Introdução à Redes Neurais Artificiais

Tópicos Especiais em Telecomunicações IV

Prof. João Marcos Meirelles da Silva

www.telecom.uff.br/~jmarcos

Departamento de Engenharia de Telecomunicações

Escola de Engenharia

Universidade Federal Fluminense

Redes de Função de Base Radial

- Introdução
  - O uso da técnica “back-propagation” para o projeto de redes neurais pode ser interpretado como um algoritmo recursivo da Estatística conhecido como Aproximação Estocástica (Haykin,1999).
  - Nesta parte, vamos adotar uma forma diferente de implementar uma rede neural vendo-a como um Problema de Ajuste de Curvas em espaços de grandes dimensões.

$\Rightarrow$ Aprendizado = encontrar uma superfície em um espaço de grandes dimensões que provê o melhor ajuste possível aos dados de treinamento, sendo o “melhor ajuste” medido através de algum modo estatístico.

Redes de Função de Base Radial

- Introdução
  De forma correspondente:
  - Generalização $\equiv$ usar a hiper-superfície encontrada para interpolar os dados de teste !
  - Este ponto de vista é a motivação para o uso de *Funções de Base Radial*.

Redes de Função de Base Radial

- Por que utilizar espaços de grandes dimensões? $\rightarrow$ Teorema de Cover (1965)

"Um problema complexo de classificação de padrões colocado em um espaço de dimensão elevada tem mais chance de ser resolvido linearmente (classes linearmente separáveis) do que quando colocado em um espaço de dimensão baixa."

Redes de Função de Base Radial

- O que é uma Função de Base Radial?
  - Uma função de base Radial (RBF) é uma função real cujo valor depende somente da distância da origem de um sistema de coordenadas, de forma que $\varphi(x) = \varphi(\| x \|)$, ou alternativamente, da distância a partir de um outro ponto c, chamado de centro, de forma que $\varphi(x, c) = \varphi(\| x - c \|)$.
  - $\Rightarrow$ Qualquer função φ que satisfaça a propriedade $\varphi(x) = \varphi(\| x \|)$ é uma função radial !

Redes de Função de Base Radial

- Redes neurais de base radial são redes feedfoward multicamadas cujos neurônios possuem como função de ativação as chamadas funções de base radial;
- Combinação linear de funções radiais;
- Utilizada em aproximação de funções, predição de séries temporais e controle;
- Geralmente, possui 3 camadas: entrada, escondida (não-linear) e saída (linear);
- Pode aproximar qualquer função contínua através de um número suficiente de neurônios na camada escondida.

Redes de Função de Base Radial

Alguns tipos de funções radiais utilizadas ($r = \| w - p \|$):

Gaussiana

$$\varphi(r) = e^{-\beta r^2}, \quad \beta = -\frac{1}{2\sigma^2}, \quad \sigma > 0, \quad r \in \mathbb{R} \qquad (1)$$

Multiquadrática

$$\varphi(r) = \sqrt{r^2 + c^2}, \quad c > 0, \quad r \in \mathbb{R} \qquad (2)$$

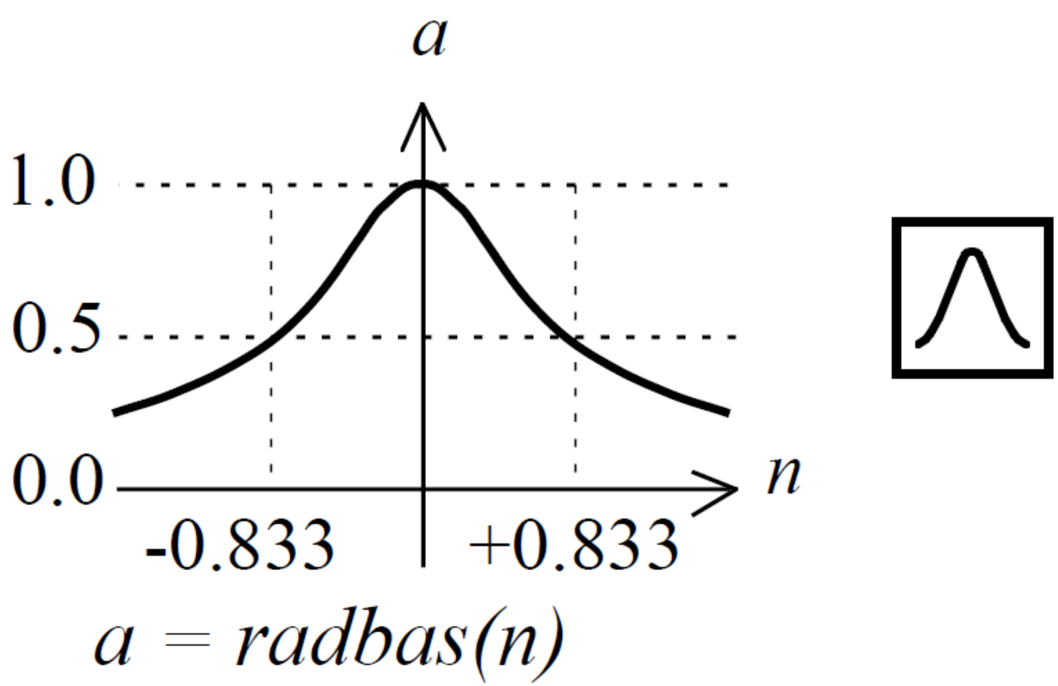
Poliharmônica

$$\varphi(r) = r^k, k = 1, 3, 5, \ldots \qquad (3)$$

$$\varphi(r) = r^k \ln(r), k = 2, 4, 6, \ldots \qquad (4)$$

Redes de Função de Base Radial

Em geral, a mais utilizada é a Gaussiana



Radial Basis Function

Redes de Função de Base Radial

Seja:

- O vetor $x = [x_1, x_2, \ldots, x_N]$, $x \in \mathbb{R}^{M \times N}$
- O vetor $t = [t_1, t_2, \ldots, t_N]$, $t \in \mathbb{R}^{P \times N}$

  onde:
  - M é a dimensão dos vetores de entrada
  - P é a dimensão dos vetores de saída
  - N é o número de vetores

Redes de Função de Base Radial

Dicotomia: Partição binária do espaço:

$$\mathbf{X} = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_N \\ | & | & & | \end{bmatrix} \qquad \mathbf{X} \in \mathbb{R}^{M \times N}$$

$$\mathbf{T} = \begin{bmatrix} | & | & & | \\ \mathbf{t}_1 & \mathbf{t}_2 & \cdots & \mathbf{t}_N \\ | & | & & | \end{bmatrix} \qquad \mathbf{T} \in \mathbb{R}^{P \times N}$$

Dimensão dos Vetores de Entrada

Número de Vetores

Dimensão dos Vetores de Saída

Classe $\mathcal{C}_0$: $\quad \mathcal{C}_0 = \{\mathbf{x}_n | t_n = 0\}$

Classe $\mathcal{C}_1$: $\quad \mathcal{C}_1 = \{\mathbf{x}_n | t_n = 1\}$

Redes de Função de Base Radial

Considere uma família de superfícies onde cada uma naturalmente divide o espaço de entradas em duas regiões. Seja $\mathbb{C}$ um conjunto de N padrões (vetores) $x_1, x_2, \ldots, x_N$, cada qual é atribuído a uma de duas classes $\mathbb{C}_1$ e $\mathbb{C}_2$. Esta *dicotomia* do espaço é dita separável em relação à família de superfícies, se uma superfície existir na família $\mathbb{C}$ de maneira que separe corretamente os pontos da classes $\mathbb{C}_1$ daqueles da classe $\mathbb{C}_2$.

Redes de Função de Base Radial

- Seja $\varphi_i(x) : \mathbb{R}^M \to \mathbb{R}, i = 1, 2, \ldots, I$

$$\varphi(x_n) = \begin{bmatrix} \varphi_1(x_n) \\ \varphi_2(x_n) \\ \vdots \\ \varphi_I(x_n) \end{bmatrix} \tag{5}$$

- $\varphi_i(x)$ recebe o nome de função escondida;
- Espaço gerado pelo conjunto de vetores $x_n \to \varphi(x_n)$: espaço escondido ou espaço de propriedades (feature space).

Redes de Função de Base Radial

- Dizemos que as classes $\mathcal{C}_0$ e $\mathcal{C}_1$ são linearmente separáveis em φ, se existe $w \in \mathbb{R}^I$ tal que:

$$w^T \varphi(x) < 0 \rightarrow x \in \mathcal{C}_0 \tag{6}$$

$$w^T \varphi(x) > 0 \rightarrow x \in \mathcal{C}_1 \tag{7}$$

- No espaço escondido (φ-space), $w^T \varphi(x) = 0$ é um (hiper)plano;
- No espaço $\mathbb{R}^M$ dos vetores x (espaço de entrada), a superfície não-linear $x : w^T \varphi(x) = 0$ é chamada de superfície de separação.

Redes de Função de Base Radial

- Redes de funções de base radial tipicamente possuem três camadas:
  - Camada de entrada
  - camada escondida com funções de ativação não-lineares do tipo funções radiais
  - Camada de saída com função de ativação linear ou sinal (sign).

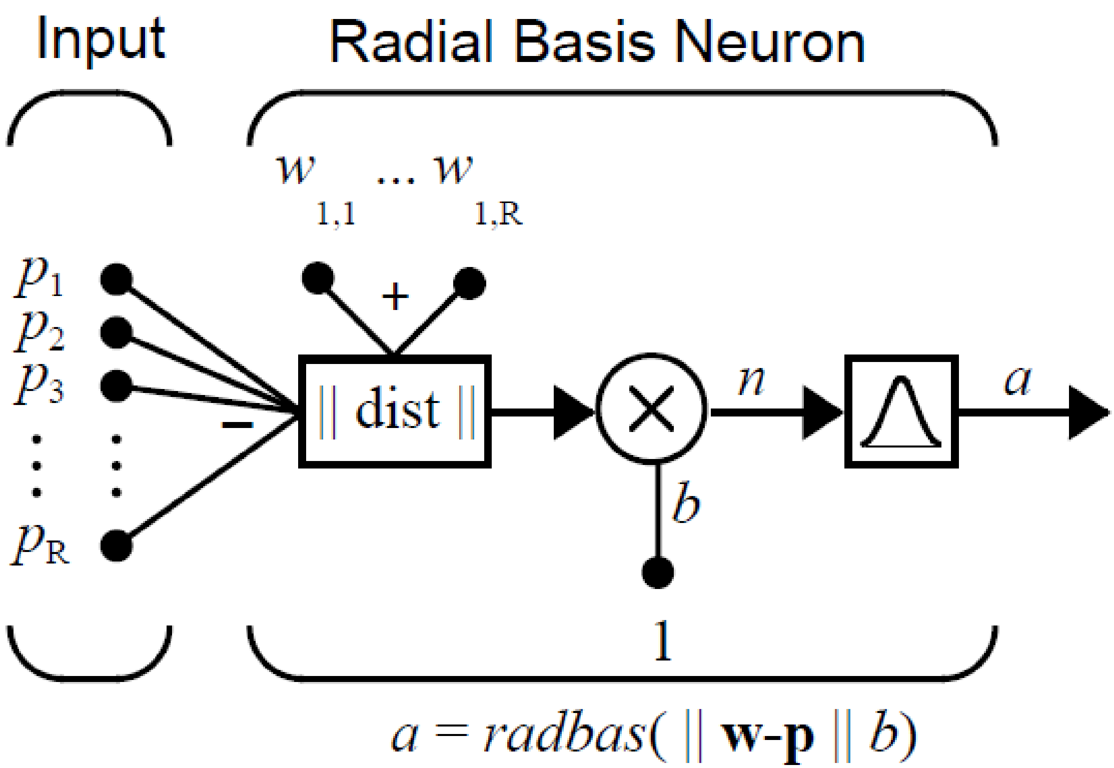
A saída, $\varphi : \mathbb{R}^n \to \mathbb{R}$, da rede é dada por:

$$F(x) = sign(\sum_{i=0}^{I} w_i \varphi(\| x - x_i \|)) \tag{8}$$

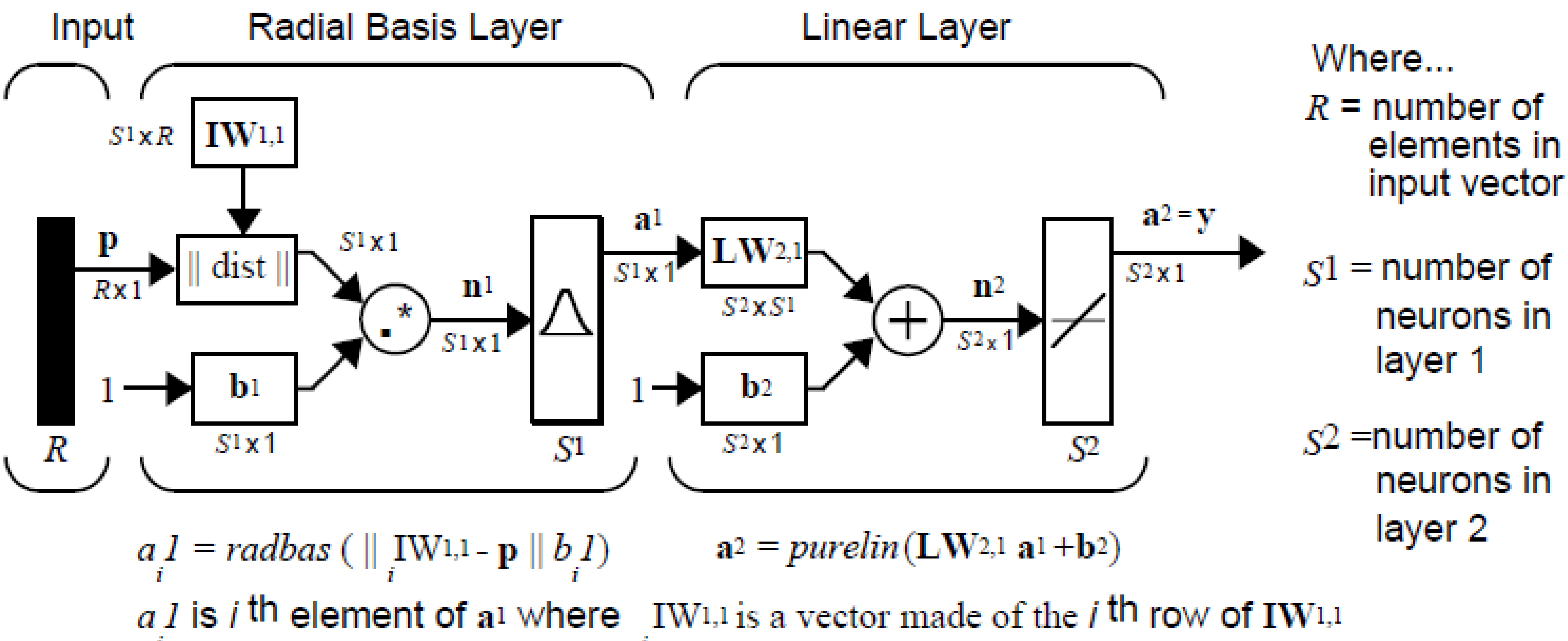
onde N é o número de neurônios da camada escondida, x_i é o vetor centro do neurônio i, e w_i são os pesos do neurônio da camada de saída.

Redes de Função de Base Radial

- Modelo de Neurônio (Matlab)



Redes de Função de Base Radial



Redes de Função de Base Radial

- Exemplo: O problema do OU-Exclusivo revisitado
  - Padrões de entrada $\rightarrow$ saída:
    - $x(1) = [0, 0] \rightarrow y(1) = 0$
    - $x(2) = [0, 1] \rightarrow y(2) = 1$
    - $x(3) = [1, 0] \rightarrow y(3) = 1$
    - $x(4) = [1, 1] \rightarrow y(4) = 0$
  - Defini-se um par de funções escondidas do tipo Gaussiana:
    - $\varphi_1(x) = e^{-e|x-t_1|^2}, \quad t_1 = [1, 1]^T$
    - $\varphi_2(x) = e^{-e|x-t_2|^2}, \quad t_1 = [0, 0]^T$

Redes de Função de Base Radial

$Entrada(x)$	$\varphi_1(x)$	$\varphi_2(x)$
$(1,1)$	1	$0,1353$
$(0,1)$	$0,3678$	$0,3678$
$(0,0)$	$0,1353$	1
$(1,0)$	$0,3678$	$0,3678$

Figura: Plano $\varphi_1 - \varphi_2$

Redes de Função de Base Radial

- Conclusão:

  O uso de Funções de Base Radial não necessariamente aumenta a dimensão do problema no espaço escondido em relação ao espaço de entrada.

  Em outras palavras, a não-linearidade exemplificada pela Função de Base Radial Gaussiana foi suficiente para transformar o problema não separável no plano (x_1, x_2) em um problema linearmente separável no plano (φ_1, φ_2).

Redes de Função de Base Radial

- O Problema da Interpolação
  Dado um conjunto contendo N pontos diferentes $\{x_i \in \mathbb{R}^{m_0} | i = 1, 2, \ldots, N\}$ e um conjunto de N números reais correspondentes $\{t_i \in \mathbb{R} | i = 1, 2, \ldots, N\}$, achar uma função $F : \mathbb{R}^N \rightarrow \mathbb{R}$ que satisfaça a condição de interpolação:

$$F(x_i) = t_i, \qquad i = 1, 2, \ldots, N. \tag{9}$$

OBS: A superfície de interpolação dada por F deve passar obrigatoriamente em todos os pontos x_i de treinamento.

Redes de Função de Base Radial

A técnica das Funções de Base Radial consiste na escolha de uma função F que possui a seguinte forma:

$$F(x) = \sum_{i=1}^{N} w_i \varphi(\| x - x_i \|) \tag{10}$$

onde $\{\varphi(\| x - x_i \|) | i = 1, 2, \ldots, N\}$ é um conjunto arbitrário (geralmente não-linear) de N funções, conhecidas como funções de base radial, $\| \cdot \|$ representa a norma (usualmente Euclidiana).

$\Rightarrow$ Os pontos conhecidos $x_i \in \mathbb{R}^{m_0}, i = 1, 2, \ldots, N$ serão tomados como os centros das funções de base radial.

Redes de Função de Base Radial

$\Rightarrow$ A norma utilizada é normalmente a distância Euclidiana, apesar de existirem outras como a métrica de Lukaszyk-karmowski (evitar problemas de condicionamento de matrizes).

$\Rightarrow$ Somas de funções radiais são empregadas para aproximar funções (fundamento da rede neural de base radial).

Redes de Função de Base Radial

Inserindo as condições de interpolação da equação (9) em (10), obtemos o conjunto de equações lineares:

$$\begin{bmatrix} \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1N} \\ \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{N1} & \varphi_{N2} & \cdots & \varphi_{NN} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix} \qquad (11)$$

onde

$$\varphi_{ji} = \varphi(\| x_j - x_i \|), \qquad (j, i) = 1, 2, \ldots, N \qquad (12)$$

Redes de Função de Base Radial

Seja

$$t = [t_1, t_2, \ldots, t_N]^T \tag{13}$$

$$w = [w_1, w_2, \ldots, w_N]^T \tag{14}$$

onde t representa o vetor de respostas desejadas e w representa o vetor de pesos.

Podemos reescrever a equação (11) como:

$$\Phi w = x \tag{15}$$

Redes de Função de Base Radial

É possível mostrar que a matriz de interpolação Φ é não-singular, desde que os pontos x_i sejam distintos (*Teorema de Micchelli*). Logo, podemos resolver a equação (15) através de:

$$w = \Phi^{-1}x \tag{16}$$

Redes de Função de Base Radial

- Treinamento[a]

Em uma rede RBF, existem três parâmetros que devem ser escolhidos para adaptar a rede a uma aplicação em particular:

1. Os vetores de centro x_i;
2. Os pesos das sinapses w_k;
3. Parâmetros de largura β_j.

[a] O treinamento neste caso é a solução de um sistema linear de equações.

Redes de Função de Base Radial

- Funções do Matlab

nnet = newrbe(P, T, spread) Cria uma rede base radial com matriz de vetores de entrada P, matriz de vetores de saída T e $\sigma^2 =$ spread;

Redes de Função de Base Radial

- Demonstrações
  - demorb1: Demonstra como uma rede de base radial é utilizada para aproximar uma função;
  - demorb3, demorb4: Demonstra como a constante de espalhamento (spread) afeta o desempenho da rede de base radial;

OBS: É importante escolher uma constante de espalhamento maior que a distância entre vetores de entrada adjacentes, a fim de obtermos uma boa generalização, mas menor do que a distância através de todo o espaço de entrada.

Redes de Função de Base Radial

- Demonstrações

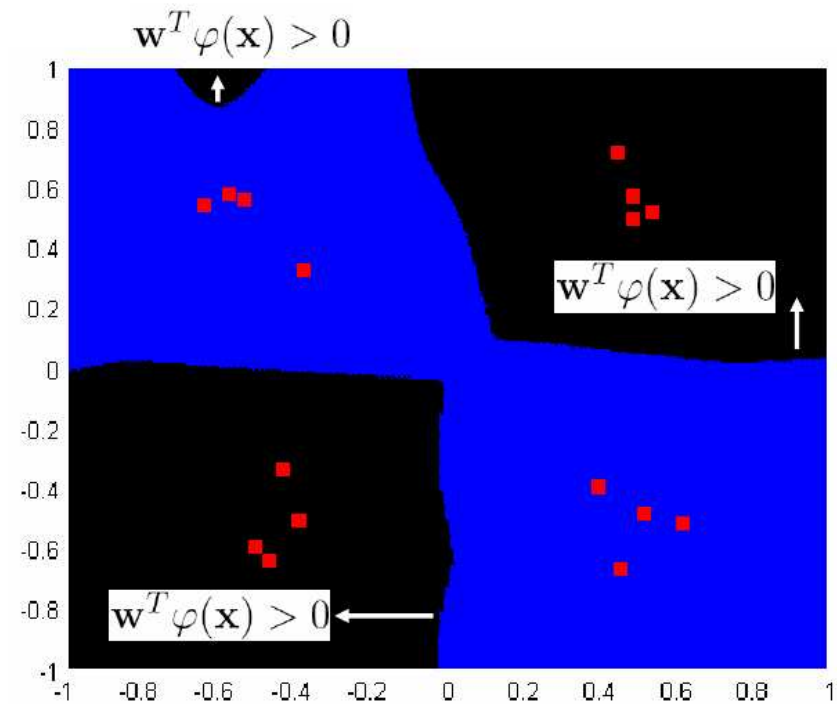
```
close all; clear all;
C = [0.5 0.5 -0.5 -0.5 ; -0.5 0.5 -0.5 0.5];
t = [-1 1 1 -1];
randn('state',0);
E = repmat(C(:,1),1,4) + 0.1*randn(2,4);
F = repmat(C(:,2),1,4) + 0.1*randn(2,4);
G = repmat(C(:,3),1,4) + 0.1*randn(2,4);
H = repmat(C(:,4),1,4) + 0.1*randn(2,4);
X = [E F G H]; t = [-1 -1 -1 -1 1 1 1 1 1 1 1 1 -1 -1 -1 -1];
% P = 1;
P = 100; for i = 1:16, for j = 1:16, Phi(i,j) = exp(-P*norm(X(:,i)-X(:,j))^2); end; end;
w = inv(Phi)*t'
X = zeros(2,10000);
for i = 1:100,
   for j = 1:100,
      X(:,100*(i-1)+j) = [(i-50.5)/50 ; (j-50.5)/50];
   end;
end;
Y = [E F G H];
figure; hold on;
for n=1:size(X,2),
o = exp(-P*sum((Y-repmat(X(:,n),1,16)).^2,1))*w;
if o < 0,
plot(X(1,n),X(2,n),'b.');
else
plot(X(1,n),X(2,n),'k.');
end;
end;
plot(Y(1,:),Y(2,:),'r.');
```

Redes de Função de Base Radial

- Demonstrações

$\beta = 100$

$$\mathbf{w} = \begin{bmatrix} -0.99 \\ -0.67 \\ -0.94 \\ -0.80 \\ 0.54 \\ 0.41 \\ 0.95 \\ 0.27 \\ 0.82 \\ 0.50 \\ 0.60 \\ 0.94 \\ 0.44 \\ -1.00 \\ -0.86 \\ -1.10 \end{bmatrix}$$



Redes de Função de Base Radial

- Maiores Informações:

  Para uma descrição mais completa sobre as redes neurais de base radial, ver livro do Haykin !

RBF $\times$ Perceptrons de Multicamadas

- Ambas as redes são do tipo feedfoward
- Ambas são utilizadas como aproximadores universais
- De fato, é sempre possível substituir uma pela outra;

No entanto:

- A RBF possui uma única camada escondida, enquanto as MLP podem possuir várias;
- A RBF possui uma capacidade de generalização menor, porém sua construção é mais rápida
- Em geral, utiliza mais neurônios que as MLP treinadas por backpropagation.

Exemplo de Aplicação

Real-Time Detection of Distributed Denial-of-Service Attacks Using RBF Networks and Statistical Features