

-
-
-
-
-

Redes de Função de Base Radial

Aluizio Fausto Ribeiro Araújo
Universidade Federal de Pernambuco
Centro de Informática



-
-
-
-
-

Conteúdo

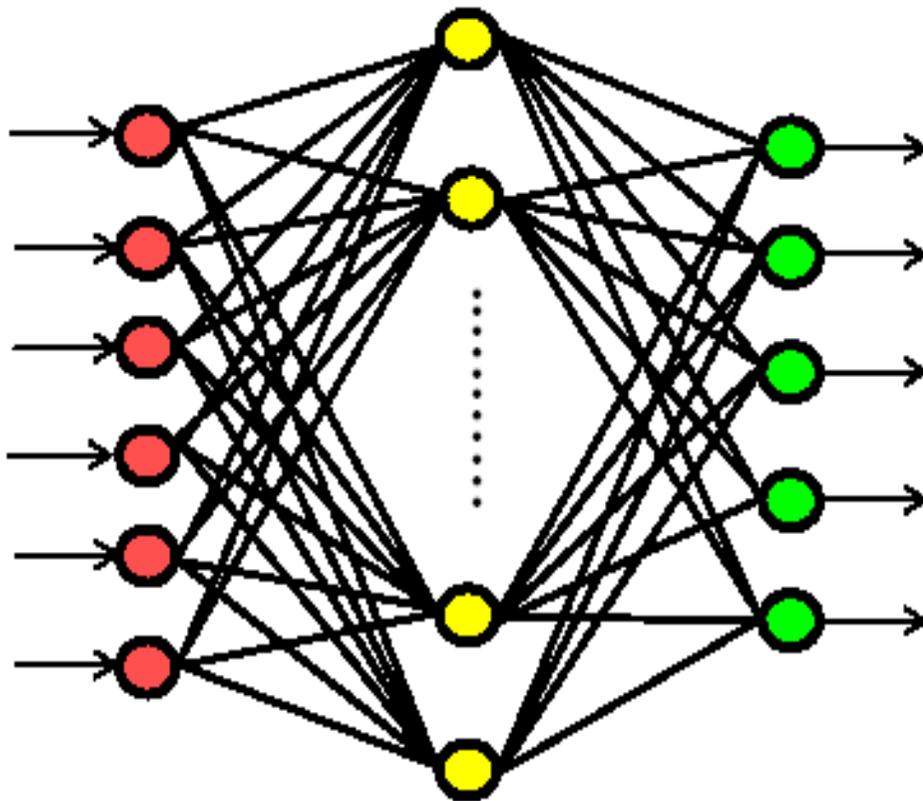
- Introdução
- Histórico
- Teoria que Fundamenta Redes RBF
- Redes de Função de Base Radial
- Aprendizagem em Redes RBF
- Desempenho das Redes RBF
- Comparando MLP-BP e RBF

Introdução

- Este capítulo trata de redes neurais com múltiplas camadas que não são treinadas por retropropagação (*backpropagation*) e que não têm unidades de processamento com função de ativação do tipo sigmoidal.
- Estas redes utilizam unidades com campos receptivos locais (*local receptive fields*), nos quais as unidades que recebem diretamente estímulos de entrada podem responder apenas a parte destas entradas.
- Esta abordagem emprega, na maioria dos casos, treinamento supervisionado e não-supervisionado.
- As redes são muito usadas como interpoladores/ aproximadores e em tarefas de classificação.

Introdução

- Arquitetura típica de uma rede RBF



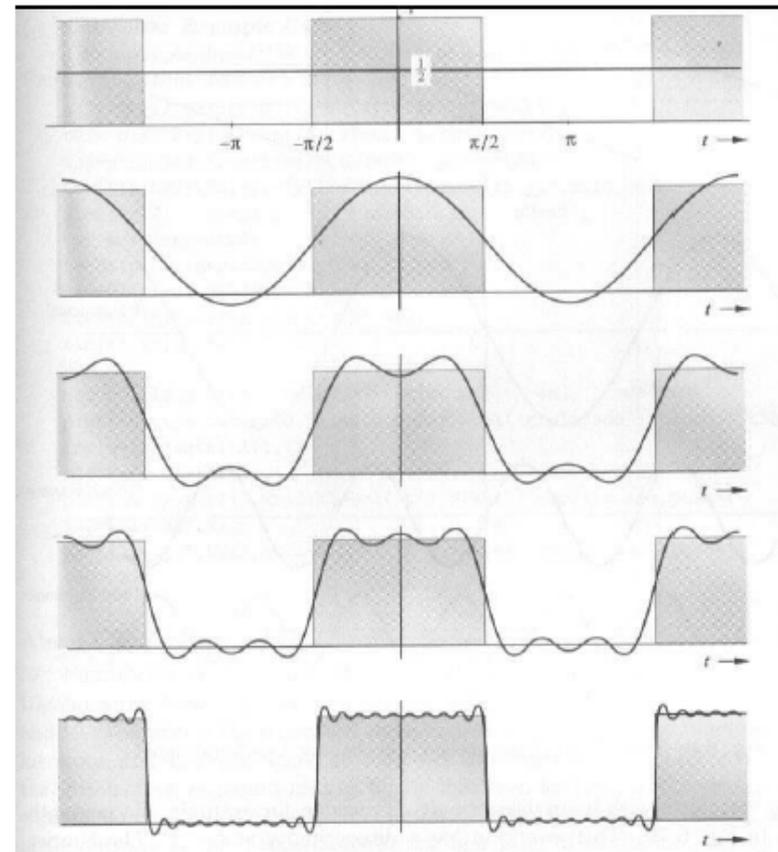
- Camada de entrada: Realiza propagação da entrada.
- Camada escondida: Realiza transformação não-linear, considerando campos receptivos locais.
- Camada de saída: Realiza combinação linear

Introdução

- Em uma rede MLP-RP, a regra de ativação de cada unidade da camada escondida é definida pelo produto interno do vetor de entrada pelo vetor de pesos.
- Em redes RBF, a regra de ativação das unidades da camada escondida é definida por uma função de base radial para a distância entre o vetor de entrada e um vetor de protótipo.
- Redes RBF podem ser caracterizadas por:
 - Proposição que é associada com outras disciplinas tais como aproximação de funções, teoria da regularização, estimação de densidade e interpolação na presença de ruídos (Bishop, 1995).
 - Permitirem interpretação clara das representações internas da camada escondida.
 - Treinamento, em geral, muito mais rápido que as MLP-BP.

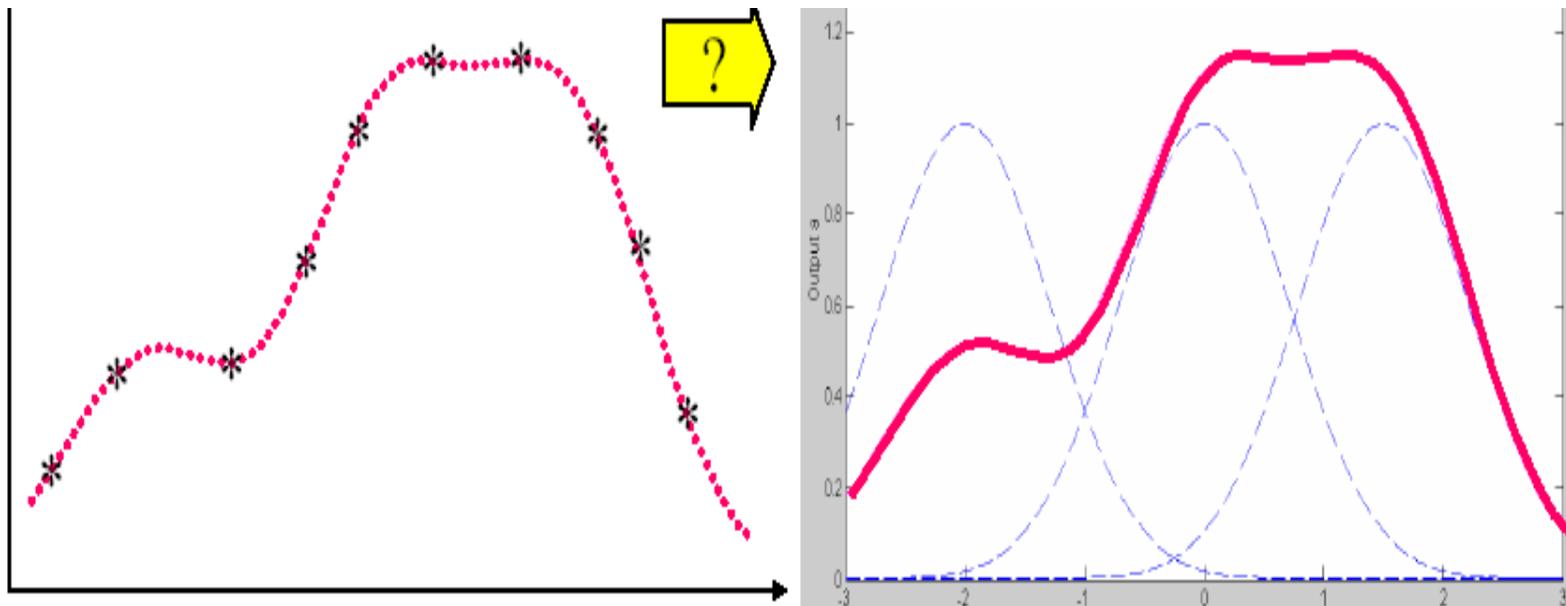
Introdução

- Construtores de funções complexas a partir de funções simples:
 - Série de Fourier
 - Transformada de Fourier
 - Transformada Wavelet
 - Redes RBF



Introdução

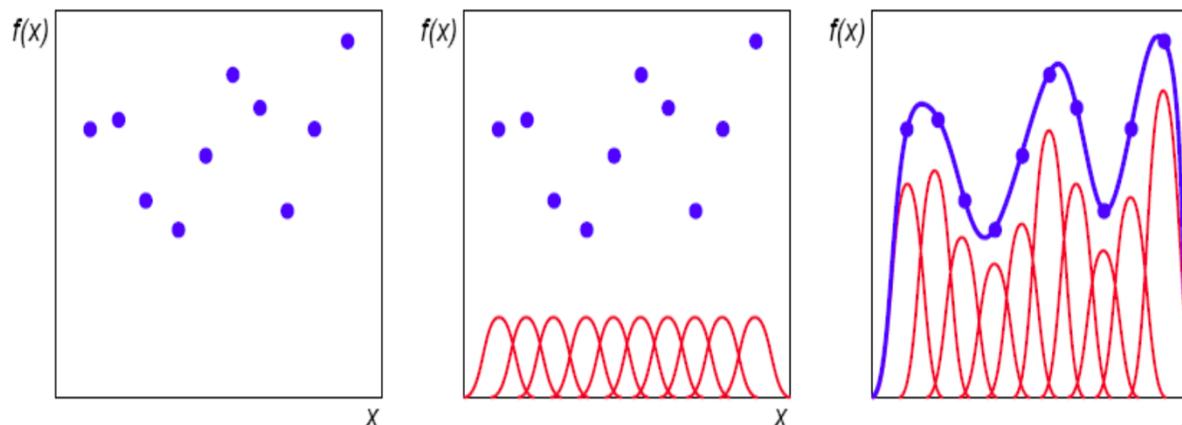
Redes RBF aproximam funções empregando funções de base radial como blocos construtores.



Introdução

- Redes RBF têm suas origens em técnicas para realizar interpolação exata de funções (Bishop, 1995).
- Abordagem RBF (Powell, 1987):
 - Empregue um conjunto de N funções de base, não-lineares, para calcular a função:

$$h(\mathbf{X}) = \sum_i w_i \Phi(\|\mathbf{X} - \mathbf{\Lambda}_i\|).$$



Histórico

- Esta abordagem é inspirada na propriedade de alguns neurônios biológicos chamada de resposta localmente sintonizada (*locally tuned response*). Tais células nervosas respondem seletivamente a um intervalo finito do espaço de sinais de entrada.
- Aprendizagem envolve encontrar uma superfície em um espaço de dimensão qualquer que produza o melhor ajuste (represente da melhor maneira) os dados de treinamento.

Histórico

- O primeiro trabalho lidando com funções de base radial foi introduzido por Medgassy (1961) cujos resultados foram posteriormente usados para interpolação (Micchelli, 1986; Powell, 1987), para estimação de densidade (Parzen, 1962; Duda e Hart, 1973; Specht, 1990) e para aproximação de funções multivariadas suaves (*smooth multivariate functions*) (Poggio and Girosi, 1989).

Histórico

- Atualmente, os modelos de RBF se diferenciam dos primeiros pois são de natureza adaptativa que permite a utilização, em muitas situações, de um número relativamente menor de unidades de processamento localmente sintonizadas.
- Redes RBF foram independentemente propostas por Broomhead e Lowe (1988), Lee e Kil (1988), Niranjana e Fallside (1988) e Moody e Darken (1989a, 1989b). Outros esquemas similares foram introduzidos por Hanson e Burr (1987), Lapedes e Faber (1987), Casdagli (1989), Poggio e Girosi (1990b), entre outros.

Teoria que Fundamenta as Redes RBF

- Teorema de Cover (1965) sobre a separabilidade de padrões: Um problema complexo de classificação de padrões tratado não-linearmente em um espaço de dimensão alta é mais provável de ser linearmente separável que em um espaço de dimensão baixa:
 - Problemas dicotômicos, com apenas duas classes (C_1 e C_2).
 - A dicotomia é separável se existir um vetor m -dimensional tal que: $\mathbf{W}_t \Phi(\mathbf{X}) > 0, \mathbf{X} \in C_1$ ou $\mathbf{W}_t \Phi(\mathbf{X}) < 0, \mathbf{X} \in C_2$
- Superfícies de separação: Hiperplanos, hiperesferas, ...

Teoria que Fundamenta as Redes RBF

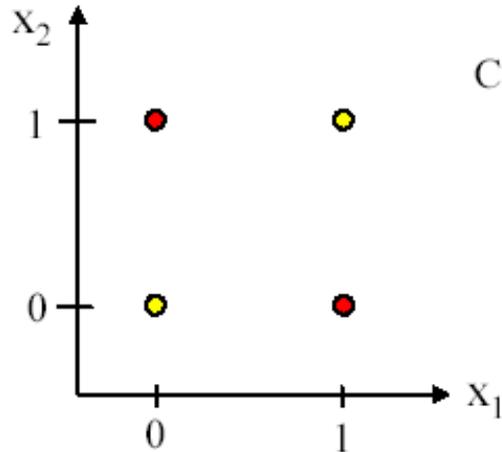
- Problema de Interpolação com Rede RBF:
 - Considere uma rede com alimentação direta de 3 camadas (entrada, escondida, saída);
 - A rede realiza mapeamento não-linear do espaço de entrada para o espaço escondido seguido por mapeamento linear do espaço escondido para o de saída;
 - A fase de treinamento consiste na otimização de um procedimento de ajuste (*fitting*) para uma superfície contruída com base nas amostras de treinamento;
 - A fase de generalização corresponde à interpolação de novos pontos não empregados durante o treinamento.

Teoria que Fundamenta as Redes RBF

- Teorema de Michelli (1986): Seja $\{\mathbf{x}_i\}$, $i=1, \dots, N$ um conjunto de pontos distintos um espaço d -dimensional. Então a matriz de interpolação Φ , com dimensão $N \times N$, cujos elementos são obtidos de uma função de base radial, $\varphi_{ij} = \varphi(\|\mathbf{x}_j - \mathbf{x}_i\|)$ é não-singular, logo é inversível.
 - O teorema é válido para qualquer valor de N , do tipo de função de base radial empregada e da natureza dos pontos.
- O procedimento de interpolação estrito não é uma boa estratégia de treinamento pois resulta em baixa capacidade de generalização.
- Aprendizagem é vista como um problema de reconstrução de hiperespaço, dado um conjunto de pontos que pode ser esparso.
- Dois problemas relacionados são ditos serem inversos se a formação de cada um deles requer conhecimento parcial ou total do outro.



Teoria que Fundamenta as Redes RBF

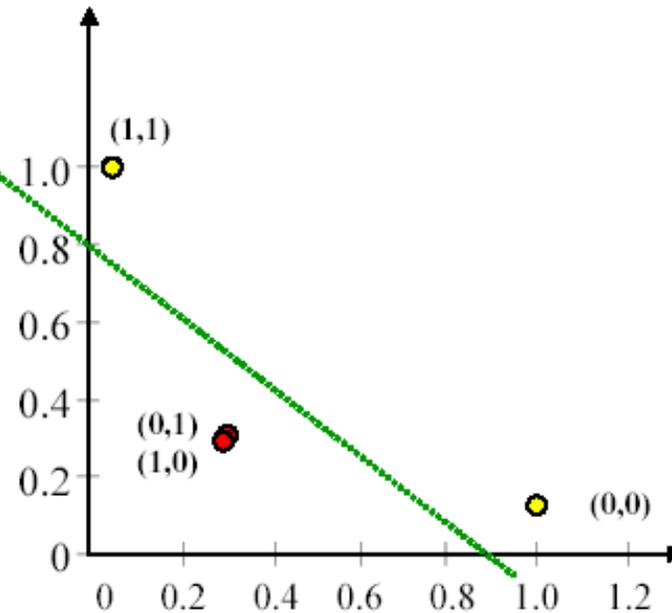
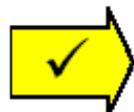


Consider the nonlinear functions to map the input vector \mathbf{x} to the ϕ_1 - ϕ_2 space

$$\mathbf{x} = [x_1 \ x_2] \quad \phi_1(\mathbf{x}) = e^{-\|\mathbf{x} - \mathbf{t}_1\|^2} \quad \mathbf{t}_1 = [1 \ 1]^T$$

$$\phi_2(\mathbf{x}) = e^{-\|\mathbf{x} - \mathbf{t}_2\|^2} \quad \mathbf{t}_2 = [0 \ 0]^T$$

Input \mathbf{x}	$\phi_1(\mathbf{x})$	$\phi_2(\mathbf{x})$
(1,1)	1	0.1353
(0,1)	0.3678	0.3678
(1,0)	0.3678	0.3678
(0,0)	0.1353	1

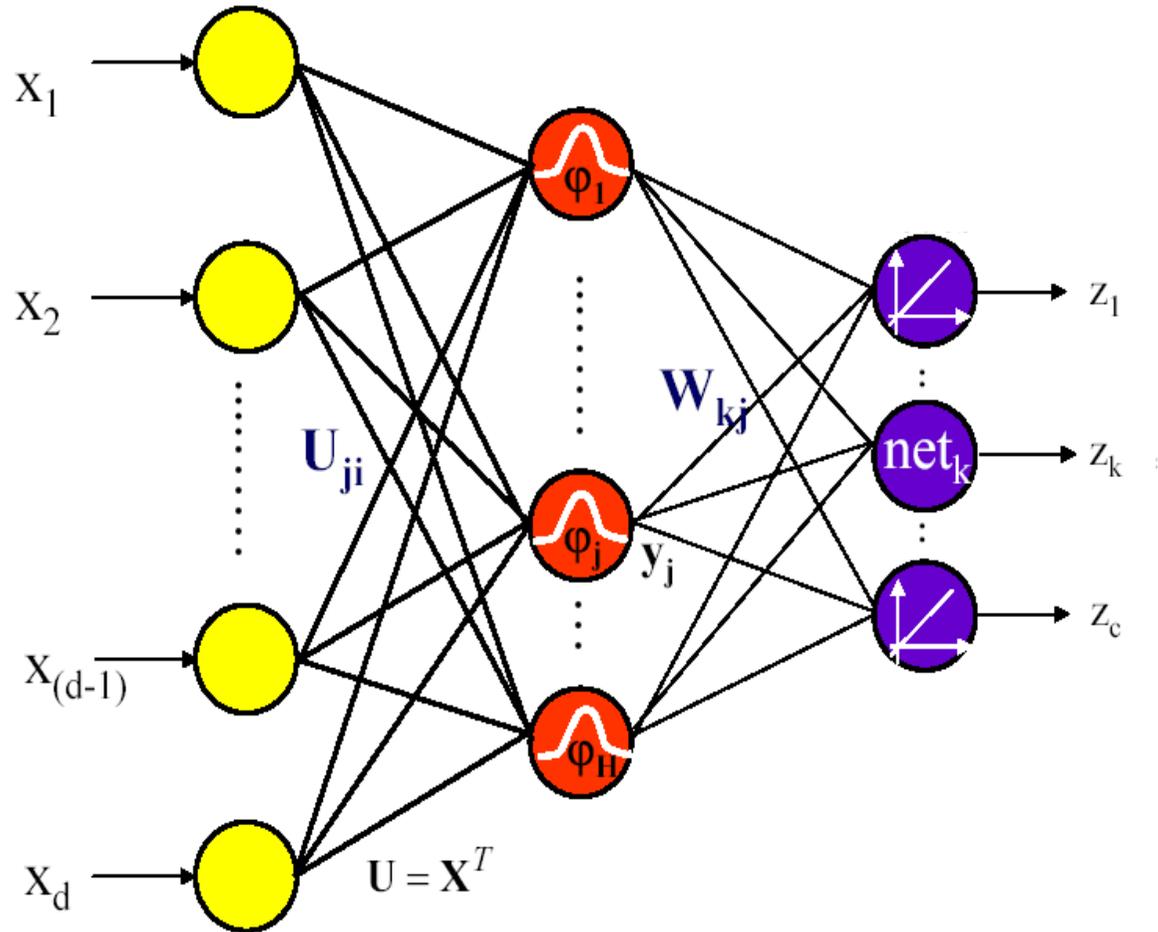


The nonlinear ϕ function transformed a nonlinearly separable problem into a linearly separable one !!!

Rede de Função de Base Radial - RBF

- As redes RBF são redes de alimentação direta (*feedforward*) consistindo tipicamente de três camadas: entrada, escondida e saída.
- Camada de entrada: propaga os estímulos.
- Camada escondida: unidades de processamento localmente sintonizáveis .
- Camada de saída: unidades de processamento lineares.

Redes RBF: Arquitetura e Funções



Redes RBF: Processamento

- Unidades escondidas recebem o vetor de entrada
- Camada de entrada: propaga os estímulos.
- Camada escondida: unidades de processamento localmente sintonizáveis .

$$y_j(\mathbf{x}_p) = F_j(\mathbf{x}_p, \boldsymbol{\mu}_j, \sigma_j)$$

onde $F_j(\cdot)$ é uma função de base radial,

$\boldsymbol{\mu}_j$ é o j -ésimo centro e

σ_j é a largura do campo receptivo para o centro.

Redes RBF: Processamento

- Camada de saída: unidades de processamento lineares.

$$z_l(\mathbf{x}_p) = \sum_{j=1}^H w_{lj} y_j(\mathbf{x}_p)$$

onde w_{lj} é o peso entre a unidade escondida j e a de saída l .

- Redes RBF realizam aproximação de uma função por superposição de funções de base radial não-ortogonais que têm forma de sino. O grau de precisão pode ser controlado por três parâmetros: o número de funções de base usadas, sua localização e sua largura do campo receptivo.

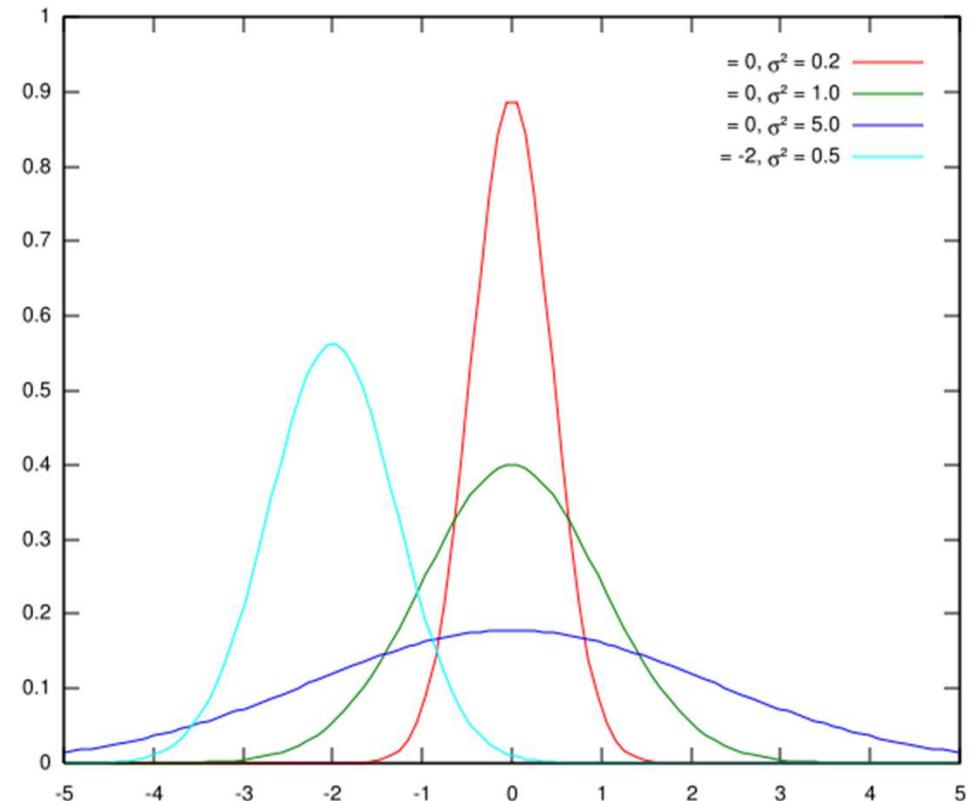
Redes RBF: Funções de Base

- Função Gaussiana

$$y_j(\mathbf{x}_p) = \exp\left(-\frac{(\mathbf{x}_p - \boldsymbol{\mu}_j)^2}{2\sigma_j^2}\right), \text{ onde}$$

$\boldsymbol{\mu}_j$ é o j -ésimo centro e

σ_j é a largura do campo receptivo.



Redes RBF: Funções de Base

- Função Multiquadrática

$$y_j(\mathbf{x}_p) = \left((\mathbf{x}_p - \boldsymbol{\mu}_j)^2 + \sigma_j^2 \right)^{1/2}$$

onde $\boldsymbol{\mu}_j$ é o j -ésimo centro e σ_j é a largura do campo.

- Função multiquadrática inversa

$$y_j(\mathbf{x}_p) = \exp\left(\frac{1}{\left((\mathbf{x}_p - \boldsymbol{\mu}_j)^2 + \sigma_j^2 \right)^{1/2}} \right)$$

onde $\boldsymbol{\mu}_j$ é o j -ésimo centro e σ_j é a largura do campo receptivo.

- Redes RBF são adequadas para aproximar mapeamentos de valores reais contínuos, ou contínuos por partes, para suficientemente pequeno. Problemas de classificação se constituem em um caso particular daquela classe de problemas.

Redes RBF: Funções de Saída

- Função Linear nas Unidades de Saída

$$z_k(\mathbf{x}_p) = f(\text{net}_k) = f\left(\sum_{j=1}^H w_{kj} y_j\right) = \sum_{j=1}^H w_{kj} y_j, \text{ onde}$$

H é o número de unidades escondidas

w_{kj} é o peso entre as unidades j e k .

Aprendizagem em Redes RBF

- O método de treinamento deve reduzir o erro na saída a valores aceitáveis por adaptação dos parâmetros livres na rede RBF: centros, larguras dos campos receptivos e pesos entre camada escondida e de saída. A aprendizagem pode ser supervisionada, não-supervisionada ou combinada.
- O treinamento que combina aprendizagem não-supervisionada (ANS) com aprendizagem supervisionada (AS) é o mais comum pois não se sabe as saídas desejadas para a camada escondida.
 - ANS determina centros e campos receptivos da camada escondida.
 - AS determina os valores dos pesos entre as camadas escondida e de saída, considerando constantes os parâmetros já definidos.

Aprendizagem em Redes RBF

- **Treinamento Não-supervisionado**

- Os centros são selecionados para se casar com a distribuição dos exemplos de treinamento no espaço de características de entrada.
- Os valores dos centros podem ser determinados por (a) seleção aleatória; (b) distribuição sobre uma grade regular; (c) técnica de agrupamento (*clustering*); (d) estimação da densidade; ou (e) outro algoritmo.
- As larguras dos campos receptivos são determinadas empregando (a) distância euclidiana média entre centros; (b) distância euclidiana entre centro e vetor de entrada; (c) distância euclidiana entre centros; ou (d) distância euclidiana entre os centros determinados pelo método k-médias.

Treinamento Não-supervisionado

- Determinação do centro por seleção aleatória :
 - Os centros são vetores de entrada aleatoriamente selecionados. Eles devem representar todo o espaço de soluções do problema. Este método é simples e direto, no entanto pode exigir grande número de unidades intermediárias, escolher centros muito próximos uns dos outros e podem acarretar funcionamento inadequado da rede.
- Determinação do centro por grade regular:
 - Os centros são fixados em uma grade regular, cobrindo todo o espaço de entrada. Em geral, método exige muitas unidades intermediárias para vetores de entrada com dimensão alta causando crescimento exponencial do número das unidades escondidas).

Treinamento Não-supervisionado

- Determinação do centro por agrupamento:
 - Algoritmo das k-médias: Divide os padrões de treinamento em grupos, encontrando o ponto central de cada um deles:

$$\boldsymbol{\mu}_j = \frac{1}{nv_j} \sum_{\mathbf{x}_p \in S_j} \mathbf{x}_p$$

onde nv_j é o número de vetores contido no agrupamento S_j .

- Mapas Auto-organizáveis (SOMs) agrupam padrões espacialmente próximos que compartilhem micro-características. No início, os centros são aleatoriamente atribuídos. O centro que resultar maior produto escalar com um vetor de entrada adiciona uma versão ponderada deste vetor de entrada ao seu grupo .

Treinamento Não-supervisionado

- Determinação do centro por estimação de diversidade:
 - A posição dos centros podem ser determinadas através de modelagem da densidade do espaço de características com um modelo de misturas de gaussianas empregando o algoritmo de maximização da expectativa.
 - Os parâmetros de espalhamento para cada centro são automaticamente obtidos das matrizes de covariância dos componentes gaussianos correspondentes.

Treinamento Não-supervisionado

- Determinação da largura do campo receptivo por distância euclidiana média entre centros:

$$\sigma = \frac{1}{ng} \sum_{j=1}^{ng} \|\boldsymbol{\mu}_j - \boldsymbol{\mu}_{j(mprox)}\|$$

onde $\boldsymbol{\mu}_{j(mprox)}$ é o centro com menor distância euclidiana para $\boldsymbol{\mu}_j$.

ng é o número de grupos que serão formados.

- Determinação da largura do campo receptivo por distância euclidiana entre centro e vetor de entrada:

$$\sigma_j^2 = \frac{1}{nv} \sum_{\mathbf{x}_p \in \Psi_j} \|\boldsymbol{\mu}_j - \mathbf{x}_p\|^2$$

onde Ψ_j é o conjunto dos nv vetores mais próximos a $\boldsymbol{\mu}_j$.

Treinamento Não-supervisionado

- Determinação da largura do campo receptivo por distância euclidiana entre centros:

$$\sigma = \alpha \|\boldsymbol{\mu}_j - \boldsymbol{\mu}_{j(mprox)}\|$$

onde $1 \leq \alpha \leq 1,5$.

- Determinação da largura do campo receptivo por distância euclidiana entre os centros determinados pelo método k-médias:

$$\sigma_j^2 = \frac{1}{na} \sum_{\mathbf{x}_p \in S_j} \|\boldsymbol{\mu}_j - \mathbf{x}_p\|^2$$

onde S_j é o agrupamento contendo na vetores de entrada.

Aprendizagem em Redes RBF

- **Treinamento Supervisionado**

- Os pesos entre a camada escondida e a de saída são calculados de modo a minimizar o erro entre a saída desejada e a saída obtida.
- Os pesos são determinados por um método que resolva o problema de minimização do erro: (a) método dos mínimos quadrados; (b) método da regra delta; ou (c) matriz pseudo-inversa.

Treinamento Supervisionado

- Esta etapa compreende a determinação dos pesos entre a camada escondida e a de saída. Inicia-se pelo cálculo do erro que é função da resposta dada pela rede comparada com a resposta que se deseja dela. Há algumas maneiras de calcular o erro.
 - Soma dos erros quadráticos (SSE – sum of squared error)

$$SSE = \sum_{j=1}^{npad} \left\| \mathbf{z}_d^{(i)} - \mathbf{z}_o^{(i)} \right\|^2$$

- Erro quadrático médio (MSE - mean squared error)

$$MSE = \frac{1}{npad} \sum_{j=1}^{npad} \left\| \mathbf{z}_d^{(i)} - \mathbf{z}_o^{(i)} \right\|^2$$

Treinamento Supervisionado

- Erro relativo médio (MRE – mean relative error)

$$MRE = \frac{1}{npad} \sum_{j=1}^{npad} \left\| \frac{\mathbf{z}_d^{(i)} - \mathbf{z}_o^{(i)}}{\mathbf{z}_d^{(i)}} \right\|^2$$

- Raiz do erro quadrático médio (RMSE – root mean squared error)

$$RMSE = \sqrt{\frac{1}{npad} \sum_{j=1}^{npad} \left\| \mathbf{z}_d^{(i)} - \mathbf{z}_o^{(i)} \right\|^2}$$

onde $\mathbf{z}_d^{(i)}$ e $\mathbf{z}_o^{(i)}$ são o i -ésimo padrão desejado e o obtido e

$npad$ é o número total de padrões.

- Depois de calculado o erro, este é minimizado por procedimentos tais como a regra delta, o método dos mínimos quadrados, e a matriz pseudo-inversa (os dois últimos, métodos lineares).

Estratégias de Treinamento

- Busca-se o compromisso entre precisão e generalização. Para tal, duas estratégias de treinamento podem ser empregadas: *hold-out* e validação cruzada (*crossvalidation*).
 - *Hold-out*: O conjunto de padrões é dividido em três grupos: treinamento, validação e teste. Cada topologia com seus centros tem seu desempenho testado com respeito aos três conjuntos.
 - *Crossvalidation*: Todos os padrões (em geral poucos) são considerados para treinamento. Assim, divide-se os padrões em ng grupos, seleciona-se aleatoriamente $ng-1$ conjuntos para treinamento e testa-se a rede com o conjunto não selecionado. Repete-se o processo até todos os conjuntos serem usados para testes. A partir daí, calcula-se o erro (E).

$$E = \frac{1}{ng} \sum_{j=1}^{ng} E_j^2$$

Desempenho das Redes RBF

- As redes RBF foram aplicadas com sucesso na aproximação de funções (Broomhead e Lowe, 1988; Lee e Kil, 1988; Casdagli, 1989; Moody e Darken, 1989a, 1989b) e em problemas de classificação (Nirajan e Fallside, 1988; Nowlan, 1990; Lee, 1991; Wettschereck e Dieterich, 1992; Vogt, 1993).
- Em tarefas difíceis de aproximação/interpolação (por exemplo, predição da série caótica de Mackey-Glass T instantes de tempo no futuro, $T > 50$), redes RBF que empregam a técnica de agrupamento no posicionamento dos campos receptivos podem alcançar desempenho comparável ao das redes MLP-BP, enquanto requerem tempo de treinamento algumas ordens de grandeza menor.

Desempenho das Redes RBF

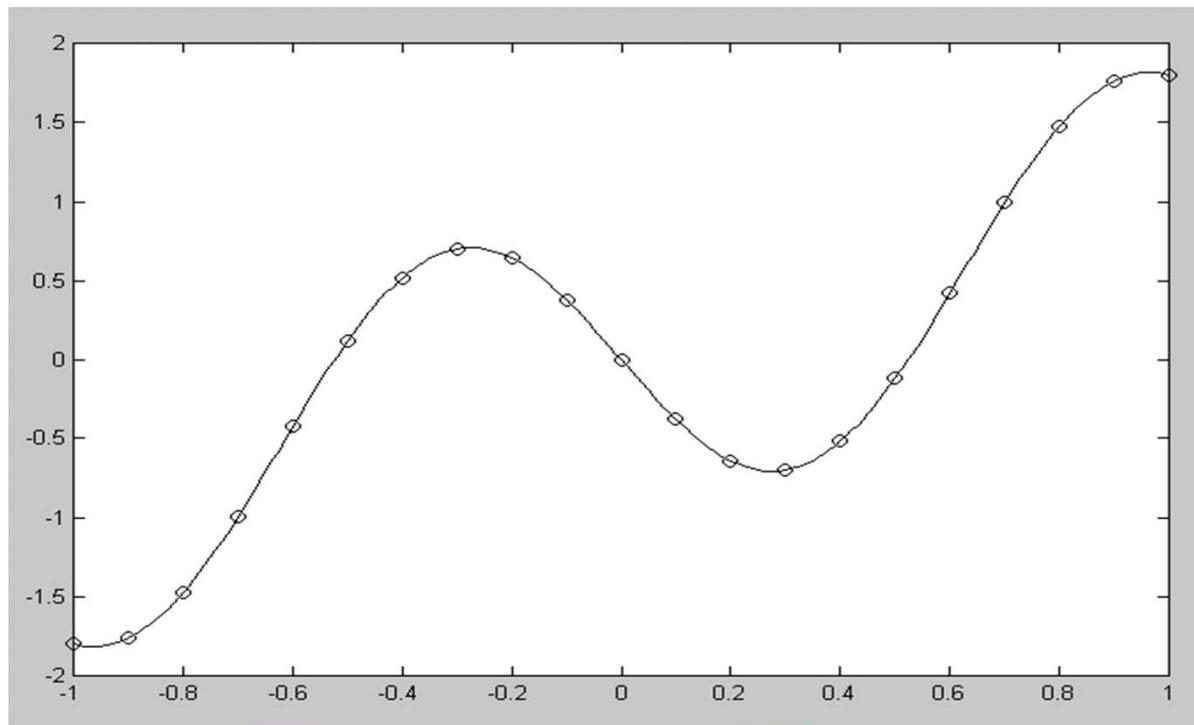
- Redes RBF geralmente precisam de ao menos dez vezes mais dados de treinamento para atingir a acuracidade das redes MLP-BP. Em tarefas difíceis de classificação, redes RBF com número suficiente de padrões de treinamento e de unidades escondidas podem obter melhores taxas de classificação e menores erros de classificações positivas falsas que redes MLP-BP.
- As redes RBF têm tempo de treinamento muito menor porque:
 - Apenas uma pequena fração de unidades escondidas responde a um dado padrão de entrada, pois são unidades localmente sintonizáveis, sensíveis apenas a padrões próximos de seus campos receptivos.
 - Todas as unidades de uma rede MLP-BP são avaliadas e têm seus pesos ajustados para cada vetor de entrada.
 - Empregam treinamento separado na camada escondida e na de saída.

Desempenho das Redes RBF

- Quando utilizadas em aproximação de funções as redes MLP-BP tendem a obter maior capacidade de generalização (ou extrapolação) que as RBF pois ajustam globalmente os padrões de entrada enquanto estas últimas fazem um ajuste local.
- Pelo mesmo motivo, em problemas de classificação, redes RBF tendem a cometer menos erros de falsa classificação positiva do que as redes MLP-BP.
- Pode-se dizer que, em geral, é melhor o uso de redes MLP-BP quando os padrões de entrada são custosos (ou difíceis de se gerar) e/ou quando a velocidade de recuperação é crítica. No entanto, se os dados são baratos e abundantes, e se é necessário treinamento *on-line*, então as redes RBF são superiores.

Exemplo de Emprego de Redes RBF

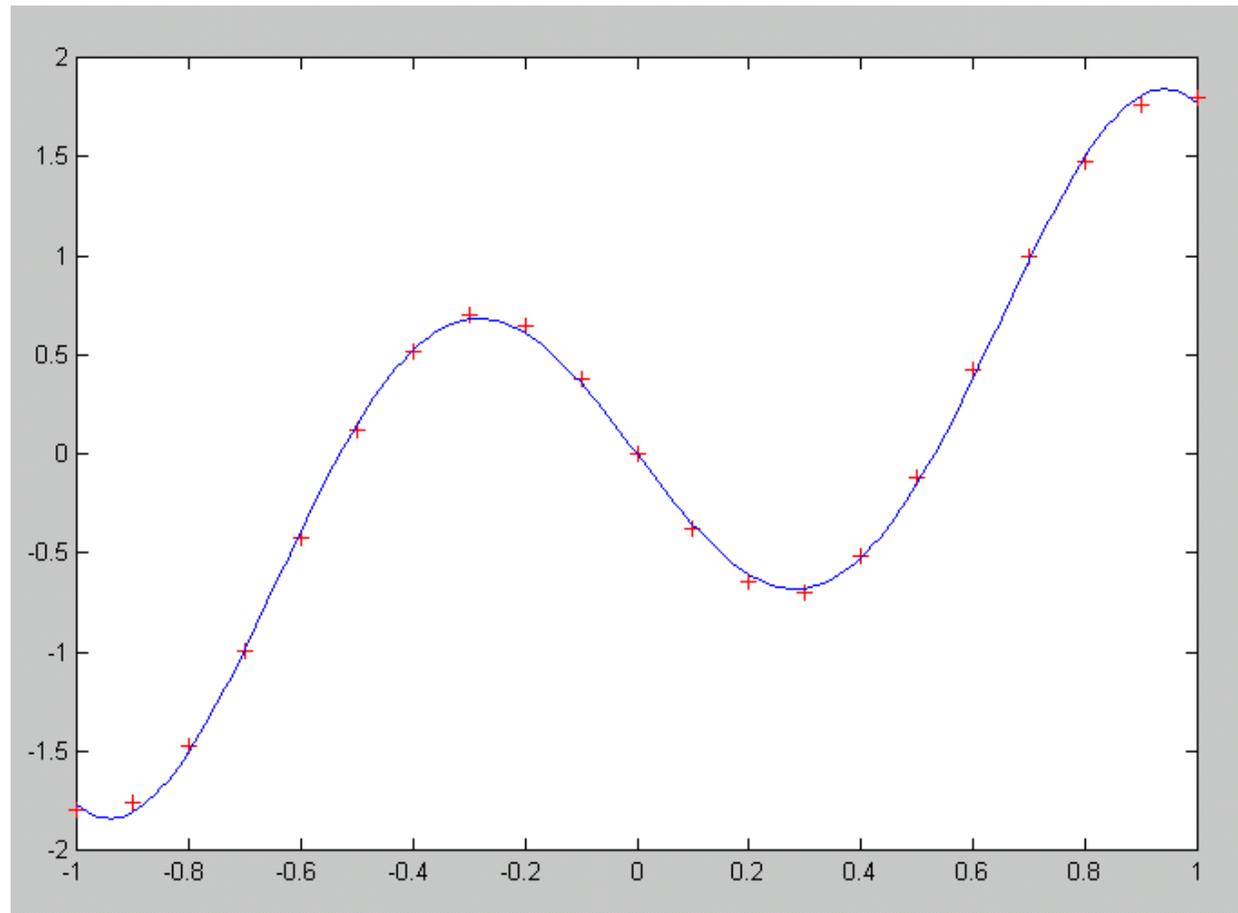
- Rede RBF para aproximar uma função e verificar papel dos parâmetros livres (os campos receptivos) afetam seu desempenho.
- A função: $y = f(x) = \text{sen}(x) - \text{sen}(5x)$ (Figura 1). Apresentam-se à rede 21 pares X-Y (círculos) onde X é a entrada e Y é sua saída desejada.



Exemplo de Emprego de Redes RBF

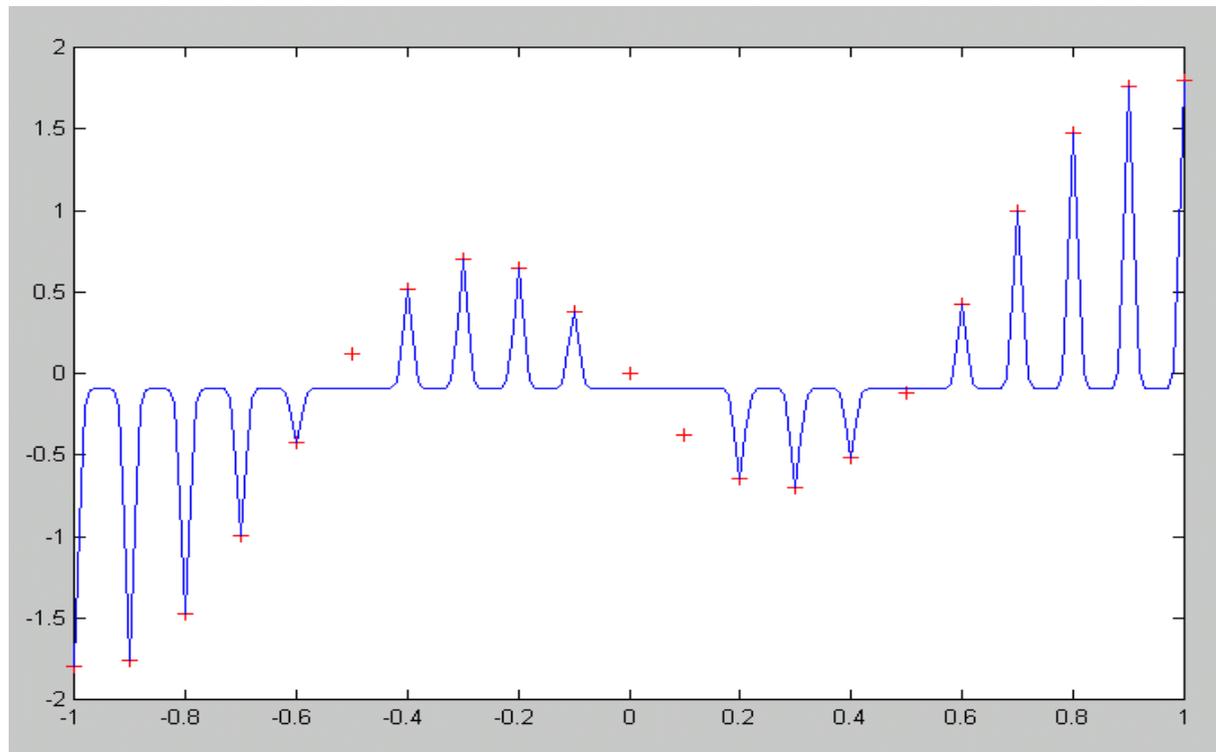
- A rede utilizada é composta por uma unidade de entrada, que recebe o valor de X , 21 unidades escondidas fixas (com centros localizados sobre os valores de entrada, que estão igualmente espaçados, ou seja, os centros estão fixados em uma “grade” regular) e uma unidade linear de saída, que produz o valor de Y . Os pesos da primeira camada são unitários e fixos e os pesos da segunda camada serão ajustados no treinamento. O erro será calculado pela função SSE.
- Depois de treinada, a rede aproxima a função dada com boa precisão e poderia ser utilizada para interpolação. O resultado está no gráfico da Figura 2, onde a linha cheia representa a saída da rede para 201 valores de entrada e as cruces representam os 21 pontos apresentados à rede durante o treinamento.

Exemplo de Emprego de Redes RBF



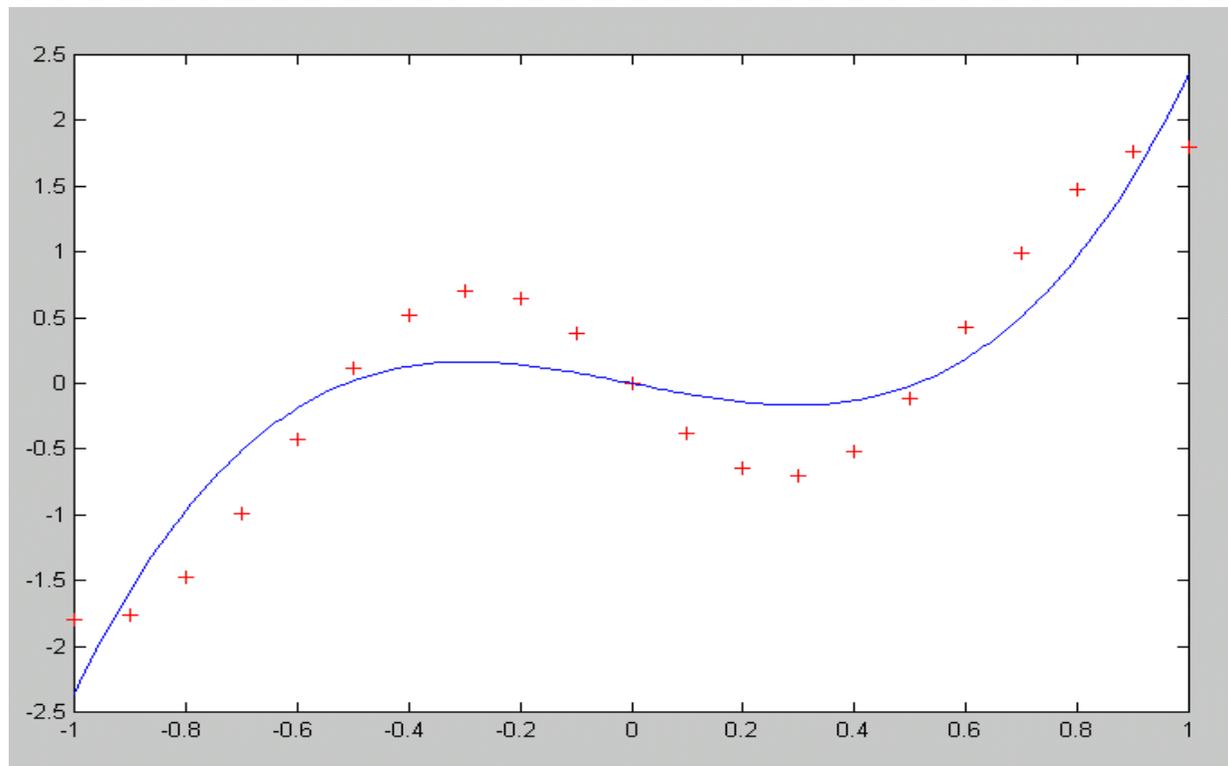
Exemplo de Emprego de Redes RBF

- Considere mesma rede anterior mas os campos receptivos são muito menores. Na Figura abaixo a linha cheia representa a saída da rede para 201 pontos de entrada e as cruces representam os 21 pontos utilizados durante o treinamento da rede.



Exemplo de Emprego de Redes RBF

- Considere a mesma rede com campos receptivos muito grandes. O resultado visto na Figura abaixo onde linha cheia representa a saída da rede para 201 pontos de entrada e as cruzes representam os 21 pontos utilizados de treinamento.

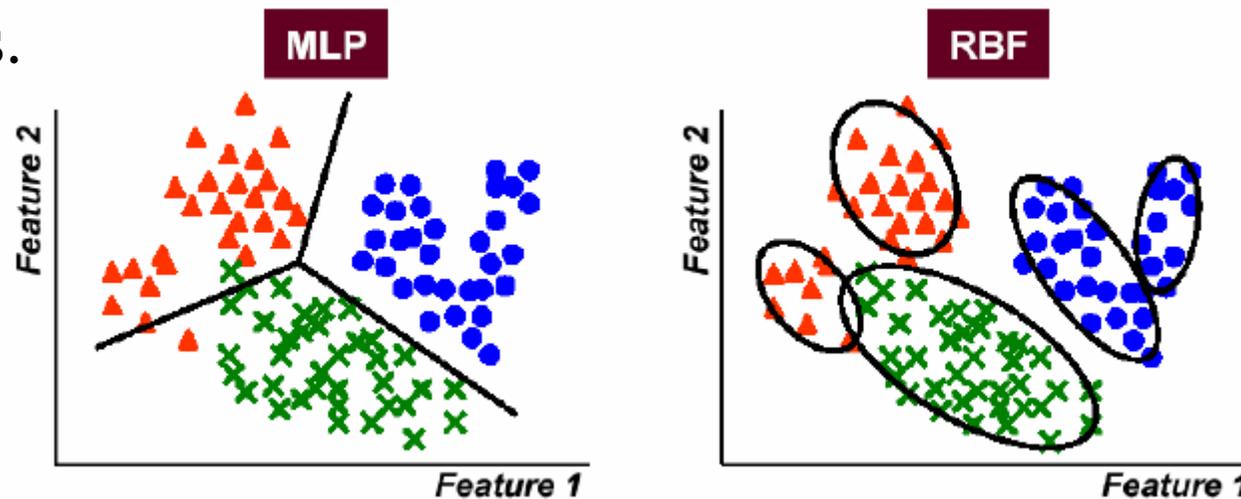


Comparando RBF e MLP-BP

- RBF só tem uma camada escondida e a MLP-BP pode ter mais.
- RBF usualmente tem mais unidades escondidas que a MLP-BP.
- RBF tem ativação na camada escondida e de saída com função de base e linear e a MLP-BP tem funções sigmoidais de ativação.
- RBF é usualmente menos sensível a desaprendizagem, devido a dados novos, que a MLP-BP.
- RBF pode necessitar de maior número de parâmetros ajustáveis que a MLP-BP para realizar mapeamento entrada-saída não-linear.
- RBF pode ser afetada pela maldição da dimensionalidade.

Comparando MLP-BP e RBF

- RBF separa classes por hiperelipsoides e a MLP-BP por hiperplanos.



- RBF e MLP-BP são aproximadores universais local e global.
- MLP-BP gera regiões globais de decisão, logo ela tende a extrapolar melhor que a RBF. Cabe salientar que extrapolação pode significar lidar com *outliers*.

Referências

- Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. IEEE Press.
- Haykin, S. (2008). *Neural Networks and Learning Machines*. Third Edition. Pearson.