

SEGMENTAÇÃO E RECONHECIMENTO DE DÍGITOS MANUSCRITOS UTILIZANDO REDES NEURAIS

Trabalho de Conclusão de Curso
Engenharia da Computação

Sandro Luiz Lourenço de Mello
Orientador: Prof. Adriano Lorena Inácio de Oliveira

Recife, dezembro de 2007



SEGMENTAÇÃO E RECONHECIMENTO DE DÍGITOS MANUSCRITOS UTILIZANDO REDES NEURAIS

Trabalho de Conclusão de Curso

Engenharia da Computação

Este Projeto é apresentado como requisito parcial para obtenção do diploma de Bacharel em Engenharia da Computação pela Escola Politécnica de Pernambuco – Universidade de Pernambuco.

Sandro Luiz Lourenço de Mello
Orientador: Prof. Adriano Lorena Inácio de Oliveira

Recife, dezembro de 2007



UNIVERSIDADE
DE PERNAMBUCO

Sandro Luiz Lourenço de Mello

**SEGMENTAÇÃO E
RECONHECIMENTO DE DÍGITOS
MANUSCRITOS UTILIZANDO
REDES NEURAIIS**

Resumo

Este trabalho apresenta uma forma de reconhecer dígitos manuscritos em documentos históricos, mais especificamente as datas contidas nesses documentos com o propósito de ordenação por data. No presente trabalho é feito um estudo comparativo entre as técnicas de Redes Neurais Artificiais (RNAs), do tipo MLP (*Multilayer perceptron*), RBFN (*Radial Basis Functions Networks*) e PNN (*Probabilistic Neural Network*), sendo que os dois últimos utilizando o algoritmo de treinamento DDA (*Dynamic Decay Adjustment*). Para a rede RBFN foi aplicado o conceito de nível de confiança para aceitação de uma decisão de classificação, que impõe uma margem de aceitação das entradas classificadas. As entradas apresentadas à rede que não forem aceitas, serão classificadas como padrões espúrios ou novidades. Capturamos as entradas a partir de duas bases: uma extraída de documentos históricos e outra da base do *MNIST Database of Handwritten Digits*.

Abstract

This work presents a form of recognizing handwritten digits in historical documents, more specifically the dates contained in those documents with the purpose of ordination for date. In the present work it is made a comparative study among the techniques of Artificial Neural Network (ANN), of the type MLP (Multilayer perceptron), RBFN (Radial Basis Functions Networks) and PNN (Probabilistic Neural Network), and the last two using the training algorithm DDA (Dynamic Decay Adjustment). To the RBFN the concept of trust level was applied for acceptance of a classification decision, that imposes a margin of acceptance of the classified entrances. The entrances presented to the network that be not accepted, will be classified as spurious patterns or innovations. We captured the entrances starting from two bases: an extracted of historical documents and another of the base of MNIST Database of Handwritten Digits.

Sumário

Índice de Figuras	v
Índice de Tabelas	vii
Tabela de Símbolos e Siglas	viii
1 Introdução	10
1.1 Definição do Problema	10
1.2 Objetivos	12
1.3 Organização	12
2 Técnicas de Pré-processamento de Imagens	13
2.1 Pré-processamento	13
2.2 Segmentação	14
2.3 Técnicas para Extração de Características	14
2.3.1 <i>Undersampled Bitmaps</i>	14
3 Redes Neurais Artificiais para Detecção de Novidades	16
3.1 Redes MLP (<i>MultiLayer Perceptron</i>)	16
3.2 Redes RBFN (<i>Radial Basis Function Network</i>)	18
3.2.1 Algoritmo de treinamento: DDA (<i>Dynamic Decay Adjustment</i>)	20
3.3 Redes PNN (<i>Probabilistic Neural Network</i>)	23
4 Experimentos	25
4.1 Projeto dos Experimentos	25
4.2 Resultados Obtidos	26
4.2.1 Base Dados de Documentos Históricos	26
4.2.1.1 RBF-DDA	26
4.2.1.2 PNN-DDA	29
4.2.1.3 MLP	30
4.2.2 Base de Dados do MNIST	30
4.2.2.1 RBF-DDA	31
4.2.2.2 PNN-DDA	33
4.2.2.3 MLP	33
4.3 Análise dos Resultados	34
4.3.1 Base Dados de Documentos Históricos	34
4.3.1.1 RBF-DDA	34
4.3.1.2 PNN-DDA	38
4.3.1.3 MLP	39
4.3.2 Base de Dados do MNIST	40
4.3.2.1 RBF-DDA	40
4.3.2.2 PNN-DDA	44
4.3.2.3 MLP	44
4.4 Tempo de Treinamento dos Melhores Resultados	45
4.4.1 Base Dados de Documentos Históricos	46

4.4.2	Base de Dados do MNIST	46
5	Conclusões e Trabalhos Futuros	48
5.1	Contribuições	48
5.2	Estudo Comparativo	48
5.3	Trabalhos Futuros	49

Índice de Figuras

Figura 1. Imagem extraída de um documento histórico	11
Figura 2. Amostra de números	11
Figura 3. Exemplo de imagem extraída de um bloco intermediário	11
Figura 4. Pré-processamento: Limiarização	14
Figura 5. Redução da inclinação	14
Figura 6. Imagem dividida em blocos de 4x4	15
Figura 7. Matriz resultante depois de aplicado o <i>undersampled bitmaps</i>	15
Figura 8. Função sigmoïdal logística	16
Figura 9. Regiões definidas pelo processamento da segunda camada intermediária	17
Figura 10. Regiões definidas pelo processamento da camada de saída	17
Figura 11. Arquitetura de uma rede RBF	19
Figura 12. Exemplo de uma Rede RBF	19
Figura 13. Rede RBF junto com componente que detecta se a entrada é válida ou se é espúria	20
Figura 14. Exemplo de conflito de padrões	20
Figura 15. Exemplo da execução do algoritmo DDA	22
Figura 16. Estrutura típica de uma PNN	23
Figura 17. Comparação entre Resultados Obtidos para saídas diretas para o Método 1	35
Figura 18. Comparação entre Resultados Obtidos para saídas diretas para o Método 1	35
Figura 19. Comparação entre Resultados Obtidos para saídas diretas para o Método 2	36
Figura 20. Comparação entre Resultados Obtidos para saídas diretas para o Método 2	36
Figura 21. Comparação entre Resultados Obtidos para saídas probabilísticas para o Método 1	37
Figura 22. Comparação entre Resultados Obtidos para saídas probabilísticas para o Método 1	37
Figura 23. Comparação entre Resultados Obtidos para saídas probabilísticas para o Método 2	38
Figura 24. Comparação entre Resultados Obtidos para saídas probabilísticas para o Método 2	38
Figura 25. Resultado da rede PNN-DDA	39
Figura 26. Resultados Obtidos sem os dados negativos na base de treinamento	39

Figura 27. Resultados Obtidos com os dados negativos na base de treinamento	40
Figura 28. Comparação entre Resultados Obtidos para saídas diretas para o Método 1	40
Figura 29. Comparação entre Resultados Obtidos para saídas diretas para o Método 1	41
Figura 30. Comparação entre Resultados Obtidos para saídas diretas para o Método 2	41
Figura 31. Comparação entre Resultados Obtidos para saídas diretas para o Método 2	42
Figura 32. Comparação entre Resultados Obtidos para saídas probabilísticas para o Método 1	42
Figura 33. Comparação entre Resultados Obtidos para saídas probabilísticas para o Método 1	42
Figura 34. Comparação entre Resultados Obtidos para saídas probabilísticas para o Método 2	43
Figura 35. Comparação entre Resultados Obtidos para saídas probabilísticas para o Método 2	43
Figura 36. Resultado da rede PNN-DDA	44
Figura 37. Resultados Obtidos sem os dados negativos na base de treinamento	44
Figura 38. Resultados Obtidos com os dados negativos na base de treinamento	45

Índice de Tabelas

Tabela 1. $\theta^- = 0.1$ utilizando o Método 1	27
Tabela 2. $\theta^- = 0.1$ utilizando o Método 2	27
Tabela 3. $\theta^- = 0.01$ utilizando o Método 1	27
Tabela 4. $\theta^- = 0.01$ utilizando o Método 2	27
Tabela 5. $\theta^- = 0.1$ utilizando o Método 1	28
Tabela 6. $\theta^- = 0.1$ utilizando o Método 2	28
Tabela 7. $\theta^- = 0.01$ utilizando o Método 1	28
Tabela 8. $\theta^- = 0.01$ utilizando o Método 2	28
Tabela 9. $\theta^- = 0.001$ utilizando o Método 1	29
Tabela 10. $\theta^- = 0.001$ utilizando o Método 2	29
Tabela 11. Resultados do PNN-DDA	29
Tabela 12. Resultados do MLP sem os dados negativos na base de treinamento	30
Tabela 13. Resultados do MLP com os dados negativos na base de treinamento	30
Tabela 14. $\theta^- = 0.1$ utilizando o Método 1	31
Tabela 15. $\theta^- = 0.1$ utilizando o Método 2	31
Tabela 16. $\theta^- = 0.01$ utilizando o Método 1	31
Tabela 17. $\theta^- = 0.01$ utilizando o Método 2	32
Tabela 18. $\theta^- = 0.1$ utilizando o Método 1	32
Tabela 19. $\theta^- = 0.1$ utilizando o Método 2	32
Tabela 20. $\theta^- = 0.01$ utilizando o Método 1	32
Tabela 21. $\theta^- = 0.01$ utilizando o Método 2	33
Tabela 22. Resultados do PNN-DDA	33
Tabela 23. Resultados do MLP sem os dados negativos na base de treinamento	34
Tabela 24. Resultados do MLP com os dados negativos na base de treinamento	34
Tabela 25. Tempo de treinamento das redes utilizando a base dados de documentos históricos	46
Tabela 26. Tempo de treinamento das redes utilizando a base dados do MNIST	47

Tabela de Símbolos e Siglas

(Dispostos por ordem de aparição no texto)

TI – Tecnologia da Informação

dpi – dots per inch (pontos por polegada)

JPEG – Joint Photographic Experts Group

PCA – Principal Component Analysis

RBFN – Radial Basis Function Network

PNN – Probabilistic Neural Network

DDA – Dynamic Decay Adjustment

MLP – MultiLayer Perceptron

RNA – Rede Neural Artificial

RCE – Restricted Coulomb Energy

Agradecimentos

Quero aqui expressar os meus agradecimentos a um conjunto de pessoas que me ajudaram durante a minha formação, tornando-o um projeto de grande aprendizado, e também a Deus, por dar-me a vida e por guardar meus caminhos a cada dia, sempre me lembrando que tenho muito a aprender e a melhorar como profissional, estudante e ser humano.

Aos meus pais Adilberto e Isabel Mello, meus irmãos Marcus Vinicius e Thiago André e a minha namorada Cláudia Cadena pela compreensão e apoio incondicional em todos os momentos da minha vida, com quem pude compartilhar meus momentos de alegria e os momentos mais difíceis.

Ao meu orientador Adriano Lorena pela disponibilidade em atender minhas dúvidas quanto à elaboração da presente monografia, dentre outras questões.

À Thyago Rocha, agradeço a companhia e a ajuda prestada.

Aos amigos da Escola Politécnica de Pernambuco, pelos momentos de descontração e pelas demonstrações de amizade ao longo do curso de Engenharia da Computação.

Deixo ainda expresso os meus agradecimentos a todos os que participaram e que de alguma forma contribuíram para a realização desta monografia.

Capítulo 1

Introdução

Reconhecimento de escrita manuscrita é uma área de pesquisa bastante ativa. Apesar da disseminação do uso de Tecnologia da Informação (TI) nos últimos 20 anos, o uso de documentos manuscritos é ainda imprescindível em algumas áreas. Além de ser um campo de estudo de alguns dos principais problemas na área de reconhecimento de padrões [1]. Os avanços em termos de capacidade de processamento e de armazenamento tornaram viável a criação de soluções de reconhecimento automático de escrita manuscrita para as mais diversas finalidades – reconhecimento de cheques, reconhecimento de envelopes postais, indexação de documentos históricos, entre outras – de maneira cada vez mais condizente com a realidade da escrita humana (tolerante a variações de estilo, ruído, etc). Existem várias propostas em relação à estrutura de um sistema desse tipo na literatura, mas de uma forma geral, tais sistemas possuem quatro etapas principais: aquisição, pré-processamento, segmentação, representação e reconhecimento.

1.1 Definição do Problema

O trabalho tem o objetivo de reconhecer dígitos manuscritos extraídos de documentos históricos, mais especificamente as datas contidas nesses documentos com o propósito de ordenação por data. Nós apresentamos aqui alguns experimentos para reconhecimento desses dígitos.

Ao mesmo tempo em que buscamos reconhecer os dígitos, também buscamos reconhecer entradas espúrias, ou seja, dados que não representam números. Esses dados podem ser padrões intermediários. Capturamos as entradas a partir de imagens extraídas de documentos históricos. Um exemplo dessas imagens pode ser visto na Figura 1.

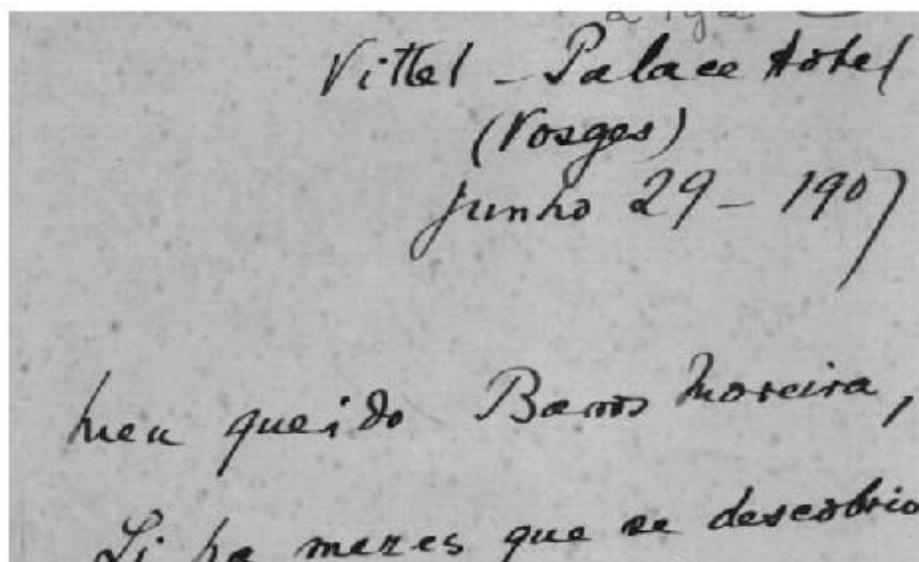


Figura 1. Imagem extraída de um documento histórico

As imagens que vão ser processadas e utilizadas no trabalho foram obtidas a partir do trabalho de Oliveira [2] (base de documentos históricos) e no site do *MNIST Database of Handwritten Digits* [3][4]. Para extração de características, utilizou-se o método de extração *undersampled bitmaps*. Maiores explicações sobre esses métodos podem ser encontradas na seção 2.

Neste trabalho, estamos lidando com números. Para extração desses números dos documentos, é necessário realizar um processo de segmentação de imagens para que as imagens utilizadas sejam divididas em blocos contendo apenas os números para posterior classificação. Em relação aos números, o processo de segmentação é fácil de ser aplicado para a maioria deles. Os números estão separados em sua maioria da forma como pode ser vista na Figura 2. Como pode ser visto, a área de cada número tem aproximadamente o mesmo tamanho, embora algumas áreas possam interceptar outras.



Figura 2. Amostra de números

Como dito anteriormente, temos o intuito de identificar entradas espúrias, ou seja, padrões que não podem ser classificados como pertencentes a nenhuma das classes treinadas. Dito isso, capturamos padrões intermediários que correspondem a blocos de imagens entre números. Um exemplo desse tipo de padrão pode ser visto na Figura 3.

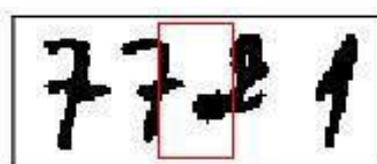


Figura 3. Exemplo de imagem extraída de um bloco intermediário

1.2 Objetivos

Este trabalho tem como objetivo reconhecer dígitos manuscritos extraídos de documentos históricos, mais especificamente as datas contidas nesses documentos com o propósito de ordenação por data. Ao mesmo tempo em que buscamos reconhecer os dígitos, também buscamos reconhecer entradas espúrias, ou seja, dados que não representam números. Esses dados podem ser padrões intermediários.

1.3 Organização

Esta monografia está organizada em cinco capítulos. Este capítulo apresenta uma breve apresentação do trabalho proposto. O Capítulo 2 apresenta algumas técnicas utilizadas para melhorar o desempenho dos sistemas de reconhecimento de dígitos. No Capítulo 3 abordaremos as Redes Neurais Artificiais utilizadas bem como a técnica utilizada para torná-la capaz de reconhecer novidades. O Capítulo 4 detalha os experimentos, relatando a questão do tratamento dos dados de entrada e a elaboração dos experimentos, bem como a análise dos resultados obtidos experimentalmente. E, por fim, o último Capítulo conclui esse trabalho e indica melhorias e trabalhos futuros.

Capítulo 2

Técnicas de Pré-processamento de Imagens

Antes de realizarmos os experimentos, foi preciso capturar dados para treinamento e teste para as Redes Neurais Artificiais (RNA's). Os dados utilizados no trabalho foram obtidos a partir do trabalho de Oliveira [2] (base de documentos históricos) e no site do *MNIST Database of Handwritten Digits* [3][4]. Com relação aos dados de dígitos manuscritos extraídos de documentos históricos, eles foram digitalizados em true color com uma resolução de 200 dpi e armazenados em formato JPEG por outro lado, as imagens dos dígitos manuscritos da base do MNIST encontravam-se pré-processadas, segmentadas, em tons de escala de cinza e armazenadas em formato JPEG.

2.1 Pré-processamento

Na etapa de pré-processamento realizam-se diversas operações para tratamento da imagem digitalizada, tais como realce de contraste, eliminação de ruídos, segmentação, limiarização, de maneira que as características da imagem possam ser realçadas. No presente estudo, as etapas de pré-processamento utilizadas nas bases de dados são compostas pelas técnicas de limiarização (*thresholding*) e rotação.

A limiarização consiste em converter a imagem original em níveis de cinza em uma imagem binária. Na imagem original, as variações de nível de cinza ocorrem em um intervalo que vai do preto (valor mínimo) ao branco (valor máximo). O extremo maior corresponde ao fundo branco do papel, enquanto que o extremo menor representa a imagem que está em primeiro plano. A tarefa de limiarização consiste em determinar o valor para um limiar de escala de cinza. Todos os valores abaixo deste limiar escolhido são mapeados na cor preto os valores acima são mapeados na cor branca. A Figura 4 apresenta a imagem original da base de dados e a imagem resultante do método de limiarização.



Figura 4. Pré-processamento: Limiarização

Na outra etapa é preciso eliminar ou pelo menos reduzir a inclinação da linha (rotação), como mostra a Figura 5 abaixo.

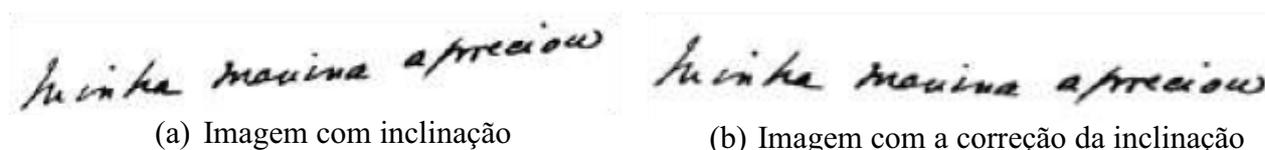


Figura 5. Redução da inclinação

2.2 Segmentação

Depois de feita a parte de pré-processamento da imagem, é feita a segmentação do texto, dividindo as imagens classificadas como texto em sub-imagens. Um exemplo disso foi mostrado na Figura 3. A base de dados utilizada em nosso trabalho foi obtida manualmente a partir da fase de segmentação de documentos e de texto. Posteriormente as imagens foram normalizadas em 32×32 pixels.

2.3 Técnicas para Extração de Características

Existem alguns métodos para a extração de características como por exemplo o *Principal Component Analysis* (PCA) e o *Undersampled Bitmaps*. Neste trabalho a técnica utilizada foi o *Undersampled Bitmap*.

2.3.1 Undersampled Bitmaps

O *Undersampled Bitmaps* foi aplicado em duas bases de dados, uma dos documentos históricos e outra do MNIST. As bases possuem caracteres diferentes, sendo a primeira com caracteres de tamanho 32×32 pixels e a segunda com caracteres de tamanho 28×28 pixels.

Undersampled bitmaps consiste em dividir imagens de 32×32 em blocos de 4×4 , como mostra a Figura 6, e em seguida é contado o número de pixels ativos em cada bloco, gerando uma matriz de entrada de 8×8 , onde cada elemento é um inteiro de 0 a 16. Dessa forma, temos 64 características por padrão a serem apresentadas para a rede neural, no caso da base de documentos históricos. No caso da base do MNIST depois de aplicada essa técnica, vamos obter 49 características por padrão a serem apresentadas para a rede neural.

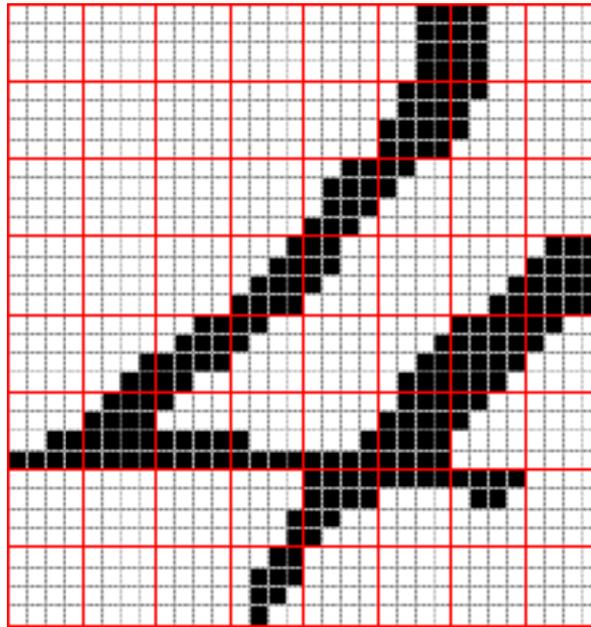


Figura 6. Imagem dividida em blocos de 4x4

0	0	0	0	0	8	8	0
0	0	0	0	0	14	4	0
0	0	0	0	11	3	0	0
0	0	0	10	5	0	3	15
0	3	11	3	0	6	15	3
6	15	9	5	5	15	3	0
0	0	0	2	11	4	6	0
0	0	0	8	0	0	0	0

Figura 7. Matriz resultante depois de aplicado o *undersampled bitmaps*

Capítulo 3

Redes Neurais Artificiais para Detecção de Novidades

Este capítulo visa explicar conceitos básicos de aprendizagem de máquinas relevantes para o trabalho. Abordaremos as técnicas de aprendizagem de máquina utilizadas, tais como: redes neurais de funções de bases radiais (RBF-DDA), probabilísticas (PNN-DDA) e perceptrons de multicamadas (MLP).

3.1 Redes MLP (*MultiLayer Perceptron*)

Para resolver problemas não-linearmente separáveis foram criadas as redes MLP, pois este tipo de rede possui pelo menos duas camadas permitindo a aproximação de qualquer função contínua. As redes MLP derivam de um modelo denominado *perceptron* proposto por Frank Rosenblatt em 1958 [7]. Com este modelo apenas problemas linearmente separáveis poderiam ser solucionados. A utilização de uma camada escondida aumentou o poder computacional das MLP. A precisão a ser obtida e a implementação da função objetivo dependem do número de neurônios utilizados nas camadas intermediárias.

Um dos principais aspectos das redes neurais artificiais é a utilização de uma função de ativação. Para redes MLP a mais empregada é a sigmoideal logística. Essa função é representada no plano cartesiano mostrado na Figura 8.

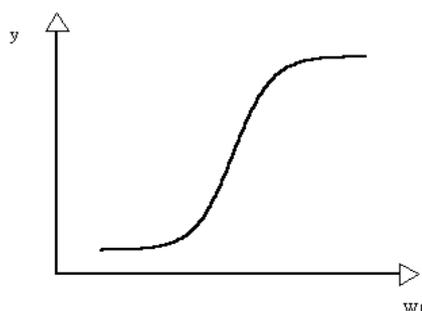


Figura 8. Função sigmoideal logística

Em uma rede multicamadas, o processamento realizado por cada nodo é definido pela combinação dos processamentos realizados pelos nodos da camada anterior, que estão conectados a ele. Para uma rede com duas camadas intermediárias pode-se dizer que o seguinte processamento ocorre em cada camada:

- Primeira camada intermediária: cada neurônio traça retas no espaço de padrões de treinamento;
- Segunda camada intermediária: cada neurônio combina as retas traçadas pelos nodos da camada anterior (primeira camada intermediária) conectados a ele, formando regiões convexas, onde o número de lados é definido pelo número de unidades conectadas a ele. Abaixo, a Figura 9, mostra um exemplo de uma região convexa:

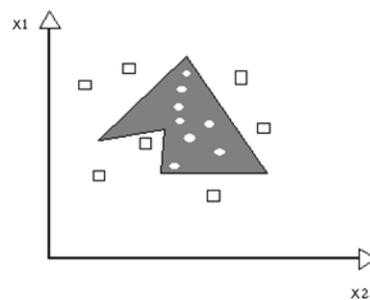


Figura 9. Regiões definidas pelo processamento da segunda camada intermediária

- Camada de saída: cada neurônio forma regiões que são combinações das regiões convexas definida pelos nodos conectados a ele da camada anterior (segunda camada intermediária). Na Figura 10 demonstramos um exemplo das combinações de regiões convexas.

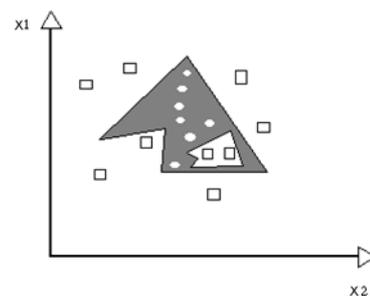


Figura 10. Regiões definidas pelo processamento da camada de saída

Pode-se dizer que as camadas intermediárias de uma MLP funcionam como detectores de características. Eles geram uma representação interna dos padrões de entrada, que é utilizada para definição da saída da rede. A camada de saída de uma MLP emprega uma técnica denominada *winner-takes-all* [7], com isso a classe fornecida pela rede corresponderá à maior saída da rede.

O número de neurônios contido em cada camada é definido empiricamente, esse número depende muito da distribuição dos padrões para treinamento e validação da rede. O número ideal de neurônios é influenciado por vários fatores, como:

- Número de exemplos de treinamento;

- Quantidade de ruído presente nos exemplos;
- Complexidade da função a ser aprendida;
- Distribuição estática dos dados de treinamento.

A alocação de unidades intermediárias (neurônios) deve ser suficiente para solucionar o sistema em questão. É preciso ter cuidado para não utilizar unidades demais, pois pode levar a rede a memorizar os padrões de treinamento, ao invés de extrair as características gerais que permitirão a generalização ou o reconhecimento de padrões que não fizeram parte do treinamento. Esse problema é denominado *overfitting*. Por outro lado, se utilizarmos poucos neurônios na camada intermediária, podemos ter *underfitting*, situação em que a rede não consegue aproximar adequadamente a função.

Uma das técnicas empregadas para evitar o *overfitting* é estimar o erro de generalização durante o processo de treinamento. Para isso a massa de dados é dividida em dois conjuntos: o de treinamento e o de validação. O conjunto de treinamento continua sendo utilizado na atualização dos pesos, enquanto que o conjunto de validação é empregado para estimar a capacidade de generalização da rede durante o processo de aprendizagem. O treinamento deve ser interrompido quando o erro de validação chegar ao seu mínimo global.

Neste trabalho foi utilizado o algoritmo de Levenberg-Marquardt [8], que é um algoritmo para ajustamento de comportamentos não-lineares, baseado no método dos mínimos-quadrados que procura encontrar o melhor ajustamento para um conjunto de dados, tentando minimizar a soma dos quadrados das diferenças entre a curva ajustada e os dados (tais diferenças são chamadas de resíduos).

3.2 Redes RBFN (*Radial Basis Function Network*)

São redes neurais que empregam funções de base radiais nos neurônios da camada escondida, RBFNs (*Radial Basis Functions*). RBF se diferencia da maioria das redes multicamadas por utilizar como argumento da função de ativação a distância entre seus vetores de entrada e de pesos, enquanto que MLPs utilizam o produto escalar do vetor de entrada e do vetor de peso para um nodo. Com a utilização dessa função na camada intermediária, RBFN é capaz de separar os padrões de classes distintas através de hiperelipsóides. As funções de bases radiais são representantes de uma classe de funções, cujo valor diminui ou aumenta em relação a um ponto central [5]. As mais comuns para serem empregadas em redes RBF são:

- Função Gaussiana: $f(u) = \exp\left(\frac{-v^2}{\sigma_i^2}\right)$ (2.1)

- Função Multiquadrática: $f(u) = \sqrt{v^2 + \sigma^2}$ (2.2)

- Função thin-plate-spline: $f(u) = v^2 \log(v)$ (2.3)

Onde $v = \|x - \mu\|$ é geralmente dado pela distância Euclidiana, x o vetor de entrada, e μ e σ representam respectivamente o centro e a largura da função radial.

A arquitetura de uma rede RBF costuma ter apenas uma camada intermediária, conforme mostrado na Figura 11. A camada intermediária utiliza funções de base radiais, agrupando os dados de entrada em *clusters*, com isso essa camada transforma um conjunto de padrões não-linearmente separáveis, ou seja, que não podem ser separados apenas traçando um plano ou uma reta, em um conjunto de padrões linearmente separáveis. A camada de saída classifica os padrões

recebidos da classe anterior. Podem ser utilizada redes do tipo *perceptron* ou *adaline* nesta camada, uma vez que seus padrões são linearmente separáveis.

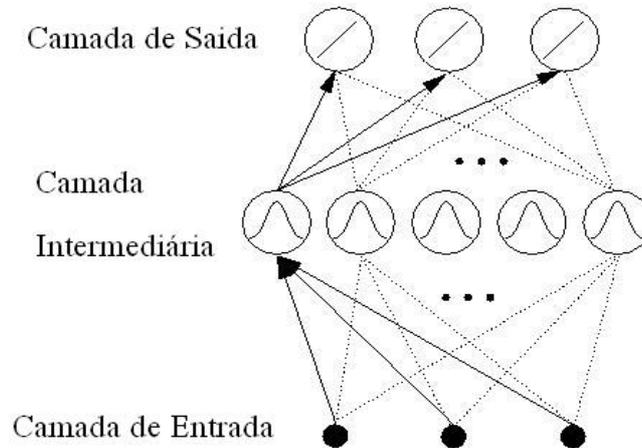


Figura 11. Arquitetura de uma rede RBF [15]

Quanto ao particionamento do espaço de entrada, as redes RBF, ao contrário das redes MLP (que empregam hiperplanos para fazer o particionamento), utilizam hiperelipsóides como mostra a Figura 12, onde cada um destes agrupa padrões de mesma classe formando regiões específicas para cada classe. Esse particionamento realizado pela rede RBF implica que a rede só poderá classificar novos padrões se esses forem das mesmas classes utilizadas para o seu treinamento, ou seja, se um determinado padrão pertencente à outra classe, que não tenha sido utilizada durante o treinamento, for apresentado à rede para ser classificado, a rede não saberá como classificar esse novo padrão [5].

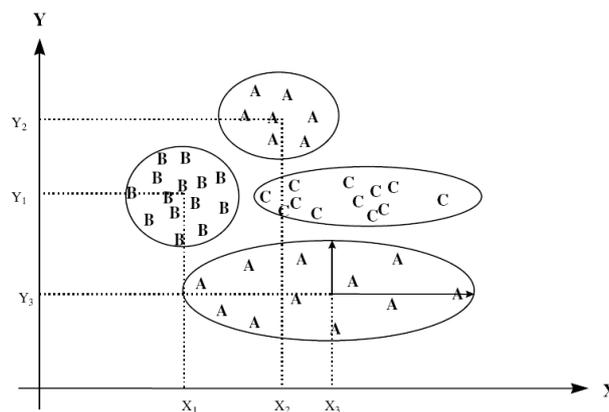


Figura 12. Exemplo de uma Rede RBF

Uma vantagem desse tipo de abordagem de RNA, é o fato de a rede não classificar um padrão discrepante, classificando-os como desconhecidos. Neste trabalho as redes RBF utilizam o algoritmo de treinamento DDA, que será descrito na subseção 3.2.1.

Para melhorar a detecção de novidades em redes RBF foi definido um critério de confiança (nível de confiança) para a aceitação da classificação. O nível de confiança é um coeficiente que impõe uma limitação para que as entradas só possam ser aceitas (classificadas em umas das classes conhecidas) quando a resposta da unidade de saída vencedora, for maior que todas as outras por uma certa margem, senão o padrão de entrada é rejeitado (ou seja,

“classificado” como desconhecido, ou novidade) [20]. Outro método utilizado foi o de a resposta da unidade de saída vencedora, for maior apenas que a segunda maior saída por uma certa margem. A Figura 13 exibe uma rede RBF aliada a um processamento da saída para detecção de novidades.

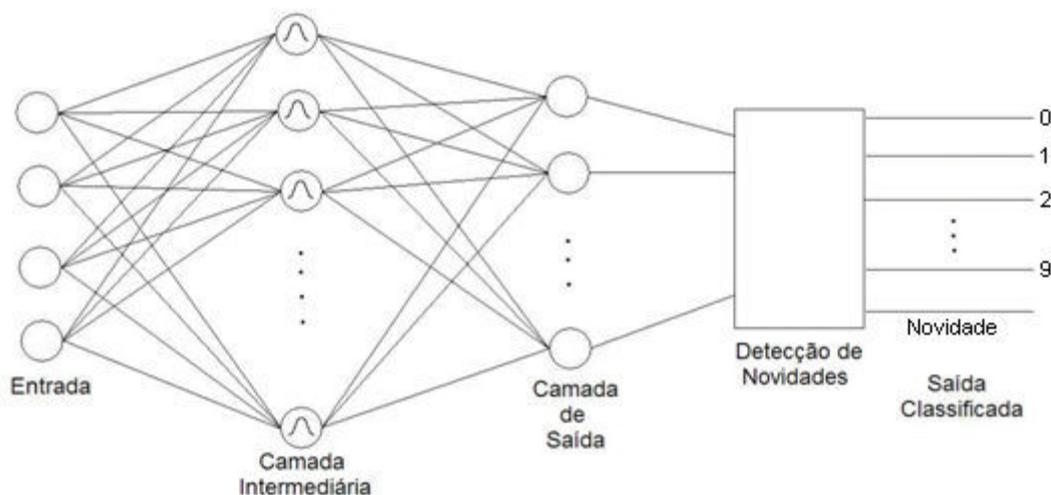


Figura 13. Rede RBF junto com componente que detecta se a entrada é válida ou se é espúria

3.2.1 Algoritmo de treinamento: DDA (*Dynamic Decay Adjustment*)

Existem vários métodos de treinamento para redes RBF. Neste trabalho vamos utilizar o algoritmo DDA (*Dynamic Decay Adjustment*) [9][10][11][12][13], um algoritmo que se baseia no algoritmo RCE (*Restricted Coulomb Energy*) [14]. O DDA corrige um problema do RCE que é o de se confundir em áreas de conflito, como ilustrado na Figura 14. Quando um padrão é apresentado à rede, ela cria uma unidade RBF para classificá-la. Se um padrão de uma outra classe for inserido no treinamento, irá também criar uma unidade RBF e isso pode gerar uma região comum às duas gaussianas, denominado área de conflito. O algoritmo RCE não trata desse caso e, por conseguinte pode não conseguir classificar corretamente esse padrão, podendo este ser atribuído a classe errada.

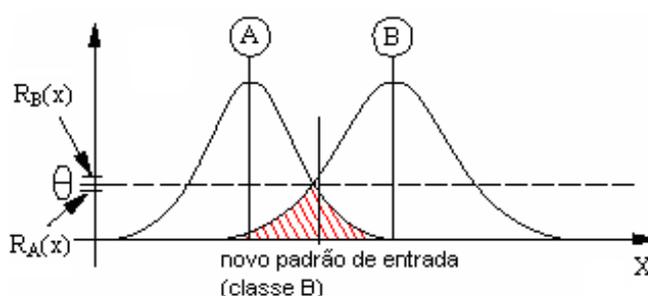


Figura 14. Exemplo de conflito de padrões [9]

O algoritmo DDA tem algumas peculiaridades que o diferenciam tanto na arquitetura usada por essas redes, como também no tempo utilizado para realização do treinamento. Quando

aplicado a uma rede RBF, os nós da camada escondida dessa rede utilizam funções gaussianas para processarem os valores de entrada [5].

O algoritmo DDA é um algoritmo construtivo, pois inicialmente é criada uma camada escondida sem neurônios. Cada gaussiana, apresentada na Figura 14, representa uma unidade na camada escondida de uma RBF, ou seja, um neurônio. A partir do momento que o treinamento ocorre, novas unidades vão sendo adicionadas dentro dessa camada escondida se houver a necessidade da inclusão de um novo neurônio, ou seja, a quantidade de neurônios que devem fazer parte da camada escondida não é determinada inicialmente, ela será determinada dinamicamente durante o treinamento.

O DDA utiliza dois parâmetros específicos para decidir se um novo neurônio deverá ser introduzido na camada intermediária da rede RBF [6]. O limiar positivo θ^+ é utilizado para verificar se, para um novo padrão usado para o treinamento da rede, existe algum protótipo (gaussiana) da mesma classe com ativação acima do θ^+ . Caso exista tal protótipo, não será adicionado novo protótipo à rede, ao invés disso o peso de um protótipo já existente será incrementado. O limiar negativo θ^- é utilizado para ajudar a solucionar o problema de conflitos de padrões que podem vir a existir durante o treinamento [9][11].

O algoritmo 1, a seguir, mostra um pseudocódigo para o DDA durante uma época de treinamento:

Algoritmo 1. Algoritmo DDA para treinamento de RBFs (uma época de treinamento)

```

//inicializa pesos com 0.0:
FORALL protótipos  $p_i^k$  DO
     $A_i^k = 0.0$ 
ENDFOR
//treina para uma época completa
FOR ALL padrão de treinamento (x, c) DO
    IF  $\exists p_i^c: R_i^c(x) \geq \theta^+$  THEN
         $A_i^c + = 1.0$ 
    ELSE
        // introduz um novo protótipo
        Adiciona um novo protótipo  $p_{mc+1}^c$  com:
         $r_{mc+1}^c = x$ 
         $\sigma_{mc+1}^c = \max_{k \neq c \wedge 1 \leq j \leq m_k} \{ \sigma: R_{mc+1}^c(r_j^k) < \theta \}$ 
         $A_{mc+1}^c = 1.0$ 
         $m_c + = 1$ 
    ENDIF
    //ajusta protótipos conflitantes
    FORALL  $k \neq c, 1 \leq j \leq m_k$  DO
         $\sigma_j^k = \max\{ \sigma: R_j^k(x) < \theta \}$ 
    ENDFOR
ENDFOR

```

Podemos verificar no algoritmo 1 que inicialmente todos os pesos recebem valor 0 para cada protótipo presente na rede. Depois da inicialização dos pesos, para cada padrão de treinamento, verifica-se o valor de sua ativação e esse valor é comparado com o parâmetro θ^+ . Se a ativação for maior ou igual ao θ^+ , não é adicionado novo protótipo à rede e o peso é incrementados de 1. Caso a ativação seja menor que o valor do parâmetro θ^- , um novo protótipo será adicionado à rede com o valor do seu centro igual ao vetor de entrada, o peso do novo protótipo será igual a 1 e o número de protótipos é incrementado de 1. O próximo passo é fazer os ajustes nos protótipos conflitantes da forma como descreve o algoritmo 1.

Temos a seguir, a Figura 15, que demonstra um exemplo de rede utilizando o algoritmo DDA. Na Figura 15 (a), é inserido um padrão de treinamento pertencente à classe A do problema

em questão, criando uma gaussiana. A seguir na Figura 15 (b), um padrão da classe B é inserido, portanto uma nova gaussiana terá que ser criada, e a gaussiana do padrão A terá que se ajustar para passar na intersecção do centro de B com o valor θ^- . No terceiro passo da Figura 15 (c), um novo padrão B é inserido, e como a intersecção do seu centro com a gaussiana do padrão B anterior é maior que o θ^+ , a gaussiana é incrementada, tendo seu peso passado para dois. No quarto passo da Figura 15 (d), um novo padrão A é inserido e a intersecção do centro desse novo padrão com a gaussiana A já existente é menor que o θ^+ , sendo assim uma nova gaussiana A teve que ser criada.

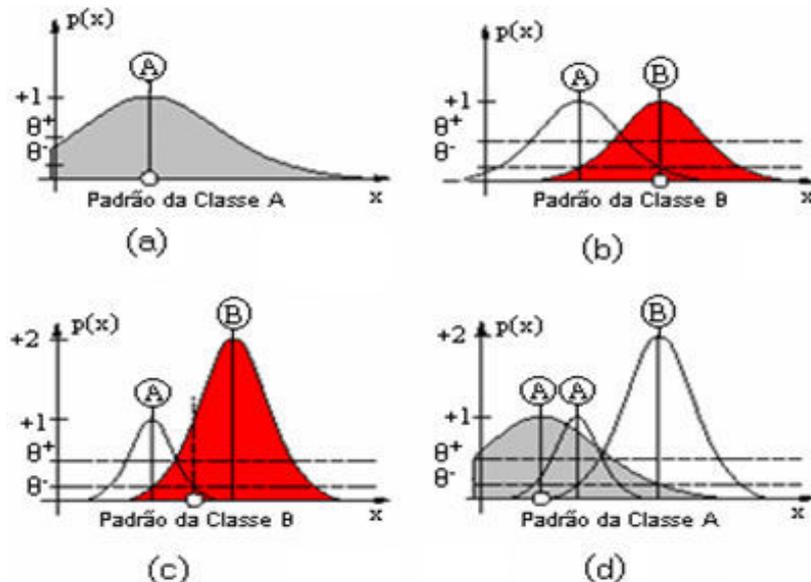


Figura 15. Exemplo da execução do algoritmo DDA [9]

Neste trabalho também utilizamos uma técnica de seleção do parâmetro θ^- [10][11][12]. O algoritmo DDA com seleção do parâmetro θ^- , propõe a utilização de valores menores que o *default* para o parâmetro θ^- , porém utiliza um método para selecionar um valor ótimo para esse parâmetro. O conjunto de dados contendo todos os padrões que serão utilizados pela rede neural RBF-DDA é dividido, inicialmente, em dois conjuntos: conjunto de treinamento e conjunto de teste. Realizada essa divisão inicial nos dados, em conjunto de treinamento e conjunto de teste, o conjunto de treinamento é novamente dividido em 2 novas partes: uma parte em um novo conjunto de treinamento, e outra parte é utilizada como um conjunto de validação. Esses dois novos conjuntos de dados serão usados para treinamento e validação da rede respectivamente, com o intuito de alcançar um valor para o parâmetro θ^- que seja ótimo, ou seja, um valor que proporcione uma taxa de generalização ideal para a rede, e essa possa classificar novos valores sem a presença de *overfitting*. O método para seleção do parâmetro θ^- usando o algoritmo DDA é mostrado no algoritmo 2.

Algoritmo 2. Algoritmo DDA com seleção de θ^- .

$$\theta_{\text{opt}}^- = \theta^- = 10^{-1}$$

Treinar uma RBF-DDA com θ^- usando o conjunto de treinamento reduzido e testa com o conjunto de validação para obter $ValError = MinValError$

REPEAT

$$\theta^- = \theta^- \times 10^{-1}$$

Treinar uma RBF-DDA com o θ^- usando o conjunto de treinamento reduzido e testar com o conjunto de validação para obter o $ValError$

IF $ValError < MinValError$

```

MinValError = ValError
θopt = θ̄
ENDIF
UNTIL ValError > MinValError OR θ̄ = 10-10
Treinar uma RBF-DDA com o θopt usando o conjunto de treinamento completo
Testar o RBF-DDA otimizado com o conjunto de teste

```

3.3 Redes PNN (*Probabilistic Neural Network*)

As Redes Neurais Probabilísticas (*Probabilistic Neural Networks – PNN*), foram criadas em 1990 por Specht [16]. PNN utiliza uma estrutura baseada em funções de base radial com um núcleo estatístico estimado [17]. O interesse por esse classificador é justificado pela característica de representar a estrutura da rede como uma função densidade de probabilidade e por ter obtido resultados superiores a outros classificadores [18]. Redes neurais probabilísticas computam a probabilidade condicional de classe $p(\text{class } k|x)$ para cada classe. A estrutura da PNN é mostrada na Figura 16.

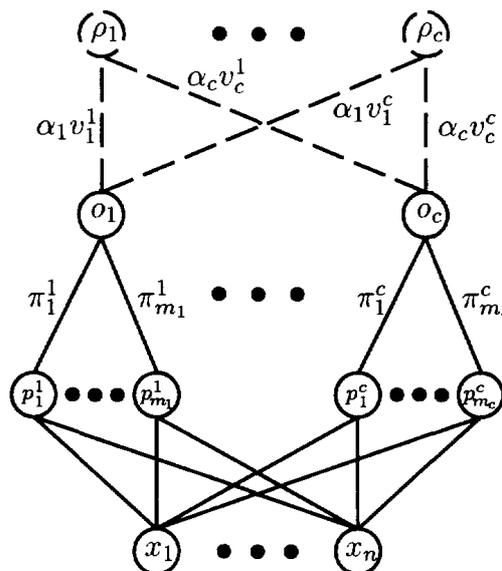


Figura 16. Estrutura típica de uma PNN [19]

A estrutura de uma PNN é similar a de uma RBF e possui:

- Um vetor de entrada n -dimensional $x=(x_1, \dots, x_n)$, é aplicado aos neurônios de entrada $x_i(1 \leq i \leq n)$;
- A primeira camada de saída que recebe a saída da camada de entrada, é composta de m_k Gaussiana que são computadas para cada classe $k(1 \leq k \leq c)$:

$$p_j^k = \frac{1}{\left((2\pi)^{\frac{n}{2}} \cdot |\Sigma_j^k|^{-\frac{1}{2}} \right)} \cdot \exp \left\{ -\frac{1}{2} \cdot (x - \mu_j^k)^T \cdot (\Sigma_j^k)^{-1} \cdot (x - \mu_j^k) \right\} \quad (2.4)$$

Onde, μ_j^k denota a media da distribuição e Σ_j^k indica a matriz de covariância.

- A segunda camada escondida calcula a aproximação da função de probabilidade da classe através da combinação de densidades multivariadas:

$$o_k(x) = \sum_{j=1}^{m_k} \pi_j^K p_j^k(x) \quad (2.5)$$

Onde, π_j^K representa a proporção máxima de mistura dentro da classe.

- A terceira camada é utilizada para computar a decisão.

Neste trabalho a rede PNN também utiliza o algoritmo de treinamento DDA, como descrito na seção 3.2.1, a diferença é que a saída da rede é probabilística.

Para a detecção de novidades na rede PNN-DDA, é acrescentado o $o_?$ que significa novidade, de tal forma que o $o_?$ é igual ao θ^- . Então a fórmula para normalizar cada uma das saídas é mostrada abaixo:

$$p(\text{class } k|x) = \frac{o_k(x)}{\sum_{l=1}^c o_l(x) + o_?} \text{ e } p(?|x) = \frac{o_?}{\sum_{l=1}^c o_l(x) + o_?} \quad (2.6)$$

Depois de feita a normalização de cada uma das saídas, verifica-se qual a saída que tem a maior probabilidade, e a classifica como dígito ou novidade.

Capítulo 4

Experimentos

Este capítulo visa explicar o projeto dos experimentos, bem como os resultados obtidos, uma análise dos resultados obtidos e por fim o tempo de treinamento dos melhores resultados.

Vale salientar que o objetivo é que ao mesmo tempo em que se segmenta a string de números, a rede detecta as novidades e reconhece os dígitos.

4.1 Projeto dos Experimentos

Foram realizados experimentos utilizando diferentes configurações para cada técnica e foram analisados os impactos dessas alterações, tendo como parâmetros de desempenho: a taxa de erro para a classificação de dígitos, a taxa para o reconhecimento de novidades, a complexidade da rede e o número de épocas para o treinamento.

Para as RBF variamos apenas o valor do θ^- , que como visto no trabalho de Oliveira [10], pode ter influência, no desempenho da rede neural em determinados problemas. Utilizamos o θ^- com os valores 0,1; 0,1; 0,001 e 0,0001.

Outra variável analisada neste trabalho foi o nível ou limiar de confiança. O nível de confiança é utilizado para verificar a validade de uma entrada, ou seja, se ela pode ser classificada como dígito ou se deve ser considerada novidade. Para o limiar de confiança, realizamos experimentos com 0,1; 0,15; 0,25; 0,35; 0,45 e 0,55.

Aplicamos dois métodos de verificação de aceitação para uma decisão de classificação, explanados na seção 3.2. O primeiro método, chamado de método 1, consiste em selecionar o maior valor fornecido pela rede, após a classificação de uma dada entrada, e comparar com a soma de todos os outros valores. Se a diferença entre a maior saída e a soma de todas as outras for maior que o limiar de confiança, a entrada é classificada como dígito e é dito a qual classe ela pertence (levando em conta que a maior saída é a saída esperada, senão a rede comete um erro de classificação). Caso contrário, a entrada é classificada como novidade, ou seja, não é um número, trata-se de uma região intermediária entre dois números.

O segundo método, denominado método 2, segue um mecanismo semelhante. No entanto, ao invés de comparar a maior saída com a soma de todas as outras, comparamos a maior saída com a segunda maior fornecida (também consideramos apenas o caso da maior saída fornecida pela rede ser a classe desejada como resposta). Se a diferença for maior que o limiar de confiança, o dígito é classificado, senão a entrada é considerada novidade.

Por último, consideramos dois tipos de saída fornecidos pela rede RBF. Realizamos a verificação dos métodos 1 e 2 sobre saídas diretas e sobre saídas probabilísticas (a soma das saídas é igual a 1).

O treinamento com PNN-DDA teve como alteração de parâmetro apenas o valor do parâmetro negativo θ^- , os valores assumidos por esse parâmetro foram: 0,1; 0,01; 0,001; 0,0001; 0,00001; 0,000001 até 10^{-18} .

O terceiro e último treinamento com o MLP teve como alteração de parâmetro o número de neurônios escondidos que foi de 10, a quantidade de épocas que foi de 30, 60, 80, 90 e 120 e por fim o limiar que foi de 0,2; 0,3 e 0,4.

4.2 Resultados Obtidos

Nesta seção mostrarei os resultados obtidos do treinamento com as diferentes bases utilizadas, sendo uma das bases de documentos históricos e outra da base do MNIST, como já foi explanado na seção 1.1.

4.2.1 Base Dados de Documentos Históricos

Foram extraídas 20 imagens para representar cada dígito de 0 a 9, resultando num arquivo de treinamento com 200 padrões de entradas e 10 classes possíveis. Como base de testes, utilizou-se mais 10 imagens para cada dígito de 0 a 9, diferentes das utilizadas na etapa de treinamento, para verificar a capacidade de reconhecimento das redes RBF, PNN e MLP. Também foram utilizadas 100 imagens representando *outliers*, ou seja, padrões que não podem ser classificados como números. Esses *outliers* representam novidades, ou entradas espúrias. Um exemplo dessas imagens já foi exibido na Figura 3. Resumindo, obtivemos uma base de testes com 400 padrões.

Com a rede MLP foi criada mais uma base de treinamento contendo dados espúrios, o qual tem 20 imagens para representar cada dígito de 0 a 9, e mais 200 imagens representando *outliers* diferentes da base de teste, totalizando no final 400 padrões.

4.2.1.1 RBF-DDA

Primeiramente realizou-se experimentos considerando saídas diretas geradas pelas Redes RBF. Obteve-se resultados para os dois métodos. Os resultados considerados são o percentual total de acerto de classificação da rede RBF, incluindo a capacidade de detecção de entradas espúrias, o percentual de acerto das classes utilizadas (no caso os dígitos) no treinamento, e a capacidade de detecção de novidade da rede neural em porcentagem.

A Tabela 1 exibe os resultados dos experimentos utilizando o θ^- igual a 0,1 e para o método 1, variando o valor do limiar de confiança. Como θ^- igual a 0,1; o DDA encontrou uma rede em 4 épocas e obteve 162 neurônios escondidos.

Tabela 1. $\theta^- = 0,1$ utilizando o Método 1

limiar	acerto total	acerto classes	% de novidades
0,1	62,5%	25%	100%
0,15	61,5%	23%	100%
0,25	59,5%	19%	100%
0,35	58,5%	17%	100%
0,45	57,5%	15%	100%
0,55	57%	14%	100%

Tabela 2. $\theta^- = 0,1$ utilizando o Método 2

limiar	acerto total	acerto classes	% de novidades
0,1	73,5%	81%	66%
0,15	78%	78,5%	77,5%
0,25	81%	74,5%	87,5%
0,35	82,75%	72%	93,5%
0,45	82,25%	67%	97,5%
0,55	80,75%	63,5%	98%

A Tabela 2 exibe os resultados dos experimentos com o método 2, com a mesma configuração de rede RBF da Tabela 1.

Utilizando o θ^- igual a 0,01, obtivemos 195 neurônios escondidos e a rede neural foi treinada em 3 épocas. A Tabela 3 e a Tabela 4 exibem os resultados para os métodos 1 e 2.

Tabela 3. $\theta^- = 0,01$ utilizando o Método 1

limiar	acerto total	acerto classes	% de novidades
0,1	79,25%	59,5%	99%
0,15	75%	50%	100%
0,25	68,75%	37,5%	100%
0,35	64%	28%	100%
0,45	60%	20%	100%
0,55	57%	14%	100%

Tabela 4. $\theta^- = 0,01$ utilizando o Método 2

limiar	acerto total	acerto classes	% de novidades
0,1	81,75%	65%	98,5%
0,15	77%	54%	99,5%
0,25	70,25%	40,5%	100%
0,35	64,75%	29,5%	100%
0,45	61%	22%	100%
0,55	57%	16%	100%

O resultado do treinamento com o θ^- igual a 0,001; 0,0001 e 0,00001 encontram-se no anexo.

Por último, realizei experimentos considerando saídas probabilísticas geradas pela rede RBF, para os dois métodos. A Tabela 5, Tabela 6, Tabela 7, Tabela 8, Tabela 9 e a Tabela 10 exibem os resultados obtidos com a aplicação dos métodos 1 e 2.

Tabela 5. $\theta^- = 0,1$ utilizando o Método 1

limiar	acerto total	acerto classes	% de novidades
0,1	58,75%	17,50%	100,00%
0,15	56,50%	13,00%	100,00%
0,25	53,75%	7,50%	100,00%
0,35	51,75%	3,50%	100,00%
0,45	50,00%	0,00%	100,00%
0,55	50,00%	0,00%	100,00%

Tabela 6. $\theta^- = 0,1$ utilizando o Método 2

limiar	acerto total	acerto classes	% de novidades
0,1	72,00%	72,00%	72,00%
0,15	75,75%	65,50%	86,00%
0,25	74,25%	51,50%	97,00%
0,35	64,00%	28,00%	100,00%
0,45	56,50%	13,00%	100,00%
0,55	53,25%	6,50%	100,00%

Tabela 7. $\theta^- = 0,01$ utilizando o Método 1

limiar	acerto total	acerto classes	% de novidades
0,1	80,75%	79,50%	82,00%
0,15	80,50%	77,50%	83,50%
0,25	80,25%	72,50%	88,00%
0,35	77,50%	64,00%	91,00%
0,45	75,50%	57,99%	93,00%
0,55	73,00%	50,00%	96,00%

Tabela 8. $\theta^- = 0,01$ utilizando o Método 2

limiar	acerto total	acerto classes	% de novidades
0,1	56,50%	85,00%	28,00%
0,15	63,00%	84,00%	42,00%
0,25	72,75%	81,50%	64,00%
0,35	78,75%	77,00%	80,50%
0,45	79,00%	70,00%	88,00%
0,55	76,75%	62,00%	91,50%

Tabela 9. $\theta^- = 0,001$ utilizando o Método 1

limiar	acerto total	acerto classes	% de novidades
0,1	70,75%	92,50%	49,00%
0,15	72,00%	91,00%	53,00%
0,25	75,00%	90,50%	59,50%
0,35	79,25%	86,50%	72,00%
0,45	79,75%	80,50%	79,00%
0,55	81,75%	78,00%	85,50%

Tabela 10. $\theta^- = 0,001$ utilizando o Método 2

limiar	acerto total	acerto classes	% de novidades
0,1	54,00%	86,00%	22,00%
0,15	57,75%	86,00%	29,50%
0,25	62,25%	84,50%	40,00%
0,35	66,75%	84,00%	49,50%
0,45	73,00%	81,50%	64,50%
0,55	76,75%	79,00%	74,50%

Também omitiremos os resultados encontrados utilizando o θ^- igual a 0,0001 e 0,00001, pois observamos que a capacidade de detecção de novidades da rede continua a diminuir.

4.2.1.2 PNN-DDA

Os resultados obtidos com o PNN-DDA estão demonstrados na Tabela 11. Os resultados considerados são o percentual total de acerto de classificação da rede PNN, incluindo a capacidade de detecção de entradas espúrias, o percentual de acerto das classes utilizadas no treinamento, e a capacidade de detecção de novidade da rede neural em porcentagem. Omitiremos os resultados encontrados utilizando o θ^- igual 10^{-10} até 10^{-18} , pois observamos que a capacidade de detecção de novidades da rede continua a diminuir.

Tabela 11. Resultados do PNN-DDA

theta-	acerto total	acerto classes	% de novidades	qtd. de neurônios	épocas
10^{-1}	66,75%	33,5%	100%	162	3
10^{-2}	75,25%	52%	98,5%	195	3
10^{-3}	83,75%	69,5%	98%	199	3
10^{-4}	84,5%	72%	97%	199	3
10^{-5}	86%	77%	95%	200	2
10^{-6}	85,25%	78,5%	92%	200	2
10^{-7}	85%	79,5%	90,5%	200	2
10^{-8}	82,75%	79,5%	86%	200	2
10^{-9}	82,75%	80,5%	85%	200	2

4.2.1.3 MLP

Primeiramente realizou-se experimentos sem os dados negativos na base de treinamento. Os resultados considerados são o percentual total de acerto de classificação da rede MLP, incluindo a capacidade de detecção de entradas espúrias, o percentual de acerto das classes utilizadas no treinamento, e a capacidade de detecção de novidade da rede neural em porcentagem.

A Tabela 12 exibe os resultados dos experimentos utilizando 10 neurônios na camada escondida, variando o valor do limiar e o número de épocas.

Tabela 12. Resultados do MLP sem os dados negativos na base de treinamento

épocas	limiar	acerto total	acerto classes	% de novidades
30	0,2	37,5%	64%	11%
30	0,3	33,75%	32%	35,5%
30	0,4	41,5%	47%	36%
80	0,2	37,75%	52%	23,5%
80	0,3	52,5%	74,5%	30,5%
80	0,4	43,5%	59%	28%
120	0,2	42,5%	68,5%	16,5%
120	0,3	40,75%	66%	15,5%
120	0,4	40%	52,5%	27,5%

A Tabela 13 exibe os resultados dos experimentos com os dados negativos na base de treinamento, utilizando 10 neurônios na camada escondida, variando o valor do limiar e o número de épocas.

Tabela 13. Resultados do MLP com os dados negativos na base de treinamento

épocas	limiar	acerto total	acerto classes	% de novidades
30	0,2	53%	50%	56%
30	0,3	50,5%	45%	56%
30	0,4	56%	37%	75%
80	0,2	39,5%	36,5%	42,5%
80	0,3	49,25%	43,5%	55%
80	0,4	54,25%	36%	72,5%
120	0,2	41,25%	38%	44,5%
120	0,3	47,5%	46,5%	48,5%
120	0,4	46,75%	46%	47,5%

4.2.2 Base de Dados do MNIST

Foram extraídas 200 imagens para representar cada dígito de 0 a 9, resultando num arquivo de treinamento com 2000 padrões de entradas e 10 classes possíveis. Como base de testes, utilizou-se mais 100 imagens para cada dígito de 0 a 9, diferentes das utilizadas na etapa de treinamento, para verificar a capacidade de reconhecimento das redes RBF, PNN e MLP. Também foram utilizadas 1000 imagens representando *outliers*, ou seja, padrões que não podem ser classificados como números. Esses *outliers* representam novidades, ou entradas espúrias. Resumindo, obtivemos uma base de testes com 2000 padrões.

Com a rede MLP foi criada mais uma base de treinamento contendo dados espúrios, o qual tem 100 imagens para representar cada dígito de 0 a 9, e mais 1000 imagens representando *outliers* diferentes da base de teste, totalizando no final com 2000 padrões.

4.2.2.1 RBF-DDA

Primeiramente realizou-se experimentos considerando saídas diretas geradas pelas Redes RBF. Obtivemos resultados para os dois métodos. Os resultados considerados são o percentual total de acerto de classificação da rede RBF, incluindo a capacidade de detecção de entradas espúrias, o percentual de acerto das classes utilizadas no treinamento, e a capacidade de detecção de novidade da rede neural em porcentagem.

A Tabela 14 exhibe os resultados dos experimentos utilizando o θ^- igual a 0,1 e para o método 1, variando o valor do limiar de confiança. Como θ^- igual a 0,1; o DDA encontrou uma rede em 4 épocas e obteve 1548 neurônios escondidos.

Tabela 14. $\theta^- = 0,1$ utilizando o Método 1

limiar	acerto total	acerto classes	% de novidades
0,1	74,35%	48,69%	100,00%
0,15	74,10%	48,19%	100,00%
0,25	73,35%	46,70%	100,00%
0,35	73,05%	46,10%	100,00%
0,45	72,50%	45,00%	100,00%
0,55	72,20%	44,40%	100,00%

Tabela 15. $\theta^- = 0,1$ utilizando o Método 2

limiar	acerto total	acerto classes	% de novidades
0,1	89,95%	80,90%	99,00%
0,15	89,75%	79,70%	99,80%
0,25	88,90%	77,80%	100,00%
0,35	87,80%	75,60%	100,00%
0,45	87,10%	74,20%	100,00%
0,55	86,60%	73,20%	100,00%

A Tabela 15 exhibe os resultados dos experimentos com o método 2, com a mesma configuração de rede RBF da Tabela 14.

Utilizando o θ^- igual a 0,01; obtivemos 1892 neurônios escondidos e a rede neural foi treinada em 3 épocas. A Tabela 16 e a Tabela 17 exibem os resultados para os métodos 1 e 2.

Tabela 16. $\theta^- = 0,01$ utilizando o Método 1

limiar	acerto total	acerto classes	% de novidades
0,1	84,75%	69,50%	100,00%
0,15	82,15%	64,30%	100,00%
0,25	77,30%	54,60%	100,00%
0,35	74,20%	48,40%	100,00%
0,45	71,20%	42,40%	100,00%
0,55	69,25%	38,50%	100,00%

Tabela 17. $\theta^- = 0,01$ utilizando o Método 2

limiar	acerto total	acerto classes	% de novidades
0,1	86,80%	73,60%	100,00%
0,15	83,80%	67,60%	100,00%
0,25	79,20%	58,40%	100,00%
0,35	75,50%	51,00%	100,00%
0,45	72,50%	45,00%	100,00%
0,55	70,60%	41,19%	100,00%

O resultado do treinamento com o θ^- igual a 0,001; 0,0001 e 0,00001 encontram-se no anexo.

Por último, realizei experimentos considerando saídas probabilísticas geradas pela rede RBF, para os dois métodos. As Tabela 18, Tabela 19, Tabela 20 e Tabela 21 exibem os resultados obtidos com a aplicação dos métodos 1 e 2.

Tabela 18. $\theta^- = 0,1$ utilizando o Método 1

limiar	acerto total	acerto classes	% de novidades
0,1	67,90%	43,20%	92,60%
0,15	67,50%	40,50%	94,50%
0,25	65,90%	34,80%	97,00%
0,35	63,60%	28,79%	98,40%
0,45	61,25%	23,59%	98,90%
0,55	59,10%	18,60%	99,60%

Tabela 19. $\theta^- = 0,1$ utilizando o Método 2

limiar	acerto total	acerto classes	% de novidades
0,1	59,30%	71,50%	47,09%
0,15	64,65%	66,90%	62,40%
0,25	67,80%	57,40%	78,20%
0,35	68,15%	47,50%	88,80%
0,45	67,20%	39,20%	95,19%
0,55	64,25%	30,40%	98,10%

Tabela 20. $\theta^- = 0,01$ utilizando o Método 1

limiar	acerto total	acerto classes	% de novidades
0,1	76,95%	80,80%	73,10%
0,15	78,75%	79,30%	78,20%
0,25	80,25%	75,00%	85,50%
0,35	81,60%	71,39%	91,80%
0,45	80,30%	65,60%	95,00%
0,55	78,74%	60,09%	97,39%

Tabela 21. $\theta^- = 0,01$ utilizando o Método 2

limiar	acerto total	acerto classes	% de novidades
0,1	57,10%	86,10%	28,10%
0,15	62,10%	84,39%	39,80%
0,25	70,05%	80,50%	59,59%
0,35	75,95%	76,80%	75,10%
0,45	79,10%	72,20%	86,00%
0,55	80,40%	67,50%	93,30%

Também omitiremos os resultados encontrados utilizando o θ^- igual a 0,001; 0,0001 e 0,00001, pois o mesmo se encontra no anexo.

4.2.2.2 PNN-DDA

Os resultados obtidos com o PNN-DDA estão demonstrados na Tabela 22. Os resultados considerados são o percentual total de acerto de classificação da rede PNN, incluindo a capacidade de detecção de entradas espúrias, o percentual de acerto das classes utilizadas no treinamento, e a capacidade de detecção de novidade da rede neural em porcentagem.

Tabela 22. Resultados do PNN-DDA

theta-	acerto total	acerto classes	% de novidades	qtd. de neurônios	épocas
10^{-1}	51,50%	3,00%	100,00%	1548	4
10^{-2}	57,50%	15,00%	100,00%	1892	3
10^{-3}	65,75%	31,50%	100,00%	1948	3
10^{-4}	73,75%	47,50%	100,00%	1971	2
10^{-5}	80,00%	59,99%	100,00%	1981	3
10^{-6}	83,10%	66,20%	100,00%	1988	2
10^{-7}	85,40%	70,80%	100,00%	1989	2
10^{-8}	87,20%	74,40%	100,00%	1990	2
10^{-9}	88,05%	76,10%	100,00%	1994	2
10^{-10}	89,05%	78,10%	100,00%	1996	2
10^{-11}	89,85%	79,70%	100,00%	1996	2
10^{-12}	90,45%	80,90%	100,00%	1996	2
10^{-13}	90,80%	81,60%	100,00%	1996	2
10^{-14}	91,20%	82,39%	100,00%	1996	2
10^{-15}	91,50%	83,00%	100,00%	1996	2
10^{-16}	91,65%	83,30%	100,00%	1996	2
10^{-17}	91,75%	83,50%	100,00%	1996	2
10^{-18}	91,80%	83,60%	100,00%	1996	2

4.2.2.3 MLP

Primeiramente realizamos experimentos sem os dados negativos na base de treinamento. Os resultados considerados são o percentual total de acerto de classificação da rede MLP, incluindo a capacidade de detecção de entradas espúrias, o percentual de acerto das classes utilizadas no treinamento, e a capacidade de detecção de novidade da rede neural em porcentagem.

A Tabela 23 exibe os resultados dos experimentos utilizando 10 neurônios na camada escondida, variando o valor do limiar e o número de épocas.

Tabela 23. Resultados do MLP sem os dados negativos na base de treinamento

épocas	limiar	acerto total	acerto classes	% de novidades
30	0,2	47,55%	55,3%	39,8%
30	0,3	40,25%	71,4%	9,1%
30	0,4	67,5%	73,1%	61,9%
60	0,2	43,45%	61,1%	25,8%
60	0,3	45,75%	79,5%	12%
60	0,4	50,1%	71,3%	28,9%
90	0,2	60,55%	70,9%	50,2%
90	0,3	51,8%	77,4%	26,2%
90	0,4	54,3%	68,8%	39,8%

A Tabela 24 exibe os resultados dos experimentos com os dados negativos na base de treinamento, utilizando 10 neurônios na camada escondida, variando o valor do limiar e o número de épocas.

Tabela 24. Resultados do MLP com os dados negativos na base de treinamento

épocas	limiar	acerto total	acerto classes	% de novidades
30	0,2	69,65%	48,5%	90,8%
30	0,3	75,9%	59,7%	92,1%
30	0,4	82,7%	69,9%	95,5%
60	0,2	81,15%	65%	97,3%
60	0,3	82,15%	69,3%	95%
60	0,4	74,6%	57,8%	91,4%
90	0,2	77,2%	49,2%	96%
90	0,3	86,45%	73,7%	99,2%
90	0,4	62,5%	33,3%	91,7%

4.3 Análise dos Resultados

4.3.1 Base Dados de Documentos Históricos

4.3.1.1 RBF-DDA

A Figura 17 exibe o desempenho total da rede neural, ou seja, sua capacidade de reconhecer dígitos e ao mesmo tempo detectar padrões espúrios, treinada com θ^- igual a 0,1 e utilizando o método 1 para detecção de novidades em função do valor do limiar de confiança. Além disso, também exibe o acerto da rede em classificar números e em detectar novidades. Lembrando que a Figura 17 exibe um gráfico levando em consideração as saídas diretas da rede RBF.

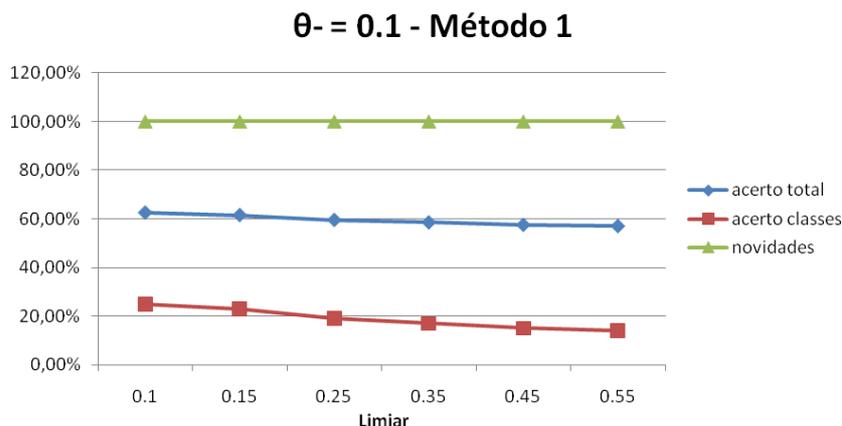


Figura 17. Comparação entre Resultados Obtidos para saídas diretas para o Método 1

Observamos que com a combinação do θ^- igual a 0,1 e utilizando o método 1, a rede não apresenta um desempenho satisfatório, pois apesar de detectar todos os padrões espúrios, seu desempenho de classificação é muito pequeno.

Com o mesmo método e com θ^- igual a 0,01; a rede já obtém melhores resultados. Para o limiar de confiança igual a 0,1; obteve um desempenho geral de 79,25%, como visto na Tabela 3, tendo com isso 59,50% de acerto no reconhecimento dos dígitos. Como ilustrado na Figura 18.

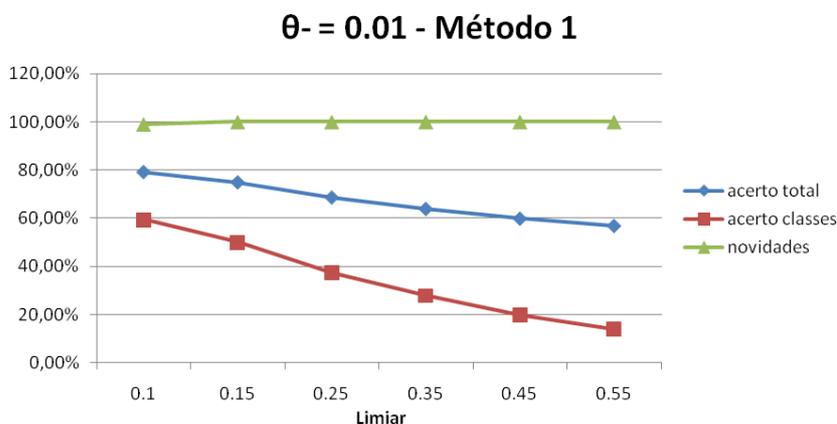


Figura 18. Comparação entre Resultados Obtidos para saídas diretas para o Método 1

Os resultados acima foram os melhores obtidos para saídas diretas da rede e com o método 1, e foi obtido com um valor de limiar de confiança muito baixo. Com a diminuição do θ^- , a tendência é da rede perder cada vez mais sua capacidade de classificar dígitos, considerando todas as entradas como espúrias.

A Figura 19 e a Figura 20 exibem os resultados da rede neural em função do valor do limiar de confiança, considerando saídas diretas da rede neural. Para saídas diretas o método 2, teve um desempenho melhor que o método 1. Para o θ^- igual 0,1 e limiar 0,15; obteve acerto de classificação total de 78%, como visto na Tabela 2, sendo 78,5% de acerto de classificação de dígitos e 77,5% de detecção de entradas espúrias. Para o limiar igual a 0,35, o desempenho total da rede foi de 82,75% de acerto, tendo 72% de classificações corretas e 93,5% de detecção de espúrios.

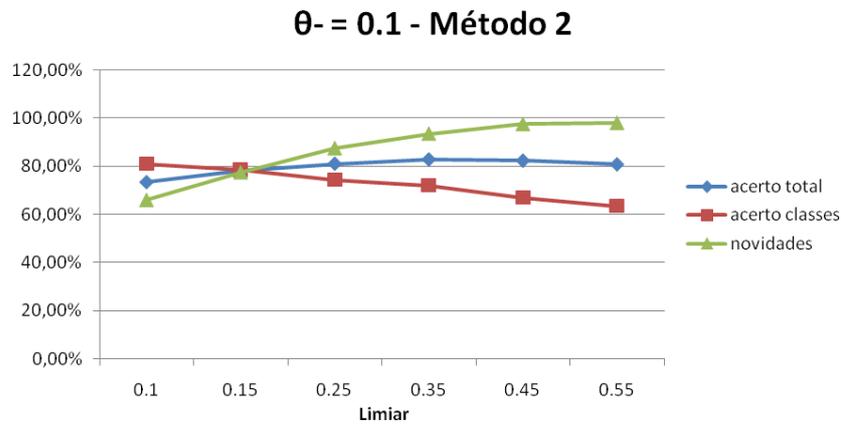


Figura 19. Comparação entre Resultados Obtidos para saídas diretas para o Método 2

Com o θ^- igual a 0,01; o método 2, já não obteve um desempenho tão bom. Para um limiar de confiança de 0,1; obteve 81,75% de acerto total, com 65% de classificações corretas para dígitos e 98,5% para detecção de novidades, como visto na Tabela 4. Para valores menores de θ^- , a rede neural perde cada vez mais sua capacidade de classificação, considerando todas as entradas como espúrias. A Figura 20 exibe uma tendência de diminuição de acerto de classificação de dígitos com o aumento do limiar de confiança.

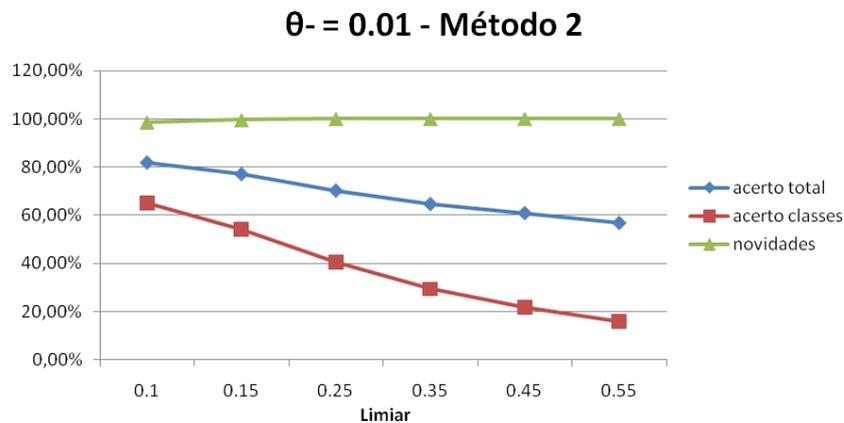


Figura 20. Comparação entre Resultados Obtidos para saídas diretas para o Método 2

Em relação às saídas probabilísticas, o método 1 não obteve um bom resultado, considerando o θ^- igual a 0,1. Observe a Figura 21.

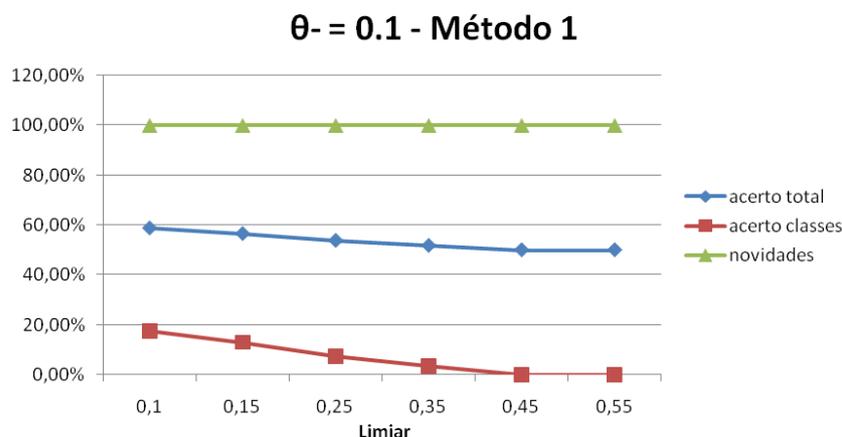


Figura 21. Comparação entre Resultados Obtidos para saídas probabilísticas para o Método 1

Entretanto com o θ^- igual a 0,01; a rede obteve um acerto total de 80,75%, classificando 79,5% de dígitos corretos e detectando 82% das entradas espúrias, como visto na Tabela 7. No entanto esse valor foi obtido para um limiar de confiança baixo. Com o aumento do limiar a capacidade de classificação de dígitos tende a diminuir, como exibe o gráfico da Figura 22.

Observamos também, que utilizando saídas probabilísticas, e havendo a diminuição do θ^- , a tendência é de aumentar a capacidade de classificação correta dos dígitos. Entretanto a rede perde cada vez mais sua capacidade de detectar novidades. Essa tendência é observada na Tabela 7 e na Tabela 9.

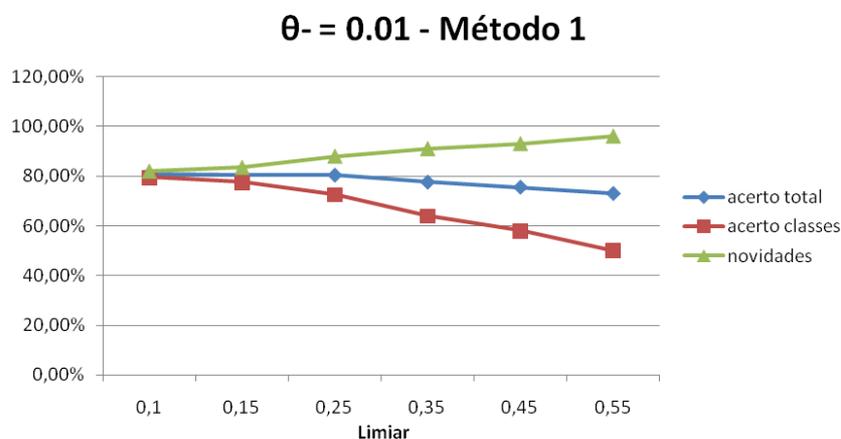


Figura 22. Comparação entre Resultados Obtidos para saídas probabilísticas para o Método 1

A Figura 23 e a Figura 24 exibem os resultados obtidos pela rede neural em função do valor do limiar, com θ^- igual a 0,1 e analisando saídas probabilísticas com o método 2.

Para o limiar 0,15 a rede obteve desempenho razoável, com 75,75% de classificações corretas, tendo 65,5% de acerto nas classes e 86% de detecção de novidades. Entretanto com o aumento do limiar a classificação dos dígitos tende a diminuir, como mostra a Figura 23.

Entretanto o resultado mais balanceado é obtido com limiar 0,1; com 72% de classificações corretas, tendo 72% de classificação de dígitos e 72% de detecção de novidades.

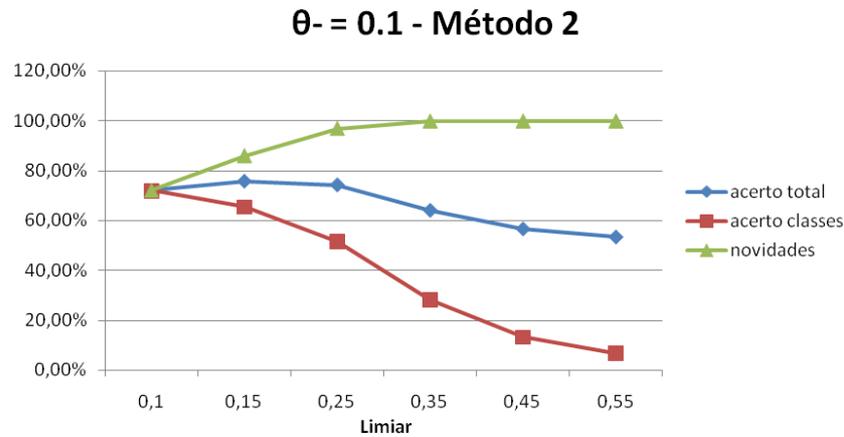


Figura 23. Comparação entre Resultados Obtidos para saídas probabilísticas para o Método 2

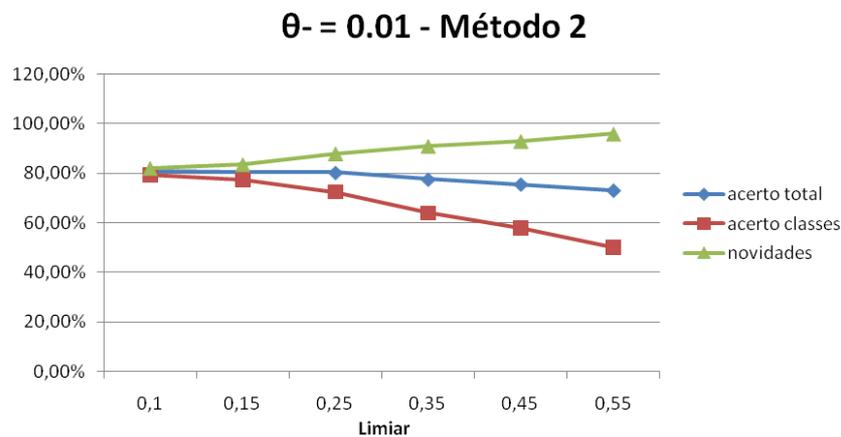


Figura 24. Comparação entre Resultados Obtidos para saídas probabilísticas para o Método 2

Para o θ^- igual a 0,01; a rede obtém um melhor desempenho geral, usando limiar igual 0,1; com 80,75% de classificações corretas. Entretanto sua capacidade de classificação de dígito diminui para 79,5%, em troca de uma boa capacidade de detecção de novidades, 82%.

4.3.1.2 PNN-DDA

A Figura 25 exibe o desempenho total da rede neural, ou seja, sua capacidade de reconhecer dígitos e ao mesmo tempo detectar padrões espúrios, treinada variando o θ^- de 0,1 até 10^{-9} . Além disso, também exibe o acerto da rede em classificar números e em detectar novidades.

Observamos que com θ^- igual a 0,1; a rede não apresenta um desempenho satisfatório, pois apesar de detectar todos os padrões espúrios, seu desempenho de classificação é muito pequeno.

Entretanto com o θ^- igual a 10^{-5} , a rede obteve um acerto total de 86%, classificando 77% de dígitos corretos e detectando 95% das entradas espúrias, como visto na Tabela 11. No entanto esse valor foi obtido para um θ^- baixo. Com a diminuição do θ^- a capacidade de classificação de dígitos tende a diminuir, como exibe o gráfico da Figura 25.

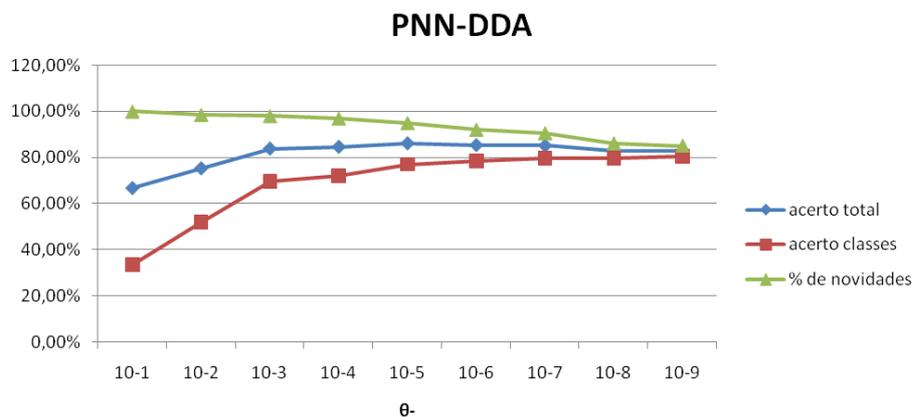


Figura 25. Resultado da rede PNN-DDA

4.3.1.3 MLP

A Figura 26 exibe o desempenho total da rede neural, ou seja, sua capacidade de reconhecer dígitos e ao mesmo tempo detectar padrões espúrios, treinada com 80 épocas variando o valor do limiar. Além disso, também exibe o acerto da rede em classificar números e em detectar novidades. Lembrando que a Figura 26 exibe um gráfico levando em consideração que na base de treinamento não tinha os dados negativos.

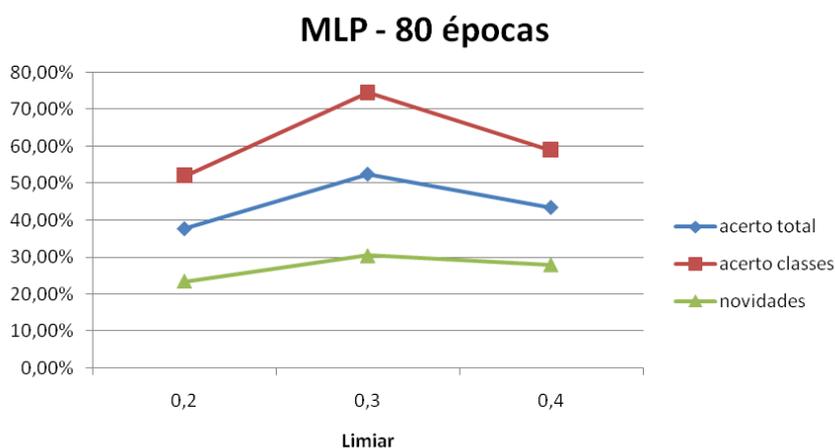


Figura 26. Resultados Obtidos sem os dados negativos na base de treinamento

Em relação com os dados negativos na base de treinamento, a rede obteve bom desempenho, e o que obteve o melhor desempenho foi para o valor de 90 épocas do treinamento,

Observamos que com 80 épocas para o treinamento, a rede não apresenta um desempenho satisfatório, pois foi a que obteve o melhor resultado variando o número de épocas do treinamento, e com o limiar igual a 0,3 onde obteve um desempenho geral de 52,5%, como visto na Tabela 12, tendo com isso 74,5% de acerto no reconhecimento dos dígitos.

Em relação com os dados negativos na base de treinamento, a rede não apresenta um desempenho satisfatório, e o que obteve o melhor resultado foi para o valor de 80 épocas no treinamento, como visto na Tabela 13. Para o limiar igual a 0,4; obteve um desempenho geral de 54,25%, tendo com isso 36% de acerto no reconhecimento dos dígitos e 72,5% de detecção de entradas espúrias. A Figura 27 exibe os resultados.

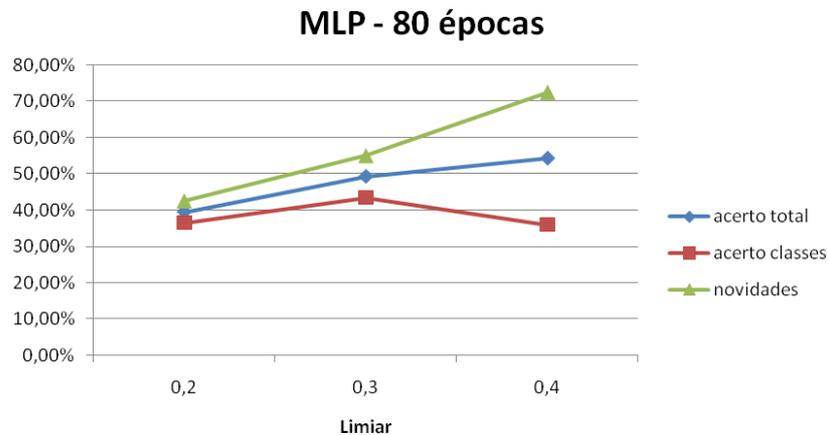


Figura 27. Resultados Obtidos com os dados negativos na base de treinamento

4.3.2 Base de Dados do MNIST

4.3.2.1 RBF-DDA

A Figura 28 exibe o desempenho total da rede neural, ou seja, sua capacidade de reconhecer dígitos e ao mesmo tempo detectar padrões espúrios, treinada com θ^- igual a 0,1 e utilizando o método 1 para detecção de novidades em função do valor do limiar de confiança. Além disso, também exibe o acerto da rede em classificar números e em detectar novidades. Lembrando que a Figura 28 exibe um gráfico levando em consideração as saídas diretas da rede RBF.

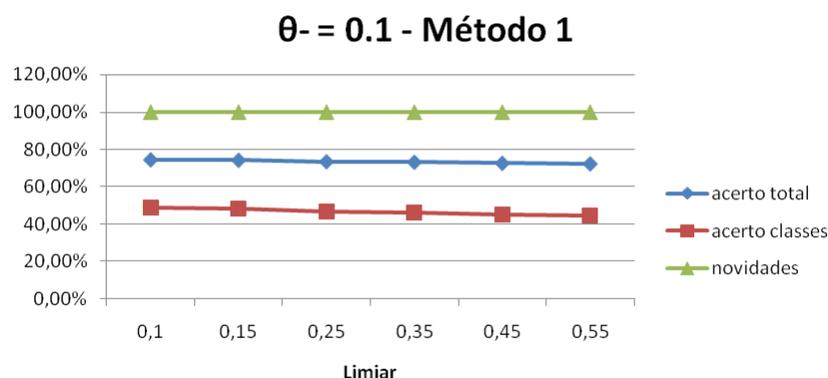


Figura 28. Comparação entre Resultados Obtidos para saídas diretas para o Método 1

Observamos que com a combinação do θ^- igual a 0,1 e utilizando o método 1, a rede não apresenta um desempenho satisfatório, pois apesar de detectar todos os padrões espúrios, seu desempenho de classificação é muito pequeno.

Com o mesmo método e com θ^- igual a 0,01; a rede já obtém melhores resultados. Para o limiar de confiança igual a 0,1, obteve um desempenho geral de 84,75%, como visto na Tabela 16, tendo com isso 69,50% de acerto no reconhecimento dos dígitos. A Figura 29 exibe os resultados.

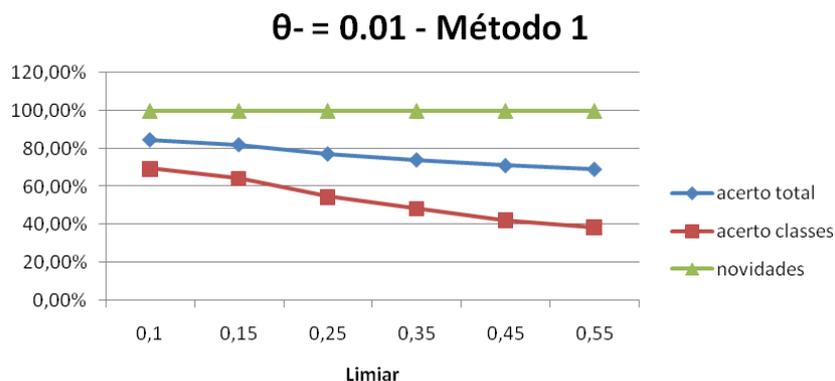


Figura 29. Comparação entre Resultados Obtidos para saídas diretas para o Método 1

Os resultados acima foram os melhores obtidos para saídas diretas da rede e com o método 1, e foi obtido com um valor de limiar de confiança muito baixo. Com a diminuição do θ^- , a tendência é da rede perder cada vez mais sua capacidade de classificar dígitos, considerando todas as entradas como espúrias.

A Figura 30 e a Figura 31 exibem os resultados da rede neural em função do valor do limiar de confiança, considerando saídas diretas da rede neural. Para saídas diretas o método 2, teve um desempenho melhor que o método 1. Para o θ^- igual 0,1 e limiar 0,1; obteve acerto de classificação total de 89,95%, como visto na Tabela 15, sendo 80,9% de acerto de classificação de dígitos e 99% de detecção de entradas espúrias.

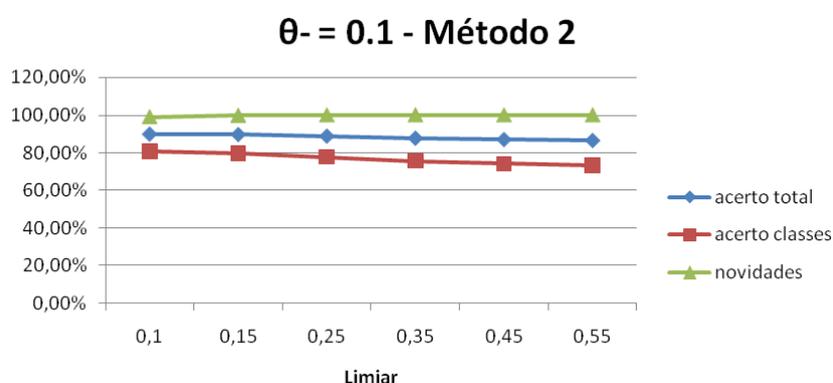


Figura 30. Comparação entre Resultados Obtidos para saídas diretas para o Método 2

Com o θ^- igual a 0,01; o método 2, já não obteve um desempenho tão bom. Para um limiar de confiança de 0,1; obteve 86,8% de acerto total, com 73,6% de classificações corretas para dígitos e 100% para detecção de novidades, como visto na Tabela 17. Para valores menores de θ^- , a rede neural perde cada vez mais sua capacidade de classificação, considerando todas as entradas como espúrias. A Figura 31 exhibe uma tendência de diminuição de acerto de classificação de dígitos com o aumento do limiar de confiança.

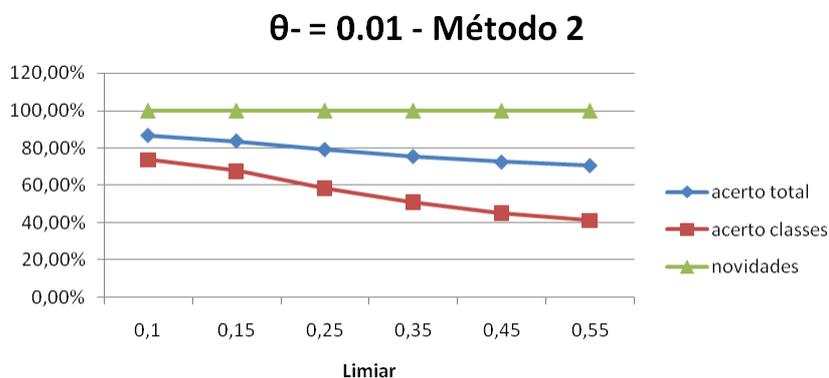


Figura 31. Comparação entre Resultados Obtidos para saídas diretas para o Método 2

Em relação às saídas probabilísticas, o método 1 não obteve um bom resultado, considerando o θ^- igual a 0,1. Observe a Figura 32.

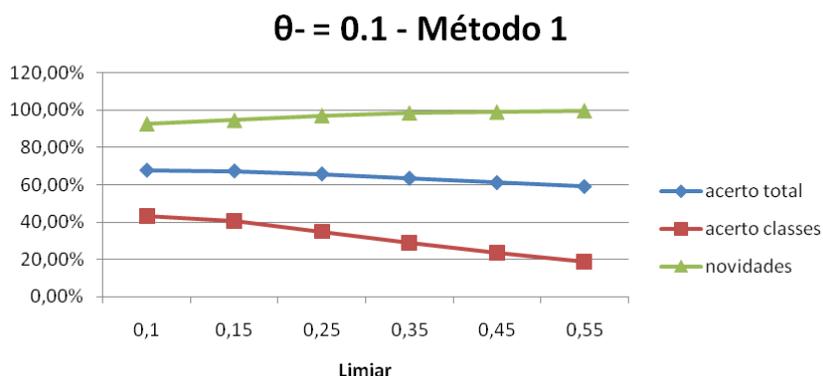


Figura 32. Comparação entre Resultados Obtidos para saídas probabilísticas para o Método 1

Entretanto com o θ^- igual a 0,01; a rede obteve um acerto total de 81,6%, classificando 71,39% de dígitos corretos e detectando 91,8% das entradas espúrias, como visto na Tabela 16. No entanto esse valor foi obtido para um limiar de confiança baixo. Com o aumento do limiar a capacidade de classificação de dígitos tende a diminuir, como exibe o gráfico da figura 10.

Observamos também, que utilizando saídas probabilísticas, e havendo a diminuição do θ^- , a tendência é de aumentar a capacidade de classificação correta dos dígitos. Entretanto a rede perde cada vez mais sua capacidade de detectar novidades. Essa tendência é observada na Tabela 16.

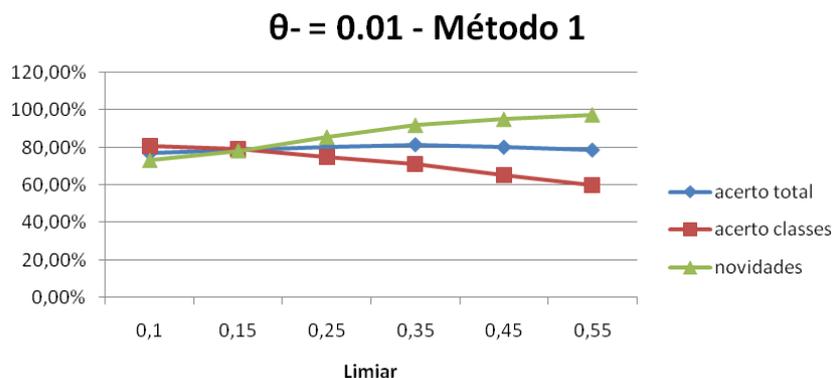


Figura 33. Comparação entre Resultados Obtidos para saídas probabilísticas para o Método 1

A Figura 34 e a Figura 35 exibem os resultados obtidos pela rede neural em função do valor do limiar, com θ^- igual a 0,1 e analisando saídas probabilísticas com o método 2.

Para o limiar 0,35 a rede obteve desempenho razoável, com 68,15% de classificações corretas, tendo 47,5% de acerto nas classes e 88,8% de detecção de novidades. Entretanto com o aumento do limiar a classificação dos dígitos tende a diminuir, como mostra a Figura 34.

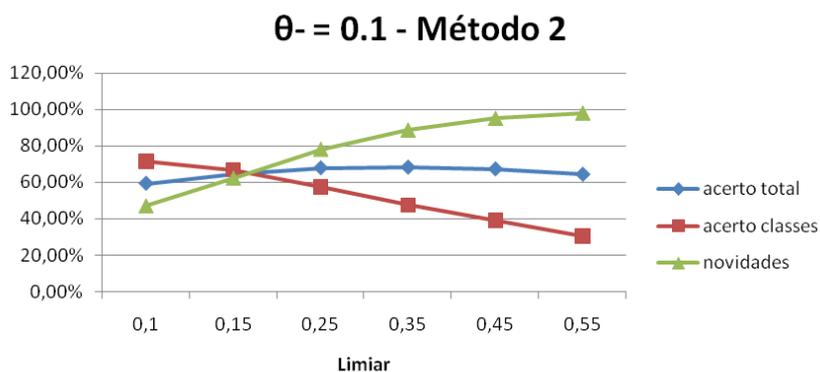


Figura 34. Comparação entre Resultados Obtidos para saídas probabilísticas para o Método 2

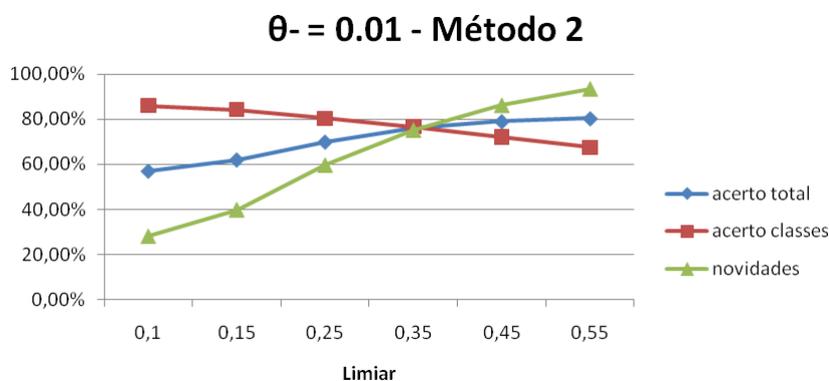


Figura 35. Comparação entre Resultados Obtidos para saídas probabilísticas para o Método 2

Para o θ^- igual a 0,01; a rede obtém um melhor desempenho geral, usando limiar igual 0,55; com 80,4% de classificações corretas. Entretanto sua capacidade de classificação de dígito aumentou para 67,5%, tendo também uma boa capacidade de detecção de novidades, 93,3%.

Entretanto o resultado mais balanceado é obtido com limiar 0,35; com 75,95% de classificações corretas, tendo 76,8% de classificação de dígitos e 75,1% de detecção de novidades.

4.3.2.2 PNN-DDA

A Figura 36 exibe o desempenho total da rede neural, ou seja, sua capacidade de reconhecer dígitos e ao mesmo tempo detectar padrões espúrios, treinada variando o θ^- de 0,1 até 10^{-18} . Além disso, também exibe o acerto da rede em classificar números e em detectar novidades.

Observamos que com θ^- igual a 0,1; a rede não apresenta um desempenho satisfatório, pois apesar de detectar todos os padrões espúrios, seu desempenho de classificação é muito pequeno.

Entretanto com o θ^- igual a 10^{-18} , a rede obteve um acerto total de 91,8%, classificando 83,6% de dígitos corretos e detectando 100% das entradas espúrias, como visto na Tabela 22. No entanto esse valor foi obtido para um θ^- muito baixo. Com a diminuição do θ^- a capacidade de classificação de dígitos tende a aumentar, como exibe o gráfico da Figura 36.

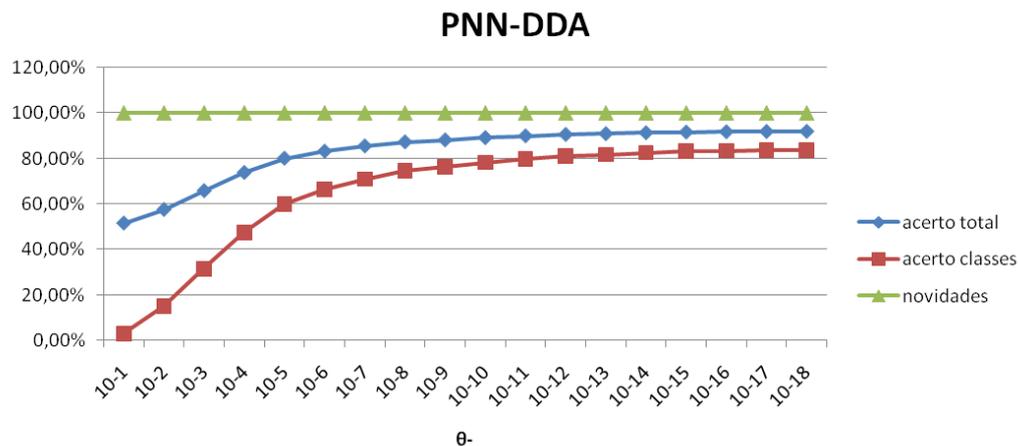


Figura 36. Resultado da rede PNN-DDA

4.3.2.3 MLP

A Figura 37 exibe o desempenho total da rede neural, ou seja, sua capacidade de reconhecer dígitos e ao mesmo tempo detectar padrões espúrios, treinada com 30 épocas variando o valor do limiar. Além disso, também exibe o acerto da rede em classificar números e em detectar novidades. Lembrando que a Figura 37 exibe um gráfico levando em consideração que na base de treinamento não tinha os dados negativos.

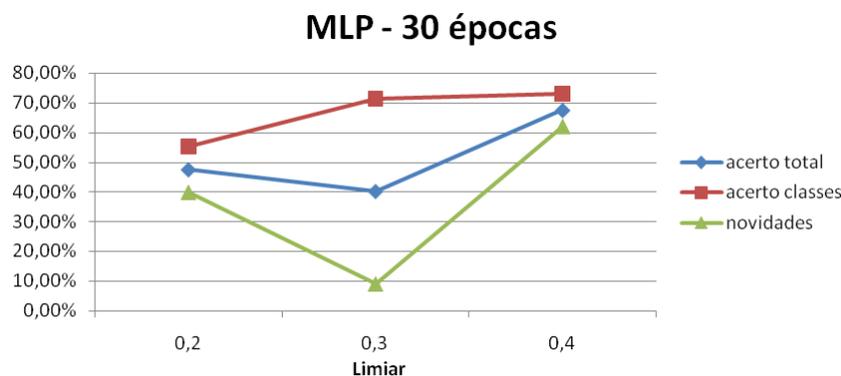


Figura 37. Resultados Obtidos sem os dados negativos na base de treinamento

Observamos que com 30 épocas para o treinamento, a rede apresenta um desempenho satisfatório, pois o melhor resultado foi com o limiar igual a 0,4 onde obteve um desempenho geral de 67,5%, como visto na Tabela 23, tendo com isso 73,1% de acerto no reconhecimento dos dígitos e 61,9% de detecção de entradas espúrias.

Alterando o valor do número de épocas do treinamento para 60 e 90, a rede não obteve bons resultados, como visto na Tabela 23.

Em relação com os dados negativos na base de treinamento, a rede obteve bom desempenho, e o que obteve o melhor desempenho foi para o valor de 90 épocas do treinamento, como mostra a Figura 38. Para o limiar igual a 0,3; obteve um desempenho geral de 86,45%, como visto na Tabela 24, tendo com isso 73,7% de acerto no reconhecimento dos dígitos e 99,2% de detecção de entradas espúrias.

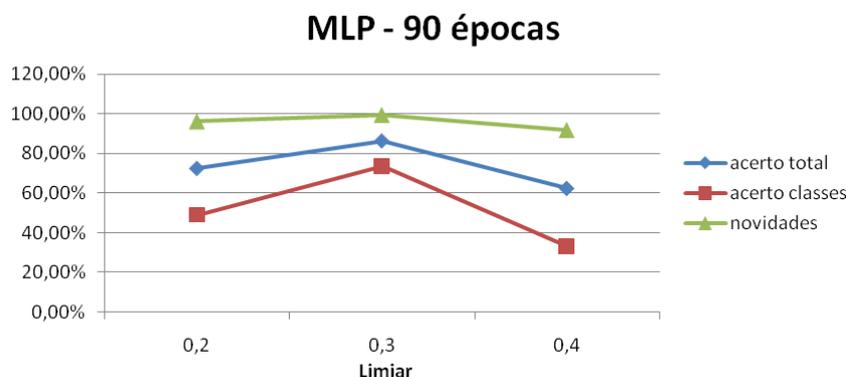


Figura 38. Resultados Obtidos com os dados negativos na base de treinamento

4.4 Tempo de Treinamento dos Melhores Resultados

Nesta seção mostrarei o tempo dos melhores resultados obtidos do treinamento com as diferentes bases utilizadas, sendo uma das bases de documentos históricos e outra da base do MNIST.

O computador utilizado na geração de todos os tempos de treinamentos possui a seguinte configuração:

- Processador: Intel Dual Core D de 2,8 GHz;

- Memória principal: 1 Gb DRR2 – 533Mhz;
- Sistema Operacional: Windows XP.

4.4.1 Base Dados de Documentos Históricos

Pegando os melhores resultados encontrados em cada uma das redes e calculamos o tempo de processamento de cada, como mostra a Tabela 25. Percebemos que o PNN obteve um tempo total de resposta abaixo de 1 segundo. No caso para a rede RBF o tempo total ficou em torno dos 13 segundos, no qual é um tempo razoável de resposta.

Já com relação a rede MLP, o tempo total de resposta já foi um pouco significativo, em um caso o tempo chegou a mais de 2 minutos, no entanto o resultado da rede também não foi muito bom para ambos os casos.

Tabela 25. Tempo de treinamento das redes utilizando a base dados de documentos históricos

Rede Neural	acerto total	acerto classes	novidades	tempo treinamento (s)	tempo teste (s)	tempo total (s)
RBF - Saídas Diretas - Método 2 - $\theta^- = 0,1$ - limiar = 0,35	82,75%	72%	93,50%	9	4	13
RBF - Saídas Probabilísticas - Método 1 - $\theta^- = 0,001$ - limiar = 0,55	81,75%	78,00%	85,50%	8	4	12
PNN - $\theta^- = 10^{-5}$	86,00%	77%	95%	<1	<1	<1
MLP - Sem os dados negativos - épocas = 80 - limiar = 0,3	52,50%	74,50%	30,50%	130	1	131
MLP - Com os dados negativos - épocas = 30 - limiar = 0,4	56%	37%	75%	48	1	49

4.4.2 Base de Dados do MNIST

Pegando os melhores resultados encontrados em cada uma das redes e calculamos o tempo de processamento de cada, como mostra a Tabela 26. Para a base do MNIST o que obteve o melhor tempo foi o PNN com um tempo total de resposta igual a 20 segundos. No caso para a rede RBF o tempo total ficou acima dos 11 minutos, no qual é um tempo bastante significativo.

Com relação a rede MLP, em um caso o tempo total de resposta chegou a 5 minutos, melhor do que o RBF, mas em compensação o resultado da rede é mediano. No outro resultado da rede MLP o tempo total de resposta já foi muito significativo, ultrapassando os 14 minutos. Vale salientar que o tempo aumentou significativamente em ambos os casos, porque o tamanho da base do MNIST comparada a do documentos históricos é cinco vezes maior.

Tabela 26. Tempo de treinamento das redes utilizando a base dados do MNIST

Rede Neural	acerto total	acerto classes	novidades	tempo treinamento (s)	tempo teste (s)	tempo total (s)
RBF - Saídas Diretas - Método 2 - $\theta^- = 0,1$ - limiar = 0,1	89,95%	80,90%	99,00%	646	100	746
RBF - Saídas Probabilísticas - Método 1 - $\theta^- = 0,001$ - limiar = 0,35	81,60%	71,39%	91,80%	591	100	691
PNN - $\theta^- = 10^{-18}$	91,80%	83,60%	100,00%	9	11	20
MLP - Sem os dados negativos - épocas = 30 - limiar = 0,4	67,50%	73,10%	61,90%	297	1	298
MLP - Com os dados negativos - épocas = 90 - limiar = 0,3	86,45%	73,70%	99,20%	856	1	857

Capítulo 5

Conclusões e Trabalhos Futuros

Neste trabalho, foi considerado o problema de reconhecimento de dígitos manuscritos em documentos históricos e na base do MNIST, para detecção de entradas espúrias, ou seja, valores indesejados, que não pertencem às classes de dígitos utilizadas na etapa de treinamento da rede neural, as redes neurais utilizadas foram o RBF-DDA, PNN-DDA e MLP.

5.1 Contribuições

A principal contribuição do trabalho é que ao mesmo tempo em que se segmenta a string de números, a rede detecta as novidades e reconhece os dígitos. Algumas outras contribuições do trabalho estão listadas a seguir:

- Estudo comparativo das técnicas utilizadas;
- A aplicação inovadora da técnica de PNN-DDA para detecção de novidades.

5.2 Estudo Comparativo

Foram utilizados dois métodos de análise das saídas geradas pela rede RBF, considerando o conceito de nível de confiança para aceitação de uma decisão de classificação. O primeiro método é baseado no trabalho de Vasconcelos [20], e compara a maior saída da RBF com a soma das menores, verificando se a diferença entre elas é maior que o nível de confiança, aceitando assim a classificação da rede.

O segundo método proposto utiliza o mesmo mecanismo do método anterior, entretanto nesse caso compara-se a maior saída com a segunda maior, e não com todas as outras.

Foram utilizadas essas duas técnicas para dois tipos de saídas fornecidas pela rede RBF. Foi considerado primeiramente as saídas diretas e posteriormente as saídas probabilísticas. Também foi analisado a influência do parâmetro θ^- na decisão de aceitação de uma classificação.

A partir dos experimentos, concluímos que tanto para a base de dados de documentos históricos como para a base do MNIST, o que obteve o melhor desempenho analisando as saídas diretas da rede RBF foi para o método 2, utilizando o valor do θ^- 0,1 para ambos os casos.

Obtivemos bons resultados, mesmo com valores altos de limiar de confiança. Observamos também que com a diminuição do θ^- , a capacidade de classificação de dígitos decai, sendo muitos considerados espúrios.

O método 1 obteve melhores resultados analisando saídas probabilísticas, tanto para a base de dados de documentos históricos como para a base do MNIST, para θ^- igual a 0,001 e para θ^- igual a 0,01 respectivamente, entretanto com a diminuição do valor de θ^- há uma perda gradativa na habilidade de detecção de padrões espúrios.

No caso da rede PNN, com a base de dados de documentos históricos, o que obteve o melhor desempenho foi com o θ^- igual 10^{-5} . Observamos também que com a diminuição do θ^- , a capacidade de classificação de dígitos espúrios decai. No caso da base do MNIST, o que obteve o melhor desempenho foi com o θ^- igual 10^{-18} .

Com relação a rede MLP, foram utilizados dois métodos de análises, a primeira sem os dados espúrios no treinamento e a segunda com os dados espúrios no treinamento. No primeiro método utilizando a base de documentos históricos, não obteve um resultado razoável, já com a base do MNIST obteve um resultado razoável, com o limiar igual a 0,4 e com 30 épocas para o treinamento.

No segundo método, utilizando os dados de documentos históricos, também não se obteve um bom resultado, no caso da base do MNIST, chegou a um bom resultado com o limiar igual a 0,3 e com 90 épocas para o treinamento.

Concluimos que para cada problema é preciso uma análise do que se espera como resultado da rede neural. Se para a base de dados de documentos históricos o objetivo for obter uma boa capacidade de classificação de dígitos (77%) e uma grande habilidade de detecção de novidades (95%), a melhor configuração seria utilizar a rede PNN com $\theta^- 10^{-5}$. Se o objetivo for obter uma grande habilidade de classificação de dígitos (81%) e apenas uma capacidade mediana de detecção de novidades (66%), a melhor configuração seria utilizar a rede RBF com saídas diretas e treinando com $\theta^- 0,1$, utilizando o método 2 com limiar de 0.1.

No caso da base do MNIST, se o objetivo for obter uma grande habilidade de classificação de dígitos (80,9%) e uma capacidade alta de detecção de novidades (99%), a melhor configuração seria utilizar redes RBF com saídas diretas e treinado com $\theta^- 0,1$, utilizando o método 2 com limiar de 0,1.

O método 2 obteve resultados superiores ao método 1 levando em consideração saídas diretas. Entretanto para saídas probabilísticas, o método 1 se saiu um pouco melhor, mas a capacidade de detecção de novidades decaiu com a diminuição do θ^- . No caso do MLP, o a base que não tinha os dados negativos no treinamento não obteve um bom resultado quando comparado com a que teve os dados negativos no treinamento. No entanto para o PNN, com a base do MNIST quanto mais diminuir o valor de θ^- maior a detecção de dígitos e padrões espúrios, enquanto na base de documentos históricos não ocorreu tal fato.

Na realidade a melhor configuração de parâmetros para cada uma das bases, deve ser definida de acordo com a demanda do problema. Alguns problemas podem exigir uma maior confiança na detecção de padrões espúrios, enquanto outras podem demandar maior potencial de classificação. Deve-se analisar com cuidado qual é o tipo de problema e qual a melhor abordagem.

5.3 Trabalhos Futuros

Como trabalhos futuros, propomos a criação de um programa que segmente e classifique os dígitos, indexando-os pelas datas. Outro trabalho proposto é a utilização de um maior número de padrões da base de dados utilizada, com isso pode melhorar ainda mais a confiabilidade dos

resultados, podendo fazer uma comparação com os resultados obtidos com as redes RBF, PNN e MLP. Também propomos a utilização de outras arquiteturas de redes neurais ou de outros tipos de classificadores e uma comparação de seu desempenho com o obtido com as redes RBF, PNN e MLP.

Bibliografia

- [1] VELLASQUES, E. *Classificação de Pontos de Segmentação de Dígitos Manuscritos*. Dissertação de Mestrado, Pontifícia Universidade Católica do Paraná, 2006.
- [2] OLIVEIRA, A., MELLO, C., SILVA, E., ALVES, V. *Optical Digit Recognition for Images of Handwritten Historical Documents*. Proceedings of the Ninth Brazilian Symposium on Neural Networks, 2006.
- [3] The MNIST Database of Handwritten Digits. Disponível on-line em <http://yann.lecun.com/exdb/mnist/>. Acesso em 23 de outubro de 2007.
- [4] Data for MATLAB hackers. Disponível on-line em <http://www.cs.toronto.edu/~7Eroweis/data.html>. Acesso em 23 de outubro de 2007.
- [5] BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. *Redes Neurais Artificiais Teoria e Aplicações*. Livros Técnicos e Científicos Editora, Rio de Janeiro, 2000.
- [6] MUKKAMALA, S.; SUNG, A. H.; ABRAHAM, A. *Intrusion detection using an ensemble of intelligent paradigms*. Journal of Network and Computer Applications, volume 28, p. 167-182, 2005.
- [7] WANG, X. G., TANG, Z., TAMURA, H., ISHII, M., e SUN, W. D.. *An improved backpropagation algorithm to avoid the local minima problem*. Neurocomputing, 56:455–460, 2004.
- [8] XU, J., HO, D.W.C., *A New Training and Pruning Algorithm Based on Node Dependence and Jacobian Rank Deficiency*, Neurocomputing, v. 70, n. 1-3, pp. 544-58, 2006.
- [9] OLIVEIRA, A. L. I.; MELO, B. J. M.; MEIRA, S. R. L. *Integrated method for constructive training of radial basis function networks*. Electronics Letters, volume 41, p. 429-430, 2005.
- [10] OLIVEIRA, A. L. I.; NETO, F. B. L.; MEIRA, S. R. L. *Improving RBF-DDA Performance on Optical Character Recognition through Parameter Selection*. Proceedings of the 17th International Conference on volume 4, p. 625–628, 2004.
- [11] OLIVEIRA A. L. I.; A. MEDEIROS E. A.; ROCHA, T. A. B. V.; BEZERRA, M. E. R.; VERAS R. C. *A Study on the Influence of Parameter θ^- on Performance of RBF Neural Networks Trained with the Dynamic Decay Adjustment Algorithm*. Fifth International Conference on Hybrid Intelligent Systems, 2005.
- [12] HUDAK, M. J. *RCE classifiers: Theory and practice*. Cybernetics and Systems, 23:483–515, 1992.
- [13] MUELLER A. *Uma aplicação de redes neurais artificiais na previsão do mercado acionário*. Dissertação de Pós-graduação. Universidade Federal de Santa Catarina. Julho, 2005.
- [14] CAMPBELL, C. *Kernel methods: a survey of current techniques*. Neurocomputing, volume 48, p. 63-84, 2002.
- [15] KUNCHEVA L. I. *Reducing the computational demand of the nearest neighbor classifier*. Mathematics Publications List. Universidade de Wales. 2001.

- [16] SPECHT D. F. *Probabilistic Neural Networks*. Neural Networks. pp. 109-118, 1990.
- [17] PARZEN E. *On the estimation of a probability density function*. Ann Math Statist. P1065-1076, 1962.
- [18] SPECHT D. F. *PNN: from fast training to fast running*. Computational Intelligence, A Dynamic System Perspective, IEEE Press. New York , p. 246-258, 1995.
- [19] BERTHOLD M. R., DIAMOND J. *Constructive training of probabilistic neural networks*. Neurocomputing. Vol. 19, p. 167-183, 1997.
- [20] VASCONCELOS, G. *An Investigation of Feedforward Neural Networks with Respect to the Deteccion of Spurius Patterns*. 1995. Tese submetida a Universidade de Kent.

Apêndice A

Resultados do treinamento das redes RBF-DDA com as saídas diretas utilizando a base de dados de documentos históricos

Neste apêndice serão mostrados todos os resultados advindos do treinamento das redes RBF-DDA com as saídas diretas utilizando a base de dados de documentos históricos.

Comparando a maior saída com todas as outras (Método 1)						
theta-	limiar	acerto total	acerto classes	% de novidades	qtd de Neurônios	Épocas
0,1	0,1	62,5%	25%	100%	162	4
0,1	0,15	61,5%	23%	100%	162	4
0,1	0,25	59,5%	19%	100%	162	4
0,1	0,35	58,5%	17%	100%	162	4
0,1	0,45	57,5%	15%	100%	162	4
0,1	0,55	57%	14%	100%	162	4
0,01	0,1	79,25%	59,5%	99%	195	3
0,01	0,15	75%	50%	100%	195	3
0,01	0,25	68,75%	37,5%	100%	195	3
0,01	0,35	64%	28%	100%	195	3
0,01	0,45	60%	20%	100%	195	3
0,01	0,55	57%	14%	100%	195	3
0,001	0,1	65,5%	31%	100%	199	3
0,001	0,15	59,75%	19,5%	100%	199	3
0,001	0,25	56%	12%	100%	199	3
0,001	0,35	54,5%	9%	100%	199	3
0,001	0,45	51,25%	2,5%	100%	199	3
0,001	0,55	50,75%	1,5%	100%	199	3
0,0001	0,1	56,5%	13%	100%	199	3
0,0001	0,15	54,5%	9%	100%	199	3
0,0001	0,25	50,5%	1%	100%	199	3

0,0001	0,35	50,5%	1%	100%	199	3
0,0001	0,45	50,25%	0,5%	100%	199	3
0,0001	0,55	50%	0%	100%	199	3
0,00001	0,1	52,5%	5%	100%	200	2
0,00001	0,15	50,5%	1%	100%	200	2
0,00001	0,25	50,5%	1%	100%	200	2
0,00001	0,35	50%	0%	100%	200	2
0,00001	0,45	50%	0%	100%	200	2
0,00001	0,55	50%	0%	100%	200	2
0,000001	0,1	50,5%	1%	100%	200	2
0,000001	0,15	50,5%	1%	100%	200	2
0,000001	0,25	50%	0%	100%	200	2
0,000001	0,35	50%	0%	100%	200	2
0,000001	0,45	50%	0%	100%	200	2
0,000001	0,55	50%	0%	100%	200	2

Comparado a maior saída com a segunda maior (Método 2)

theta-	limiar	acerto total	acerto classes	% de novidades	qtd de Neurônios	Épocas
0,1	0,1	73,5%	81%	66%	162	4
0,1	0,15	78%	78,5%	77,5%	162	4
0,1	0,25	81%	74,5%	87,5%	162	4
0,1	0,35	82,75%	72%	93,5%	162	4
0,1	0,45	82,25%	67%	97,5%	162	4
0,1	0,55	80,75%	63,5%	98%	162	4
0,01	0,1	81,75%	65%	98,5%	195	3
0,01	0,15	77%	54%	99,5%	195	3
0,01	0,25	70,25%	40,5%	100%	195	3
0,01	0,35	64,75%	29,5%	100%	195	3
0,01	0,45	61%	22%	100%	195	3
0,01	0,55	57%	16%	100%	195	3
0,001	0,1	64,75%	29,5%	100%	199	3
0,001	0,15	59,25%	18,5%	100%	199	3
0,001	0,25	55,5%	11%	100%	199	3
0,001	0,35	54,25%	8,5%	100%	199	3
0,001	0,45	51,25%	2,5%	100%	199	3
0,001	0,55	50,75%	1,5%	100%	199	3
0,0001	0,1	56%	12%	100%	199	3
0,0001	0,15	54%	8%	100%	199	3
0,0001	0,25	50,5%	1%	100%	199	3
0,0001	0,35	50,5%	1%	100%	199	3
0,0001	0,45	50,25%	0,5%	100%	199	3
0,0001	0,55	50%	0%	100%	199	3
0,00001	0,1	52%	4%	100%	200	2
0,00001	0,15	50,5%	1%	100%	200	2
0,00001	0,25	50,5%	1%	100%	200	2
0,00001	0,35	50%	0%	100%	200	2
0,00001	0,45	50%	0%	100%	200	2
0,00001	0,55	50%	0%	100%	200	2
0,000001	0,1	50,5%	1%	100%	200	2
0,000001	0,15	50,5%	1%	100%	200	2
0,000001	0,25	50%	0%	100%	200	2

0,000001	0,35	50%	0%	100%	200	2
0,000001	0,45	50%	0%	100%	200	2
0,000001	0,55	50%	0%	100%	200	2

Apêndice B

Resultados do treinamento das redes RBF-DDA com as saídas probabilísticas utilizando a base de dados de documentos históricos

Neste apêndice serão mostrados todos os resultados advindos do treinamento das redes RBF-DDA com as saídas probabilísticas utilizando a base de dados de documentos históricos.

Comparando a maior saída com todas as outras (Método 1)						
theta-	limiar	acerto total	acerto classes	% de novidades	qtd de Neurônios	Épocas
0,1	0,1	58,75%	17,50%	100,00%	162	4
0,1	0,15	56,50%	13,00%	100,00%	162	4
0,1	0,25	53,75%	7,50%	100,00%	162	4
0,1	0,35	51,75%	3,50%	100,00%	162	4
0,1	0,45	50,00%	0,00%	100,00%	162	4
0,1	0,55	50,00%	0,00%	100,00%	162	4
0,01	0,1	80,75%	79,50%	82,00%	195	3
0,01	0,15	80,50%	77,50%	83,50%	195	3
0,01	0,25	80,25%	72,50%	88,00%	195	3
0,01	0,35	77,50%	64,00%	91,00%	195	3
0,01	0,45	75,50%	57,99%	93,00%	195	3
0,01	0,55	73,00%	50,00%	96,00%	195	3
0,001	0,1	70,75%	92,50%	49,00%	199	3
0,001	0,15	72,00%	91,00%	53,00%	199	3
0,001	0,25	75,00%	90,50%	59,50%	199	3
0,001	0,35	79,25%	86,50%	72,00%	199	3
0,001	0,45	79,75%	80,50%	79,00%	199	3
0,001	0,55	81,75%	78,00%	85,50%	199	3
0,0001	0,1	65,50%	95,00%	36,00%	199	3
0,0001	0,15	66,50%	94,00%	39,00%	199	3

0,0001	0,25	70,00%	93,00%	47,00%	199	3
0,0001	0,35	71,75%	91,50%	52,00%	199	3
0,0001	0,45	74,50%	90,50%	58,50%	199	3
0,0001	0,55	75,50%	87,50%	63,50%	199	3
0,00001	0,1	60,50%	97,00%	24,00%	200	2
0,00001	0,15	62,00%	95,50%	28,49%	200	2
0,00001	0,25	66,50%	94,50%	38,50%	200	2
0,00001	0,35	68,00%	93,00%	43,00%	200	2
0,00001	0,45	70,00%	92,00%	48,00%	200	2
0,00001	0,55	72,00%	90,50%	53,50%	200	2
0,000001	0,1	58,50%	98,00%	19,00%	200	2
0,000001	0,15	60,75%	97,00%	24,50%	200	2
0,000001	0,25	61,25%	95,50%	27,00%	200	2
0,000001	0,35	64,00%	94,00%	34,00%	200	2
0,000001	0,45	67,00%	93,00%	41,00%	200	2
0,000001	0,55	69,00%	92,00%	46,00%	200	2

Comparado a maior saída com a segunda maior (Método 2)

theta-	limiar	acerto total	acerto classes	% de novidades	qtd de Neurônios	Épocas
0,1	0,1	72,00%	72,00%	72,00%	162	4
0,1	0,15	75,75%	65,50%	86,00%	162	4
0,1	0,25	74,25%	51,50%	97,00%	162	4
0,1	0,35	64,00%	28,00%	100,00%	162	4
0,1	0,45	56,50%	13,00%	100,00%	162	4
0,1	0,55	53,25%	6,50%	100,00%	162	4
0,01	0,1	56,50%	85,00%	28,00%	195	3
0,01	0,15	63,00%	84,00%	42,00%	195	3
0,01	0,25	72,75%	81,50%	64,00%	195	3
0,01	0,35	78,75%	77,00%	80,50%	195	3
0,01	0,45	79,00%	70,00%	88,00%	195	3
0,01	0,55	76,75%	62,00%	91,50%	195	3
0,001	0,1	54,00%	86,00%	22,00%	199	3
0,001	0,15	57,75%	86,00%	29,50%	199	3
0,001	0,25	62,25%	84,50%	40,00%	199	3
0,001	0,35	66,75%	84,00%	49,50%	199	3
0,001	0,45	73,00%	81,50%	64,50%	199	3
0,001	0,55	76,75%	79,00%	74,50%	199	3
0,0001	0,1	50,50%	87,00%	14,00%	199	3
0,0001	0,15	54,00%	86,50%	21,50%	199	3
0,0001	0,25	59,50%	85,50%	33,50%	199	3
0,0001	0,35	62,50%	85,50%	39,50%	199	3
0,0001	0,45	65,75%	84,50%	47,00%	199	3
0,0001	0,55	69,25%	83,00%	55,50%	199	3
0,00001	0,1	49,75%	88,00%	11,50%	200	2
0,00001	0,15	51,25%	86,50%	16,00%	200	2
0,00001	0,25	55,25%	86,00%	24,50%	200	2
0,00001	0,35	60,75%	85,50%	36,00%	200	2
0,00001	0,45	63,00%	85,50%	40,50%	200	2
0,00001	0,55	65,50%	84,00%	47,00%	200	2
0,000001	0,1	48,75%	88,00%	9,50%	200	2
0,000001	0,15	49,75%	87,00%	12,50%	200	2

0,000001	0,25	52,50%	85,50%	19,50%	200	2
0,000001	0,35	56,00%	84,50%	27,50%	200	2
0,000001	0,45	59,00%	84,50%	33,50%	200	2
0,000001	0,55	63,25%	84,50%	42,00%	200	2

Apêndice C

Resultados do treinamento das redes RBF-DDA com as saídas diretas utilizando a base de dados do MNIST

Neste apêndice serão mostrados todos os resultados advindos do treinamento das redes RBF-DDA com as saídas diretas utilizando a base de dados do MNIST.

Comparando a maior saída com todas as outras (Método 1)						
theta-	limiar	acerto total	acerto classes	% de novidades	qtd de Neurônios	Épocas
0,1	0,1	74,35%	48,69%	100,00%	1548	4
0,1	0,15	74,10%	48,19%	100,00%	1548	4
0,1	0,25	73,35%	46,70%	100,00%	1548	4
0,1	0,35	73,05%	46,10%	100,00%	1548	4
0,1	0,45	72,50%	45,00%	100,00%	1548	4
0,1	0,55	72,20%	44,40%	100,00%	1548	4
0,01	0,1	84,75%	69,50%	100,00%	1892	3
0,01	0,15	82,15%	64,30%	100,00%	1892	3
0,01	0,25	77,30%	54,60%	100,00%	1892	3
0,01	0,35	74,20%	48,40%	100,00%	1892	3
0,01	0,45	71,20%	42,40%	100,00%	1892	3
0,01	0,55	69,25%	38,50%	100,00%	1892	3
0,001	0,1	71,70%	43,40%	100,00%	1948	3
0,001	0,15	68,30%	36,60%	100,00%	1948	3
0,001	0,25	63,90%	27,80%	100,00%	1948	3
0,001	0,35	61,40%	22,80%	100,00%	1948	3
0,001	0,45	59,70%	19,40%	100,00%	1948	3
0,001	0,55	57,95%	15,90%	100,00%	1948	3
0,0001	0,1	62,45%	24,90%	100,00%	1971	3
0,0001	0,15	59,95%	19,90%	100,00%	1971	3
0,0001	0,25	57,10%	14,20%	100,00%	1971	3
0,0001	0,35	55,10%	10,20%	100,00%	1971	3
0,0001	0,45	54,40%	8,79%	100,00%	1971	3
0,0001	0,55	54,00%	8,00%	100,00%	1971	3

0,00001	0,1	57,10%	14,20%	100,00%	1981	3
0,00001	0,15	55,30%	10,60%	100,00%	1981	3
0,00001	0,25	54,10%	8,20%	100,00%	1981	3
0,00001	0,35	53,30%	6,60%	100,00%	1981	3
0,00001	0,45	52,65%	5,30%	100,00%	1981	3
0,00001	0,55	52,35%	4,70%	100,00%	1981	3
0,000001	0,1	54,65%	9,30%	100,00%	1988	3
0,000001	0,15	53,85%	7,70%	100,00%	1988	3
0,000001	0,25	52,75%	5,50%	100,00%	1988	3
0,000001	0,35	52,15%	4,30%	100,00%	1988	3
0,000001	0,45	51,80%	3,59%	100,00%	1988	3
0,000001	0,55	51,60%	3,20%	100,00%	1988	3

Comparado a maior saída com a segunda maior (Método 2)

theta-	limiar	acerto total	acerto classes	% de novidades	qtd de Neurônios	Epocas
0,1	0,1	89,95%	80,90%	99,00%	1548	4
0,1	0,15	89,75%	79,70%	99,80%	1548	4
0,1	0,25	88,90%	77,80%	100,00%	1548	4
0,1	0,35	87,80%	75,60%	100,00%	1548	4
0,1	0,45	87,10%	74,20%	100,00%	1548	4
0,1	0,55	86,60%	73,20%	100,00%	1548	4
0,01	0,1	86,80%	73,60%	100,00%	1892	3
0,01	0,15	83,80%	67,60%	100,00%	1892	3
0,01	0,25	79,20%	58,40%	100,00%	1892	3
0,01	0,35	75,50%	51,00%	100,00%	1892	3
0,01	0,45	72,50%	45,00%	100,00%	1892	3
0,01	0,55	70,60%	41,19%	100,00%	1892	3
0,001	0,1	71,75%	43,50%	100,00%	1948	3
0,001	0,15	68,40%	36,80%	100,00%	1948	3
0,001	0,25	63,75%	27,50%	100,00%	1948	3
0,001	0,35	61,25%	22,50%	100,00%	1948	3
0,001	0,45	59,65%	19,30%	100,00%	1948	3
0,001	0,55	57,95%	15,90%	100,00%	1948	3
0,0001	0,1	61,05%	22,10%	100,00%	1971	3
0,0001	0,15	58,85%	17,70%	100,00%	1971	3
0,0001	0,25	56,50%	13,00%	100,00%	1971	3
0,0001	0,35	54,80%	9,60%	100,00%	1971	3
0,0001	0,45	54,20%	8,40%	100,00%	1971	3
0,0001	0,55	53,80%	7,60%	100,00%	1971	3
0,00001	0,1	55,90%	11,79%	100,00%	1981	3
0,00001	0,15	54,75%	9,50%	100,00%	1981	3
0,00001	0,25	53,80%	7,60%	100,00%	1981	3
0,00001	0,35	53,10%	6,20%	100,00%	1981	3
0,00001	0,45	52,50%	5,00%	100,00%	1981	3
0,00001	0,55	52,35%	4,70%	100,00%	1981	3
0,000001	0,1	54,25%	8,50%	100,00%	1988	3
0,000001	0,15	53,60%	7,19%	100,00%	1988	3
0,000001	0,25	52,60%	5,20%	100,00%	1988	3
0,000001	0,35	52,15%	4,30%	100,00%	1988	3
0,000001	0,45	51,80%	3,59%	100,00%	1988	3
0,000001	0,55	51,60%	3,20%	100,00%	1988	3

Apêndice D

Resultados do treinamento das redes RBF-DDA com as saídas probabilísticas utilizando a base de dados do MNIST

Neste apêndice serão mostrados todos os resultados advindos do treinamento das redes RBF-DDA com as saídas probabilísticas utilizando a base de dados do MNIST.

Comparando a maior saída com todas as outras (Método 1)						
theta-	limiar	acerto total	acerto classes	% de novidades	qtd de Neurônios	Epocas
0,1	0,1	67,90%	43,20%	92,60%	1548	4
0,1	0,15	67,50%	40,50%	94,50%	1548	4
0,1	0,25	65,90%	34,80%	97,00%	1548	4
0,1	0,35	63,60%	28,79%	98,40%	1548	4
0,1	0,45	61,25%	23,59%	98,90%	1548	4
0,1	0,55	59,10%	18,60%	99,60%	1548	4
0,01	0,1	76,95%	80,80%	73,10%	1892	3
0,01	0,15	78,75%	79,30%	78,20%	1892	3
0,01	0,25	80,25%	75,00%	85,50%	1892	3
0,01	0,35	81,60%	71,39%	91,80%	1892	3
0,01	0,45	80,30%	65,60%	95,00%	1892	3
0,01	0,55	78,74%	60,09%	97,39%	1892	3
0,001	0,1	66,35%	91,50%	41,19%	1948	3
0,001	0,15	67,70%	90,40%	45,00%	1948	3
0,001	0,25	71,55%	87,70%	55,40%	1948	3
0,001	0,35	74,00%	85,10%	62,90%	1948	3
0,001	0,45	76,45%	81,10%	71,80%	1948	3
0,001	0,55	78,10%	77,50%	78,70%	1948	3
0,0001	0,1	61,35%	95,00%	27,70%	1971	3
0,0001	0,15	62,90%	94,39%	31,40%	1971	3
0,0001	0,25	65,20%	92,90%	37,50%	1971	3

0,0001	0,35	68,00%	91,10%	44,90%	1971	3
0,0001	0,45	70,75%	88,30%	53,20%	1971	3
0,0001	0,55	73,25%	84,60%	61,90%	1971	3
0,00001	0,1	58,40%	96,70%	20,10%	1981	3
0,00001	0,15	59,85%	96,39%	23,30%	1981	3
0,00001	0,25	62,20%	95,10%	29,29%	1981	3
0,00001	0,35	64,45%	94,10%	34,80%	1981	3
0,00001	0,45	66,30%	91,50%	41,09%	1981	3
0,00001	0,55	69,10%	89,40%	48,80%	1981	3
0,000001	0,1	56,65%	98,00%	15,29%	1988	3
0,000001	0,15	58,15%	97,39%	18,90%	1988	3
0,000001	0,25	60,10%	96,30%	23,90%	1988	3
0,000001	0,35	61,80%	95,19%	28,40%	1988	3
0,000001	0,45	63,85%	93,40%	34,30%	1988	3
0,000001	0,55	65,80%	91,70%	39,90%	1988	3

Comparado a maior saída com a segunda maior (Método 2)

theta-	limiar	acerto total	acerto classes	% de novidades	qtd de Neurônios	Epoas
0,1	0,1	59,30%	71,50%	47,09%	1548	4
0,1	0,15	64,65%	66,90%	62,40%	1548	4
0,1	0,25	67,80%	57,40%	78,20%	1548	4
0,1	0,35	68,15%	47,50%	88,80%	1548	4
0,1	0,45	67,20%	39,20%	95,19%	1548	4
0,1	0,55	64,25%	30,40%	98,10%	1548	4
0,01	0,1	57,10%	86,10%	28,10%	1892	3
0,01	0,15	62,10%	84,39%	39,80%	1892	3
0,01	0,25	70,05%	80,50%	59,59%	1892	3
0,01	0,35	75,95%	76,80%	75,10%	1892	3
0,01	0,45	79,10%	72,20%	86,00%	1892	3
0,01	0,55	80,40%	67,50%	93,30%	1892	3
0,001	0,1	52,90%	88,00%	17,80%	1948	3
0,001	0,15	55,85%	87,40%	24,30%	1948	3
0,001	0,25	61,60%	85,70%	37,50%	1948	3
0,001	0,35	67,30%	84,10%	50,50%	1948	3
0,001	0,45	72,00%	82,39%	61,60%	1948	3
0,001	0,55	74,90%	78,50%	71,30%	1948	3
0,0001	0,1	50,15%	86,70%	13,60%	1971	3
0,0001	0,15	52,45%	86,40%	18,50%	1971	3
0,0001	0,25	56,45%	85,50%	27,40%	1971	3
0,0001	0,35	60,40%	84,60%	36,19%	1971	3
0,0001	0,45	64,35%	83,30%	45,40%	1971	3
0,0001	0,55	68,50%	81,89%	55,10%	1971	3
0,00001	0,1	47,25%	83,89%	10,60%	1981	3
0,00001	0,15	49,40%	83,80%	15,00%	1981	3
0,00001	0,25	52,95%	82,89%	23,00%	1981	3
0,00001	0,35	55,80%	82,50%	29,09%	1981	3
0,00001	0,45	58,85%	81,30%	36,40%	1981	3
0,00001	0,55	62,25%	80,40%	44,10%	1981	3
0,000001	0,1	44,85%	82,10%	7,60%	1988	3
0,000001	0,15	47,00%	82,00%	12,00%	1988	3
0,000001	0,25	50,00%	81,39%	18,60%	1988	3
0,000001	0,35	53,05%	81,10%	25,00%	1988	3

0,000001	0,45	55,30%	80,40%	30,20%	1988	3
0,000001	0,55	58,45%	79,60%	37,30%	1988	3
