SCE 5809
REDES NEURAIS

Profa. Roseli Ap. Francelin Romero

Redes RBF

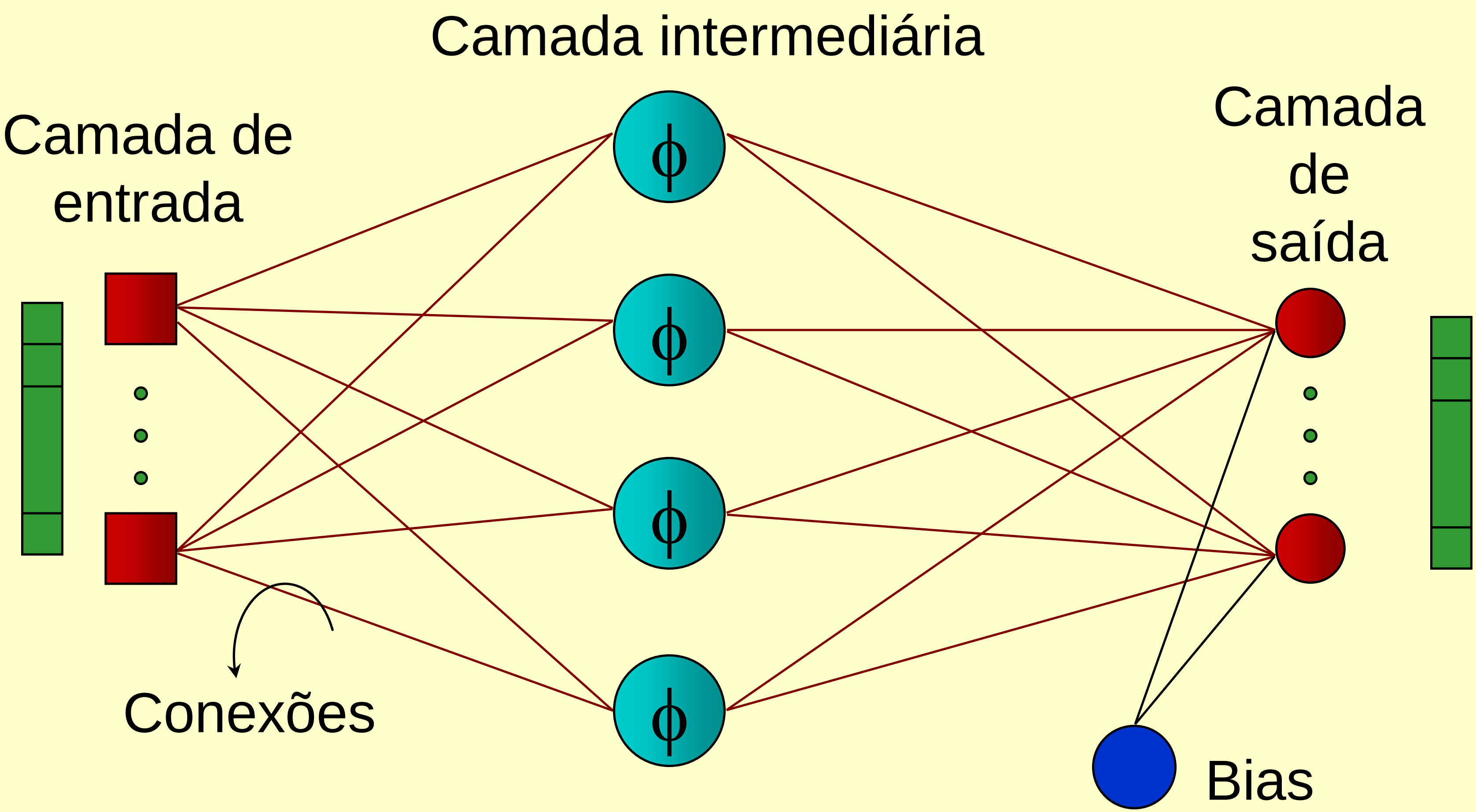
REDE RBF – Radial Basis Function foi proposta por Broomhead e Lowe (1988)

- Regressão
- Classificação
- Previsão de séries temporais

- Característica: Função de Ativação

  utiliza como argumento a função distância entre as entradas e um vetor protótipo.

Rede RBF



Camada intermediária
Camada de entrada
Camada de saída
ϕ
ϕ
ϕ
ϕ
Conexões
Bias

Teorema de Cover

As redes RBF formam inspiradas no teorema de Cover, que diz o seguinte:

"Um problema de classificação de padrões complexo, adaptado eu um espaço de dimensão mais alta, os padrões não linearmente separáveis são mais prováveis de serem separados do que em espaços de dimensão mais baixa."

Teorema de Cover

O teorema diz que se temos um vetor **x** com dimensionalidade alta e aplicamos um conjunto de funções não lineares φ_i **,** $i=1..m_0$, neste padrão gerando um vetor de dimensão mais alta, temos que a probabilidade deste vetor ser linearmente separável é maior.

Seja: $\mathbf{x} = (x_1, x_2, ..., x_N)$

Separar linearmente **x** em duas dicotomias, significa encontrar **w** tal que:

$$\mathbf{w}^T\mathbf{x} > 0, \mathbf{x} \in H_1$$
$$\mathbf{w}^T\mathbf{x} < 0, \mathbf{x} \in H_2$$

Teorema de Cover

Se aplicarmos φ em **x**, geramos um novo vetor φ(**x**) com m_1 componentes:

$$\varphi(\mathbf{x}) = [\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \ldots, \varphi_{m_1}(\mathbf{x})]^T$$

Separar linearmente este vetor, significa, da mesma forma, encontrar um **w**, tal que:

$$\mathbf{w}^T \varphi(\mathbf{x}) > 0, \mathbf{x} \in H_1$$

$$\mathbf{w}^T \varphi(\mathbf{x}) < 0, \mathbf{x} \in H_2$$

O teorema de Cover diz (em outras palavras) que a probabilidade de separação linear aumenta, a medida que m_1 aumenta.

Exemplo: Separar XOR

$$\mathbf{x} = (x_1, x_2)$$

$$\varphi : \Re^2 \rightarrow \Re^2$$

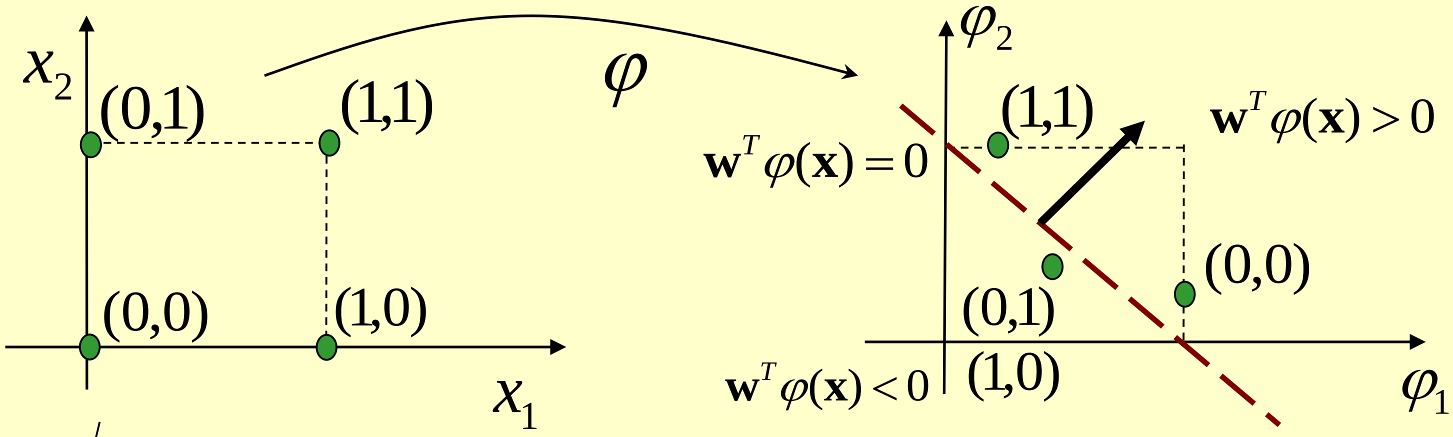
$$\varphi(\mathbf{x}) = \left(\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x})\right) = \left(\varphi_1(x_1, x_2), \varphi_2(x_1, x_2)\right)$$

Escolhendo:

$$\varphi_1(\mathbf{x}) = e^{-\|\mathbf{x}-\mathbf{t_1}\|^2}, \mathbf{t_1} = (1{,}1)$$

$$\varphi_2(\mathbf{x}) = e^{-\|\mathbf{x}-\mathbf{t}_2\|^2}, \mathbf{t_2} = (0{,}0)$$

Exemplo: Separar XOR



$$\nexists \mid \mathbf{w}^T\mathbf{x} > 0, \mathbf{x} \in H_1$$
$$\mathbf{w}^T\mathbf{x} < 0, \mathbf{x} \in H_2$$

$$\mathbf{w}^T\varphi(\mathbf{x}) > 0, \mathbf{x} \in H_1$$
$$\mathbf{w}^T\varphi(\mathbf{x}) < 0, \mathbf{x} \in H_2$$

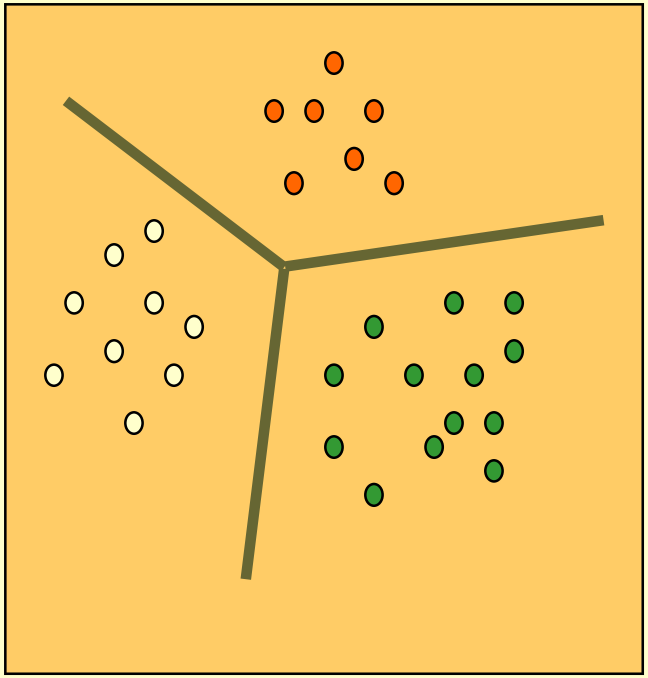
Redes RBF

- Evolução das redes MLP
- Redes de duas camadas
  - Primeira camada
    - Utiliza funções de ativação não lineares (funções de de base radial – funções de Green)
  - Segunda camada
    - Utiliza funções de ativação lineares
    - Semelhante à rede madaline

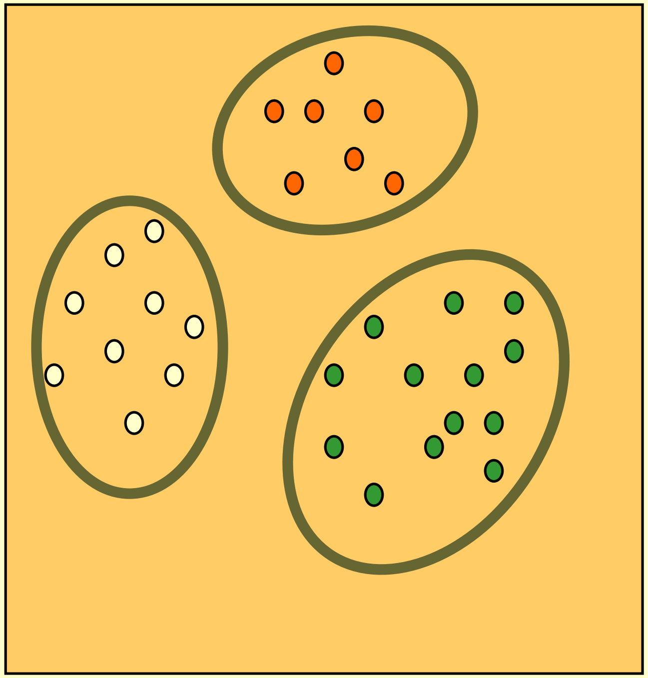
Redes RBF

- Redes RBF *X* MLP
  - MLP utiliza hiperplanos para particionar espaço de entradas (camada intermediária)
    - Definidos por funções da forma $f(\sum w_i x_i)$
  - RBF utiliza hiper-elipsóides para particionar o espaço de entradas (camada intermediária)
    - Definidos por funções da forma $\phi(\|x_i - \mu_i\|)$
    - Distância do vetor de entrada ao centro de um *cluster*

Fronteiras de decisão



MLP



RBF

Redes RBF

- É comum associar redes RBF com apenas 01 cam. intermed., mas existe redes RBF com mais de uma camada intermediária.
- Ela é usada para mapear um conj. x_i num conj. t_i Aprendizado Supervisionado

Redes RBF

- A primeira camada, cujos neurônios utilizam funções de base radiais, agrupa os dados de entrada em clusters. Esta camada transforma dados não linearmente separáveis em conjunto de saídas linearmente separáveis.

Redes RBF

- A segunda camada , camada de saída, procura classificar as entradas recebidas da camada anterior. Esta camada pode ser do tipo ADALINE ou Madaline.

Redes RBF

- Função de ativação das unidades escondidas
  - Não linear
  - Valor aumenta ou diminui com relação à distância a um ponto central
  - Funções típicas:
    - Função Gaussiana

$$\phi(v) = e^{\left(-\frac{v^2}{2\sigma^2}\right)}$$

$v = ||x - \mu||$
x: vetor de entrada
μ: centro da função radial
σ: largura da função radial

Redes RBF

- Função de ativação
  - Funções típicas
    - Função Multiquadrática

$$\phi(v) = \sqrt{(v^2 + \sigma^2)}$$

    - Função Thin-Plate-Spline

$$\phi(v) = v^2 \log(v)$$

Redes RBF

- Escolha da função depende de:
  – Nível de conhecimento sobre os dados
    - Funções devem cobrir pontos uniformemente distribuídos do espaço de entradas
  – Conhecimento da estrutura geral das entradas
    - Levar a estrutura em consideração na escolha das funções

Redes RBF

- Função de saída
  - Função de identidade $y_i = a_i$
- Treinamento
  - Geralmente dois estágios
    - Primeiro estágio: camada intermediária - não supervisionado
    - Segundo estágio: camada de saída - supervisionado

Primeiro estágio

- Determina parâmetros das funções de base radial
  - Número de bases
  - Centros das bases
  - Larguras das bases

Número de funções base

- Geralmente definido por tentativa e erro
- Sejam M o número de funções base e N o número de vetores de entrada:
  - M << N (M = N pode levar a overfitting)
  - Número de funções de base deve ser determinada pela complexidade dos dados
  - Número de funções base radial = número de clusters

Definição dos centros

- Existem várias abordagens
  - Seleção aleatória dos centros
  - Clustering
    - *K-means*
    - SOM
    - Algoritmos Genéticos

K-means

- O no. de clusters deve ser estipulado
- Supor que o espaço de padrões pode ser dividido em k regiões ou grupos. Esta partição é realizada do seguinte modo:
- k vetores de entrada são escolhidos aleatoriamente para serem os centros dos k grupos

K-means

- Os outros vetores de entrada são atribuídos ao grupo que tiver o centro mais próximo.
- Distância euclidiana
- Os centros são recalculados para que sejam o vetor médio do seu grupo.
- Novamente, os vetores de entrada são apresentados e associados ao centro + prox.
- Este processo é repetido até que os centros não precisem mais serem modificados.

Definição das larguras

- Heurísticas para definir larguras (raios) das funções radiais (σ)
  - Heurística 1: todas as larguras iguais à média sobre todas as distâncias Euclidianas entre o centro de cada unidade N_i e o centro da unidade N_j mais próxima

$$\sigma = \frac{1}{m}\sum_{i=1}^{m} \|\mu_i - \mu_j\|$$

μ_j é o centro mais próximo de μ_i

Definição das larguras

- Heurística 2: atribuir a cada unidade uma largura diferente baseada na distância do seu centro ao centro da unidade mais próxima

$$\sigma_i = \alpha \, || \mu_i - \mu_j ||$$

μ_j é o centro mais próximo de μ_j

$1.0 < \alpha < 1.5$

Definição das larguras

- Heurística 3: atribuir a cada σ_j a distância média de seu centro aos N vetores de entrada mais próximos
- Heurística 4: atribuir a σ_j um valor constante (geralmente 1)

Segundo estágio

- Determina pesos da camada de saída
  - Recebe vetores linearmente separáveis
  - Supervisionado
  - Classificação/regressão dos vetores de entrada
- Métodos para ajustar pesos
  - Regra delta - (M)adaline

Algoritmo Rede RBF

1. Inicialização do vetor peso

    Para j $\leftarrow$ 1 ate m faça $w_j \leftarrow 0$

2. Atualização dos pesos

    Para n $\leftarrow$ 1 até Maxciclos faça

        Para i $\leftarrow$ 1 até p faça

            $y_i^{(n)} \leftarrow \Sigma\, w_j \; \varphi_j(x_i)$

            Para j=1 até m faça

                $w_j \leftarrow w_j + \eta(t - y_i^{(n)})\, \varphi_j(x_i)$

            Fim_para

        Fim_para

    Fim_para

Este modelo e mais rapido que MLP

- Referencias:

- Broomhead, D.S., Lowe, D., " Multivariable functional interpolation and adaptive networks", Complex Systems, v.2, p. 321-355, 1988

- X. He and A. Lapedes, " Nonlinear modelling and prediction by sucessive approximations using RBF functions" , TR, Los Alamos Lab., 1991.

- Alguns slides foram cedidos pelo prof. Andre de Carvalho (andre@icmc.usp.br)