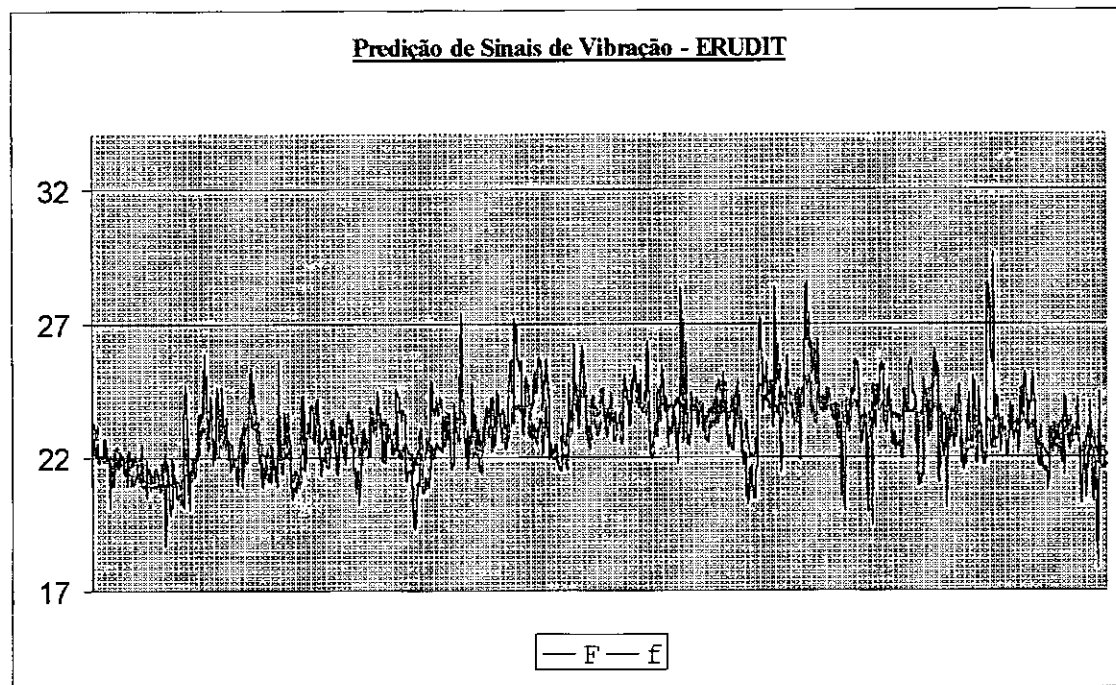


Figura 5.14 – Classificação da Base de Dados Predição de Sinais de Vibração pelo algoritmo da Pseudo-Inversa Matriz

5.6 - Análise dos Resultados

Comparando-se os resultados obtidos pelos dois algoritmos para a mesma base de dados, observa-se que não há diferenças significativas entre os resultados alcançados. Analisando-os, para cada base de dados através das tabelas e gráficos apresentados no itens 5.3, 5.4 e 5.5, verifica-se que:

⇒ Base de Dados IRIS

A classificação correta de todos os padrões pelo algoritmo que utiliza a Regra Delta para estimar os pesos. Para o algoritmo da Pseudo-Inversa Matriz o erro é de apenas 1,9% na classificação dos padrões.

Observa-se, desta forma, que os resultados alcançados pelos dois algoritmos são praticamente os mesmos. Verifica-se também, que as duas redes possuem configurações bem próximas, em relação ao número de funções de bases radiais usadas. Destaca-se porém, a utilização da tangente hiperbólica na camada de saída da rede pertencente ao CASO 2. A aplicação da tangente hiperbólica é certamente, neste caso, a explicação para o ajuste mais preciso obtido pelo algoritmo da Regra Delta.

⇒ Base de Dados Aprovação de Crédito

A classificação correta de 89,03% dos padrões para o algoritmo da pseudo-inversa matriz e 87,93% para o algoritmo da Regra Delta. Pode-se verificar que a configuração da rede para o CASO 1 possui 30 funções de base radial na camada escondida e funções de transferência linear na camada de saída. A configuração adotada para o CASO 2 possui 50 funções de base radial na camada escondida, e a função de transferência aplicada às unidades da camada de saída é a sigmóide.

Observa-se que, embora a rede do CASO 2 tenha mais funções de base radial e, portanto, é maior que a rede do CASO 1, esta alcançou resultado ligeiramente superior ao obtido pelo algoritmo da Regra Delta (CASO 2).

⇒ Base de Dados Predições de Sinais de Vibração

O erro médio quadrático e o grau de correlação obtidos pelos dois algoritmos podem ser considerados como muito bons. Verifica-se facilmente pelos gráficos do item 5.5, que são poucos os pontos em que o resultado desejado e o predito não estão totalmente superpostos.

Observa-se que a rede montada para o algoritmo da Pseudo-Inversa Matriz contém 198 funções de base radial na camada escondida e função de transferência linear na camada de saída.

A configuração da rede para o algoritmo da Regra Delta fez-se com 52 funções de base radial na camada escondida e utilizando-se a sigmóide como função de transferência para a camada de saída.

Verifica-se que a rede construída para o CASO 2 é consideravelmente menor do que a rede configurada para o CASO 1. Pode-se observar, entretanto, que os resultados alcançados pelo algoritmo da Regra Delta são superiores aos obtidos pelo da Pseudo-Inversa Matriz.

Atribui-se os resultados alcançados pelo algoritmo da Regra Delta a dois fatores:

- A capacidade do algoritmo “*Dynamic K-Means*”, em alocar com mais precisão as funções de base radial no espaço vetorial de entrada;
- A aplicação da sigmóide como função de transferência da camada de saída da rede.

5.7 - Resultados Numéricos Obtidos

Apresenta-se nas tabelas 5.4 à 5.15 os valores desejados e obtidos pela rede, bem como a diferença entre eles.

Os dados apresentados foram escolhidos aleatoriamente.

5.7.1 - Base de Dados "Iris"

IRIS SETOSA			IRIS VERSICOLOR			IRIS VIRGINICA		
Real	Estimado	Diferença	Real	Estimado	Diferença	Real	Estimado	Diferença
1,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
1,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
1,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
1,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
1,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
1,00	0,96	0,04	0,00	0,04	-0,04	0,00	0,00	0,00
1,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
0,00	0,00	0,00	1,00	0,88	0,12	0,00	0,12	-0,12
0,00	0,00	0,00	1,00	0,96	0,04	0,00	0,04	-0,04
0,00	0,00	0,00	1,00	0,54	0,46	0,00	0,46	-0,46
0,00	0,00	0,00	1,00	0,75	0,25	0,00	0,25	-0,25
0,00	0,01	-0,01	1,00	0,97	0,03	0,00	0,02	-0,02
0,00	-0,01	0,01	1,00	1,04	-0,04	0,00	-0,04	0,04
0,00	0,00	0,00	0,00	-0,06	0,06	1,00	1,06	-0,06
0,00	0,00	0,00	0,00	0,02	-0,02	1,00	0,97	0,03
0,00	0,00	0,00	0,00	-0,07	0,07	1,00	1,07	-0,07
0,00	0,00	0,00	0,00	0,01	-0,01	1,00	0,99	0,01
0,00	0,00	0,00	0,00	0,32	-0,32	1,00	0,68	0,32
0,00	0,00	0,00	0,00	0,01	-0,01	1,00	0,99	0,01
0,00	0,00	0,00	0,00	0,19	-0,19	1,00	0,82	0,18

Tabela 5.4 – Resultados do CASO 1

IRIS SETOSA			IRIS VERSICOLOR			IRIS VIRGINICA		
Real	Estimado	Diferença	Real	Estimado	Diferença	Real	Estimado	Diferença
1,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
1,00	0,99	0,01	0,00	0,01	-0,01	0,00	0,01	-0,01
1,00	1,01	-0,01	0,00	0,00	0,00	0,00	0,00	0,00
1,00	1,02	-0,02	0,00	-0,01	0,01	0,00	-0,02	0,02
1,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
1,00	1,00	0,00	0,00	0,00	0,00	0,00	-0,01	0,01
1,00	0,99	0,01	0,00	0,00	0,00	0,00	0,01	-0,01
0,00	-0,03	0,03	1,00	1,00	0,00	0,00	0,01	-0,01
0,00	-0,02	0,02	1,00	0,97	0,03	0,00	0,03	-0,03
0,00	0,00	0,00	1,00	0,99	0,01	0,00	0,00	0,00
0,00	0,00	0,00	1,00	0,95	0,05	0,00	0,03	-0,03
0,00	0,03	-0,03	1,00	0,94	0,06	0,00	0,05	-0,05
0,00	0,02	-0,02	1,00	1,07	-0,07	0,00	-0,07	0,07
0,00	-0,02	0,02	1,00	0,87	0,13	0,00	0,13	-0,13
0,00	0,01	-0,01	0,00	-0,07	0,07	1,00	1,05	-0,05
0,00	-0,02	0,02	0,00	-0,05	0,05	1,00	1,02	-0,02
0,00	-0,02	0,02	0,00	0,07	-0,07	1,00	0,95	0,05
0,00	0,01	-0,01	0,00	-0,07	0,07	1,00	1,05	-0,05
0,00	0,01	-0,01	0,00	0,21	-0,21	1,00	0,80	0,20
0,00	0,01	-0,01	0,00	0,01	-0,01	1,00	0,99	0,01

Tabela 5.5 – Resultados do CASO 2

5.7.2 - Base de Dados "Aprovação de Crédito"

APROVADO			NÃO APROVADO		
Real	Estimado	Diferença	Real	Estimado	Diferença
1,00	0,96	0,04	0,00	0,04	-0,04
0,00	0,10	-0,10	1,00	0,90	0,10
1,00	0,95	0,05	0,00	0,05	-0,05
1,00	0,93	0,07	0,00	0,07	-0,07
1,00	0,98	0,02	0,00	0,02	-0,02
0,00	0,35	-0,35	1,00	0,65	0,35
1,00	0,92	0,08	0,00	0,08	-0,08
0,00	0,05	-0,05	1,00	0,95	0,05
1,00	0,97	0,03	0,00	0,03	-0,03
1,00	0,99	0,01	0,00	0,01	-0,01
0,00	0,19	-0,19	1,00	0,81	0,19
0,00	0,41	-0,41	1,00	0,59	0,41
1,00	0,94	0,06	0,00	0,06	-0,06
0,00	0,05	-0,05	1,00	0,95	0,05
0,00	0,01	-0,01	1,00	0,99	0,01
1,00	0,93	0,07	0,00	0,07	-0,07
0,00	0,01	-0,01	1,00	0,99	0,01
0,00	0,16	-0,16	1,00	0,84	0,16
1,00	0,96	0,04	0,00	0,04	-0,04
0,00	0,93	-0,93	1,00	0,07	0,93

Tabela 5.6 – Resultados do CASO 1

APROVADO			NÃO APROVADO		
Real	Estimado	Diferença	Real	Estimado	Diferença
1,00	1,06	-0,06	0,00	-0,06	0,06
0,00	0,06	-0,06	1,00	0,93	0,07
1,00	0,71	0,29	0,00	0,31	-0,31
0,00	0,06	-0,06	1,00	0,96	0,04
1,00	0,81	0,19	0,00	0,22	-0,22
1,00	1,01	-0,01	0,00	-0,02	0,02
1,00	1,09	-0,09	0,00	-0,09	0,09
1,00	0,42	0,58	0,00	0,58	-0,58
1,00	0,83	0,17	0,00	0,17	-0,17
0,00	0,28	-0,28	1,00	0,72	0,28
1,00	0,93	0,07	0,00	0,08	-0,08
0,00	0,18	-0,18	1,00	0,82	0,18
1,00	1,01	-0,01	0,00	0,00	0,00
1,00	1,01	-0,01	0,00	-0,02	0,02
0,00	0,36	-0,36	1,00	0,63	0,37
0,00	0,49	-0,49	1,00	0,51	0,49
0,00	0,37	-0,37	1,00	0,65	0,35
0,00	0,26	-0,26	1,00	0,75	0,25
0,00	0,19	-0,19	1,00	0,82	0,18
0,00	0,04	-0,04	1,00	0,97	0,03

Tabela 5.7 – Resultados do CASO 2

5.7.3 - Base de Dados "Predição de Sinais de Vibração "

A			B			C		
Real	Estimado	Diferença	Real	Estimado	Diferença	Real	Estimado	Diferença
6,75	6,83	-0,08	3,82	3,88	-0,06	2,74	2,74	0,00
7,18	7,12	0,06	4,63	4,66	-0,03	2,69	2,68	0,01
6,86	6,61	0,25	4,87	3,92	0,95	2,82	2,74	0,08
6,71	6,40	0,31	4,02	3,96	0,06	2,87	2,85	0,02
6,59	6,48	0,11	3,84	3,95	-0,11	2,83	2,81	0,02
6,92	6,54	0,38	4,72	3,59	1,13	2,78	2,79	-0,01
6,66	6,24	0,42	4,60	3,80	0,80	2,74	2,82	-0,08
6,77	6,82	-0,05	4,17	4,30	-0,13	2,70	2,70	0,00
6,07	6,09	-0,02	4,24	4,48	-0,24	2,65	2,79	-0,14
6,00	6,02	-0,02	3,98	3,52	0,46	2,65	2,67	-0,02
5,49	5,57	-0,08	3,33	3,51	-0,18	2,77	2,77	0,00
6,43	6,53	-0,10	3,66	4,09	-0,43	2,73	2,69	0,04
7,69	7,44	0,25	4,79	4,40	0,39	2,57	2,61	-0,04
6,97	6,60	0,37	5,21	5,06	0,15	2,76	2,75	0,01
7,38	7,20	0,18	5,94	6,57	-0,63	2,64	2,61	0,03
6,31	6,69	-0,38	5,10	5,27	-0,17	2,83	2,81	0,02
6,31	6,44	-0,13	5,10	5,42	-0,32	2,82	2,81	0,01
7,02	6,86	0,16	5,34	5,10	0,24	2,92	2,90	0,02
5,92	6,06	-0,14	3,88	4,29	-0,41	2,92	2,94	-0,02
5,87	5,93	-0,06	6,45	6,12	0,33	3,09	3,09	0,00

Tabela 5.8 – Resultados do CASO 1

D			E			F		
Real	Estimado	Diferença	Real	Estimado	Diferença	Real	Estimado	Diferença
1,31	1,28	0,03	7,96	7,98	-0,02	22,68	22,12	0,56
1,21	1,23	-0,02	7,69	7,29	0,40	21,19	20,65	0,54
1,31	1,20	0,11	7,96	7,67	0,29	21,83	22,89	-1,06
1,22	1,22	0,00	8,27	8,00	0,27	24,29	25,39	-1,10
1,29	1,23	0,06	7,95	8,48	-0,53	23,29	24,22	-0,93
1,27	1,26	0,01	9,04	8,14	0,90	23,80	23,81	-0,01
1,21	1,26	-0,05	8,80	8,07	0,73	24,26	24,02	0,24
1,24	1,25	-0,01	8,21	8,21	0,00	23,82	23,86	-0,04
1,21	1,18	0,03	7,60	7,54	0,06	22,72	23,44	-0,72
1,12	1,18	-0,06	6,96	6,85	0,11	22,91	22,17	0,74
1,24	1,24	0,00	6,53	6,70	-0,17	24,55	24,92	-0,37
1,22	1,16	0,06	7,63	7,32	0,31	22,66	22,50	0,16
0,92	0,97	-0,05	8,00	8,05	-0,05	22,84	22,51	0,33
1,05	1,05	0,00	7,53	7,52	0,01	24,36	24,82	-0,46
1,05	1,01	0,04	7,18	6,95	0,23	21,59	22,32	-0,73
1,21	1,13	0,08	7,65	7,49	0,16	24,14	22,97	1,17
1,04	1,05	-0,01	6,46	6,86	-0,40	21,64	21,68	-0,04
1,06	1,09	-0,03	7,75	7,53	0,22	23,14	23,35	-0,21
1,14	1,11	0,03	8,04	7,89	0,15	25,00	24,95	0,05
1,28	1,27	0,01	9,27	9,36	-0,09	28,27	27,65	0,62

Tabela 5.9 – Resultados do CASO 1

G			H			I		
Real	Estimado	Diferença	Real	Estimado	Diferença	Real	Estimado	Diferença
1,67	1,63	0,04	1,87	1,86	0,01	3,07	3,10	-0,03
1,55	1,56	-0,01	1,85	1,81	0,04	3,03	2,99	0,04
1,36	1,33	0,03	2,18	2,11	0,07	3,12	3,14	-0,02
1,61	1,61	0,00	2,29	2,31	-0,02	3,39	3,42	-0,03
1,44	1,55	-0,11	1,94	2,02	-0,08	3,01	3,19	-0,18
1,51	1,47	0,04	2,14	1,94	0,20	3,22	3,23	-0,01
1,51	1,57	-0,06	2,03	1,91	0,12	3,10	3,21	-0,11
1,50	1,45	0,05	2,03	2,04	-0,01	3,18	3,20	-0,02
1,48	1,72	-0,24	1,71	1,68	0,03	3,08	3,17	-0,09
1,60	1,54	0,06	1,85	1,57	0,28	3,19	3,11	0,08
1,65	1,68	-0,03	1,77	1,83	-0,06	3,37	3,41	-0,04
1,52	1,54	-0,02	1,70	1,74	-0,04	3,18	3,15	0,03
1,45	1,56	-0,11	1,98	1,97	0,01	3,05	3,00	0,05
1,84	1,85	-0,01	1,91	1,99	-0,08	3,18	3,16	0,02
1,63	1,68	-0,05	1,49	1,58	-0,09	2,95	2,92	0,03
1,81	1,84	-0,03	1,65	1,55	0,10	3,16	3,08	0,08
1,82	1,82	0,00	1,36	1,40	-0,04	2,92	2,99	-0,07
1,89	1,89	0,00	1,45	1,44	0,01	3,13	3,11	0,02
2,06	2,07	-0,01	1,49	1,49	0,00	3,30	3,30	0,00
2,10	2,11	-0,01	2,07	2,04	0,03	3,61	3,56	0,05

Tabela 5.10 – Resultados do CASO 1

J			K			L		
Real	Estimado	Diferença	Real	Estimado	Diferença	Real	Estimado	Diferença
2,08	2,02	0,06	0,26	0,25	0,01	0,67	0,67	0,00
2,01	2,01	0,00	0,25	0,26	-0,01	0,66	0,63	0,03
2,19	2,04	0,15	0,25	0,27	-0,02	0,58	0,60	-0,02
2,53	2,48	0,05	0,22	0,25	-0,03	0,42	0,48	-0,06
1,83	2,06	-0,23	0,28	0,27	0,01	0,68	0,63	0,05
2,12	2,05	0,07	0,29	0,28	0,01	0,63	0,61	0,02
1,96	2,09	-0,13	0,29	0,27	0,02	0,66	0,60	0,06
2,04	2,05	-0,01	0,27	0,27	0,00	0,63	0,63	0,00
1,98	2,17	-0,19	0,28	0,26	0,02	0,64	0,61	0,03
2,09	2,02	0,07	0,26	0,27	-0,01	0,52	0,60	-0,08
2,38	2,43	-0,05	0,24	0,24	0,00	0,52	0,52	0,00
2,07	2,01	0,06	0,28	0,27	0,01	0,63	0,62	0,01
1,88	1,83	0,05	0,33	0,35	-0,02	0,63	0,63	0,00
2,22	2,18	0,04	0,27	0,27	0,00	0,53	0,54	-0,01
1,81	1,86	-0,05	0,25	0,25	0,00	0,71	0,69	0,02
2,26	2,08	0,18	0,26	0,26	0,00	0,60	0,61	-0,01
2,02	2,05	-0,03	0,23	0,24	-0,01	0,68	0,69	-0,01
2,26	2,26	0,00	0,24	0,23	0,01	0,69	0,70	-0,01
2,43	2,44	-0,01	0,25	0,24	0,01	0,63	0,63	0,00
2,87	2,82	0,05	0,22	0,22	0,00	0,56	0,55	0,01

Tabela 5.11 – Resultados do CASO 1

A			B			C		
Real	Estimado	Diferença	Real	Estimado	Diferença	Real	Estimado	Diferença
6,39	6,58	-0,19	4,99	4,47	0,52	2,78	2,75	0,03
6,48	6,49	-0,01	4,85	4,82	0,03	2,88	2,79	0,09
6,11	6,41	-0,30	5,87	4,01	1,86	2,93	2,71	0,22
6,65	6,54	0,11	1,29	3,51	-2,22	2,58	2,75	-0,17
5,76	6,43	-0,67	6,41	3,83	2,58	2,98	2,74	0,24
6,06	6,46	-0,40	4,65	3,90	0,75	2,93	2,72	0,21
5,84	6,51	-0,67	5,00	4,01	0,99	3,05	2,78	0,27
5,93	6,50	-0,57	3,87	3,88	-0,01	2,91	2,72	0,19
6,25	6,52	-0,27	4,51	3,91	0,60	2,92	2,71	0,21
5,43	6,44	-1,01	6,03	4,01	2,02	2,98	2,72	0,26
4,95	6,52	-1,57	5,77	3,86	1,91	2,98	2,74	0,24
5,77	6,42	-0,65	5,98	4,05	1,93	3,06	2,73	0,33
5,98	6,51	-0,53	4,92	3,88	1,04	2,98	2,74	0,24
6,83	6,56	0,27	4,57	3,74	0,83	2,99	2,71	0,28
6,47	6,57	-0,10	5,54	3,94	1,60	3,07	2,77	0,30
6,07	6,52	-0,45	5,05	3,92	1,13	3,03	2,78	0,25
5,71	6,33	-0,62	5,30	3,89	1,41	3,02	2,74	0,28
6,10	6,44	-0,34	4,74	3,87	0,87	2,96	2,73	0,23
5,45	6,35	-0,90	6,44	3,92	2,52	2,99	2,73	0,26
6,48	6,49	-0,01	4,25	3,87	0,38	2,91	2,72	0,19

Tabela 5.12 – Resultados do CASO 2

D			E			F		
Real	Estimado	Diferença	Real	Estimado	Diferença	Real	Estimado	Diferença
1,13	1,12	0,01	6,86	7,24	-0,38	21,89	22,94	-1,05
1,10	1,16	-0,06	7,83	7,44	0,39	21,64	22,95	-1,31
1,10	1,17	-0,07	6,56	7,48	-0,92	24,50	23,58	0,92
1,08	1,20	-0,12	7,30	7,94	-0,64	20,25	22,89	-2,64
1,19	1,20	-0,01	8,96	7,85	1,11	29,73	23,45	6,28
1,15	1,19	-0,04	7,57	7,76	-0,19	25,38	23,15	2,23
1,21	1,23	-0,02	8,53	8,28	0,25	26,86	24,43	2,43
1,14	1,20	-0,06	7,62	7,79	-0,17	23,95	22,97	0,98
1,07	1,20	-0,13	6,88	7,75	-0,87	22,75	22,78	-0,03
1,23	1,21	0,02	8,94	7,87	1,07	27,76	23,68	4,08
1,15	1,23	-0,08	8,56	8,04	0,52	28,96	23,42	5,54
1,22	1,23	-0,01	9,63	8,09	1,54	27,34	24,09	3,25
1,19	1,23	-0,04	8,61	8,13	0,48	25,56	23,80	1,76
1,19	1,23	-0,04	9,03	7,94	1,09	23,97	23,33	0,64
1,18	1,23	-0,05	9,84	8,38	1,46	25,70	24,39	1,31
1,22	1,22	0,00	8,87	8,15	0,72	27,73	23,75	3,98
1,28	1,20	0,08	9,05	7,84	1,21	27,77	23,94	3,83
1,16	1,21	-0,05	8,44	7,85	0,59	25,14	23,38	1,76
1,21	1,21	0,00	9,29	7,80	1,49	29,74	23,76	5,98
1,13	1,20	-0,07	8,34	7,71	0,63	23,31	23,12	0,19

Tabela 5.13 – Resultados do CASO 2

G			H			I		
Real	Estimado	Diferença	Real	Estimado	Diferença	Real	Estimado	Diferença
1,72	1,70	0,02	1,38	1,59	-0,21	3,02	3,02	0,00
1,86	1,74	0,12	1,36	1,66	-0,30	3,02	3,03	-0,01
2,16	1,61	0,55	1,51	1,72	-0,21	3,19	3,30	-0,11
1,72	1,61	0,11	0,55	1,70	-1,15	3,63	3,28	0,35
2,07	1,59	0,48	2,05	1,84	0,21	3,79	3,28	0,51
2,04	1,58	0,46	1,52	1,78	-0,26	3,39	3,23	0,16
2,14	1,55	0,59	1,83	2,02	-0,19	3,44	3,25	0,19
2,02	1,58	0,44	1,42	1,76	-0,34	3,23	3,21	0,02
1,95	1,56	0,39	1,47	1,78	-0,31	3,06	3,16	-0,10
2,07	1,50	0,57	1,91	1,92	-0,01	3,48	3,16	0,32
2,20	1,52	0,68	2,10	1,91	0,19	3,68	3,20	0,48
2,01	1,47	0,54	2,10	1,97	0,13	3,52	3,18	0,34
2,04	1,50	0,54	1,79	1,95	-0,16	3,42	3,21	0,21
1,91	1,42	0,49	1,68	2,01	-0,33	3,28	3,14	0,14
1,95	1,49	0,46	1,92	2,05	-0,13	3,48	3,22	0,26
2,11	1,59	0,52	1,99	1,92	0,07	3,53	3,25	0,28
2,08	1,59	0,49	2,03	1,90	0,13	3,51	3,30	0,21
2,07	1,55	0,52	1,52	1,85	-0,33	3,32	3,22	0,10
2,07	1,56	0,51	2,17	1,88	0,29	3,72	3,26	0,46
1,89	1,54	0,35	1,57	1,86	-0,29	3,21	3,15	0,06

Tabela 5.14 – Resultados do CASO 2

J			K			L		
Real	Estimado	Diferença	Real	Estimado	Diferença	Real	Estimado	Diferença
2,08	1,98	0,10	0,22	0,26	-0,04	0,73	0,64	0,09
1,93	2,05	-0,12	0,24	0,26	-0,02	0,71	0,63	0,08
2,54	2,17	0,37	0,21	0,27	-0,06	0,53	0,58	-0,05
1,76	2,06	-0,30	0,33	0,28	0,05	0,72	0,60	0,12
3,18	2,15	1,03	0,22	0,27	-0,05	0,49	0,58	-0,09
2,57	2,10	0,47	0,24	0,27	-0,03	0,60	0,59	0,01
2,74	2,14	0,60	0,21	0,28	-0,07	0,59	0,59	0,00
2,35	2,08	0,27	0,25	0,27	-0,02	0,66	0,60	0,06
2,21	2,04	0,17	0,24	0,27	-0,03	0,70	0,60	0,10
2,84	2,04	0,80	0,22	0,27	-0,05	0,58	0,61	-0,03
3,12	2,08	1,04	0,19	0,27	-0,08	0,51	0,59	-0,08
2,74	2,02	0,72	0,22	0,27	-0,05	0,54	0,63	-0,09
2,52	2,08	0,44	0,23	0,27	-0,04	0,60	0,60	0,00
2,11	2,06	0,05	0,23	0,27	-0,04	0,78	0,63	0,15
2,47	2,08	0,39	0,22	0,28	-0,06	0,63	0,61	0,02
2,79	2,14	0,65	0,21	0,27	-0,06	0,59	0,58	0,01
2,82	2,20	0,62	0,21	0,27	-0,06	0,58	0,56	0,02
2,41	2,10	0,31	0,23	0,27	-0,04	0,64	0,59	0,05
3,12	2,15	0,97	0,18	0,27	-0,09	0,50	0,58	-0,08
2,11	2,08	0,03	0,22	0,27	-0,05	0,75	0,61	0,14

Tabela 5.15 – Resultados do CASO 2

6 – CONCLUSÃO

As Redes Neurais com Função de Base vem despertando grande interesse desde os trabalhos iniciais de Broomhead e Lowe e de Moody e Darken. Atualmente já há disponível uma razoável base teórica sobre o assunto, bem como, um conjunto de resultados em aplicações práticas. Os avanços nas pesquisas, sobre o assunto, possibilitam a aplicação com sucesso das RBFN's em reconhecimento de padrões, classificação, regressão, predição de séries temporais, reconhecimento de imagens e em modelagem de fenômeno de cognição. [8, 13, 21, 22]

Os resultados alcançados neste trabalho com a aplicação de softwares que utilizam os algoritmos da pseudo-inversa matriz e da regra delta, mostram que:

- ⇒ As redes neurais com função de base radial alcançam resultados com excelente grau de precisão e treinamento bem rápido, se comparados a outros modelos;
- ⇒ Algoritmos de clusterização com habilidade de alocar mais adequadamente as funções de base radial no espaço vetorial de entrada - independentes das condições de inicialização -, permitem configurações de redes que alcançam resultados mais precisos;
- ⇒ A utilização de funções não lineares na camada de saída possibilita o aumento do desempenho da rede;

- ⇒ Pode-se construir redes menores que apresentam excelentes resultados. Para tanto, é necessário que se tenha correta alocação das funções de base radial no espaço vetorial de entrada;
- ⇒ As redes neurais com função de base radial podem ser uma interessante alternativa aos *Multi Layer Perceptron*, pois podem ser treinadas com muito mais rapidez e excelente performance nos resultados.

Fundamentado no desenvolvimento deste trabalho, propõe-se a elaboração de alguns trabalhos teóricos e de experimentação prática para as Redes Neurais com Função de Base Radial:

- ⇒ Aprofundar os estudos a partir do artigo "*Learning without local minima in Radial Basis Function*", objetivando desenvolver algoritmos que alcancem com rapidez o mínimo da função custo para algumas classes de RBFNs; [13, 28]
- ⇒ Aplicação de Algoritmos Genéticos para determinação da configuração e dos parâmetros ótimos das RBFN's;
- ⇒ Desenvolvimento de trabalhos para extração de regras fuzzy das RBFN's, levando-se em conta o caráter local das funções de base radial;
- ⇒ Elaboração de estudos para treinamento das Redes Neurais com Função de Base Radial com Regras Fuzzy; [6, 22]
- ⇒ Aprofundamento no campo da teoria da regularização afim de obter, através da álgebra linear, os parâmetros ótimos das redes; [29, 30, 31, 32]
- ⇒ Comparar o desempenho dos *Multi Layer Perceptron* com as Redes com Função de Base Radial com alocação dinâmica de centros; [33]
- ⇒ Desenvolver uma aplicação prática para testar o desempenho dos algoritmos *Dynamic Competitive Learning* e o *Dynamic Regression Tree*; [20, 21]
- ⇒ Testar o desempenho do algoritmo GNG-U (*Growing Neural Gás with Utility Criterion*), desenvolvido para aplicação em bases de dados com distribuição não estacionárias, como processos técnicos, biológicos e econômicos. [34]

REFERÊNCIAS BIBLIOGRÁFICAS

- 1 - McCULLOCH, W., PITTS, W., "A Logical Calculus of the Ideas Immanent in Nervous Activity", *Bulletin of Mathematical Biophysics*, v. 5. pp. 115-133, 1943.
- 2 - ROSEMBLATT, F., *Principles of Neurodynamics*. Spartan Book, New York. 1962.
- 3 - BROOMHEAD, D. S, LOWE, D., "Multivariable Functional Interpolation and Adaptive Networks", *Complex Systems*. Vol. 2, pp. 321-355, 1988.
- 4 - MOODY, J., DARKEN, C., *Learning with Localized Receptive Fields*. In: Touretsky D., Hilton G., Sejnowski T., editors, *Connectionist Models Summer Scholl*, Carnegie Mellon University. 1988.
- 5 - POGGIO, T. GIROSI, F., "Regularization Algorithms for Learning that are Equivalent to Multilayer Networks", *Science*. v. 247, pp. 978-982. 1990.
- 6 - JANG, J. S. R., JUN, C. T., "Functional Equivalence Between Radial Basis Function Networks and Fuzzy Inference Systems", *IEEE Transaction on Neural Network*, v. 4, n. 1, pp. 156-159. 1993.
- 7 - FURTADO, M.I.V., *Um Estudo Sobre Predição da Qualidade do Ar Utilizando Redes Neurais*. Tese M. Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil. 1999.
- 8 - ORR, M. J. L., *Introduction to Radial Basis Function Networks*. Center for Cognitive Science, University of Edinburg, 2, Buccleuch Place, Edinburg EH8 9 LW, Scotland. 1996.
- 9 - MOODY, J., DARKEN, C., "Fast Learning in Networks of Locally-tuned Processing Units". *Neural Computation*, v. 1, pp. 281-294. 1989.
- 10 - TIKHONOV, A. N., "On Solving Inverse Posed Problems and Method of Regularization", *Doklady Akademicheskoi Nauki, USSR*, v. 151, pp. 501-504. 1963.

- 11 - HAYKIN, S., *Neural Networks: a Comprehensive Foundation*. 2nd ed. New Jersey, Prentice Hall. 1999.
- 12 - SCOTT, D. W., *Multivariate Density Estimation: Theory, Practice and Visualization*, Wiley-Interscience. 1992.
- 13 - BLANZIERE, E., *Learning Algorithms for Radial Basis Function Network: Synthesis, Experiments and Cognitive Modeling*. PHD Dissertation, Center of Cognitive Science, University and Polytechnic of Turin. 1998.
- 14 - FRITZKE, B., "Incremental Neuro-Fuzzy Systems", *International Symposium on Optical Science, Engineering and Instrumentation*, San Diego, 27-01. August 1997.
- 15 - *Neural Computing: A Technology Handbook for Neuralworks Professional II/Plus and Neuralworks Explorer*. 1 ed.. Pittsburgh, Aspen Technology, Inc.. 1998.
- 16 - PLATT, J. C., "A Resource-Allocating Network for Function Interpolation". *Neural Computation*, v. 3, n. 2, pp. 213-225. 1991.
- 17 - FRITZKE, B., "Fast Learning with Incremental RBF Networks", *Neural Processing Letters*, v. 1, n.1, pp. 2-5. 1994.
- 18 - FRITZKE, B., *Growing Cell Structures - A Self-Organization Network for Unsupervised and Supervising Learning*, International Computer Science Institute, Berkeley. 1993.
- 19 - FRITZKE, B., *A Growing Neural GAS Network Learns Topologies*. In: Tesauro G., Tourestzky D. S., Leen T. K., editors, *Advances in Neural Information Processing Systems 7*, Mit Press, Cambridge MA. 1995.
- 20 - BLANZIERE, E., GRONDANA, A., *An Incremental Algorithm for Learning Radial Basis Functions Networks*. In: *Proceedings of the International Conference on Fuzzy Systems*, New Orleans. 1996.
- 21 - BLANZIERE, E., KATENKAMP, P., *Learning Radial Basis Functions Networks on-line*. In: *13-th International Conference on Machine Learning*, Bani. pp. 37-45. 1996.
- 22 - BLANZIERE, E., GRONDANA, A., *Mapping Symbolic Knowledge into Locally Receptive Fields Networks*. In: Gori, M., Sada, G., editors, *Topics in Artificial Intelligence*, Spring-Verlag. 1995.
- 23 - *User's Guide: Pattern Recognition Workbench 2.1*. 1 ed. (Lincoln) MA, Unica Technologies, Inc. 1996.

- 24 - MARTINEZ, T. M., SCHULTEN, K. J., *A Neural-Gas Network Learns Topology*. In: Kohonen, T., Makisana, K., Simula, O., Kangas, J., editors, *Artificial Neural Networks*. pp. 397-402. North-Holland, Amsterdam. 1991.
- 25 - MURPHY, P. M., *UCI Repository of Machine Learning Databases and Domain Theories*, University of California, Dec. 1999. Available from World Wide Web <<http://ftp.ics.uci.edu/pub/machine-learning-databases/iris>>.
- 26 - MURPHY, P. M., *UCI Repository of Machine Learning Databases and Domain Theories*, University of California, Dec. 1996. Available from World Wide Web <<http://ftp.ics.uci.edu/pub/machine-learning-databases/credit-screening>>.
- 27 - HECK, S., *ERUDIT Service Center*, Elite Foundation, Germany. Available from World Wide Web <<http://www.mitgmbh.de/erudit>>.
- 28 - BIANCHINI, M., FRASCONI, P., GORI, M., "Learning without Local Minim in Radial Basis Functions Networks", *IEEE Transaction on Neural Networks*, v. 6, n. 3, pp. 749-756. 1995.
- 29 - ORR, M.J.L., *Regularization in the Selection of Radial Basis Functions Centers*, Center for Cognitive Science, Edinburg University, 2, Buccleuch Place, Edinburgh EH8 9LW, UK.
- 30 - ORR, M.J.L., *Optimizing the widths of Radial Basis Functions*. Center for Cognitive Science, Edinburgh University, Edinburgh EH8 9LW, UK.
- 31 - ORR, M.J.L., *Local Smoothing of Radial Basis Functions Networks*. Center for Cognitive Science, Edinburgh University, Edinburgh EH8 9LW, UK.
- 32 - ORR, M.J.L., *Recent Advances in Radial Basis Functions Networks*. Institute for Adaptive and Neural Computation Division of Informatics, Edinburgh University, Edinburgh EH9 9LW, UK. 1999.
- 33 - WETTSCHENECH, D., DIETTERICH, T., "Improving the Performance of Radial Basis Functions Networks by Learning Center Locations", *Advances in Neural Information Processing Systems*, v. 4, pp. 1133-1140. 1992.
- 34 - FRITZKE A., *A Self-Organization Network that can Follow Non-Stationary Distributions*. In: International Conference on Artificial Neural Networks, pp. 613-618. 1997.
- 35 - FUNCHS, W. H. J., *The Theory of Functions of One Complex Variable*. Princeton, D. Van Nostrand Company, 193p. 1967.