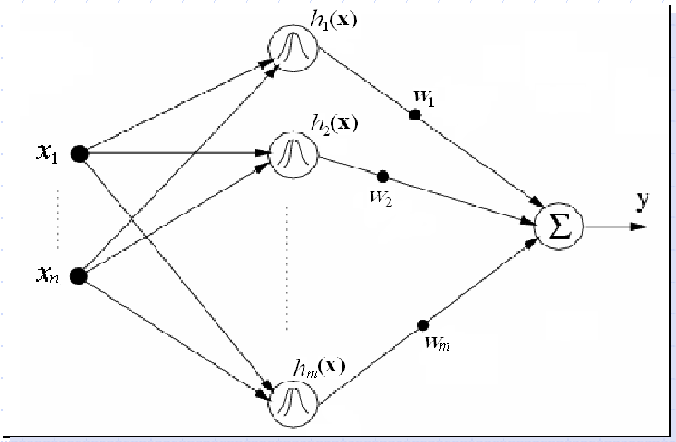
Introdução às Redes Neurais

Redes RBF:
Parte III

Prof. Ricardo J. G. B. Campello

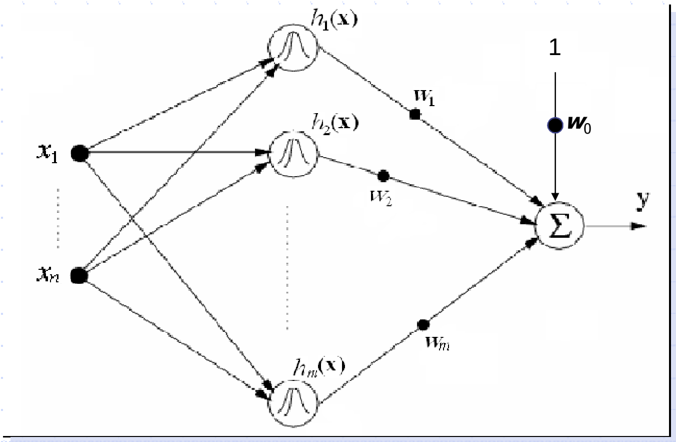
Modelo RBF (revisão)



$$y = \sum_{i=1}^{m} w_i \cdot h_i(x_1, \cdots, x_n) \quad = \quad \mathbf{h}^{\mathrm{T}} \cdot \mathbf{w}$$

2

Modelo RBF (revisão)



$$y = w_0 + \sum_{i=1}^{m} w_i \cdot h_i(x_1, \cdots, x_n) = \begin{bmatrix} 1 & \mathbf{h}^T \end{bmatrix} \cdot \begin{bmatrix} w_0 \\ \mathbf{w} \end{bmatrix}$$

3

Treinamento RBF (revisão)

- Problemas:
  - Dadas as funções de base radial, como determinar o melhor vetor de pesos **w** ?
  - Como determinar bons parâmetros para as funções de base radial (centros e aberturas) ?
    - reduzir no. de funções para dada precisão, ou
    - aumentar precisão para dado no. de funções

4

Treinamento RBF (revisão)

- Problemas:
  - **Dadas as funções de base radial, como determinar o melhor vetor de pesos w ?**
  - Como determinar bons parâmetros para as funções de base radial (centros e aberturas) ?
    - reduzir no. de funções para dada precisão, ou
    - aumentar precisão para dado no. de funções

5

Obtenção dos Pesos (revisão)

- Para k = 1, ..., N padrões de entrada $\mathbf{x}_k$ e saída $y(\mathbf{x}_k)$, deseja-se que a rede RBF possua pesos **w** tais que:

$$\underbrace{\begin{bmatrix} 1 & h_1(\mathbf{x}_1) & \cdots & h_m(\mathbf{x}_1) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & h_1(\mathbf{x}_N) & \cdots & h_m(\mathbf{x}_N) \end{bmatrix}}_{\mathbf{H}} \cdot \underbrace{\begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_m \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} y(\mathbf{x}_1) \\ \vdots \\ y(\mathbf{x}_N) \end{bmatrix}}_{\mathbf{y}}$$

o que implica:

$$\boxed{\mathbf{w} = \left(\mathbf{H}^{\mathrm{T}} \cdot \mathbf{H}\right)^{-1} \cdot \mathbf{H}^{\mathrm{T}} \cdot \mathbf{y}}$$

6

Treinamento RBF (revisão)

- Problemas:
  - Dadas as funções de base radial, como determinar o melhor vetor de pesos **w** ?
  - **Como determinar bons parâmetros para as funções de base radial (centros e aberturas)?**
    - **reduzir no. de funções para dada precisão, ou**
    - **aumentar precisão para dado no. de funções**

7

Determinação das Funções Radiais

- **Método 1** (Heurística Simples):
  - Distribuir as funções de maneira homogênea sobre o domínio das variáveis de entrada
  - Por exemplo, se a rede possui duas entradas, x_1 e x_2, que assumem valores no intervalo [-10,+10], então o domínio das funções são todos os pontos (x_1,x_2) tais que $x_1 \in [-10,+10]$ e $x_2 \in [-10,+10]$

8

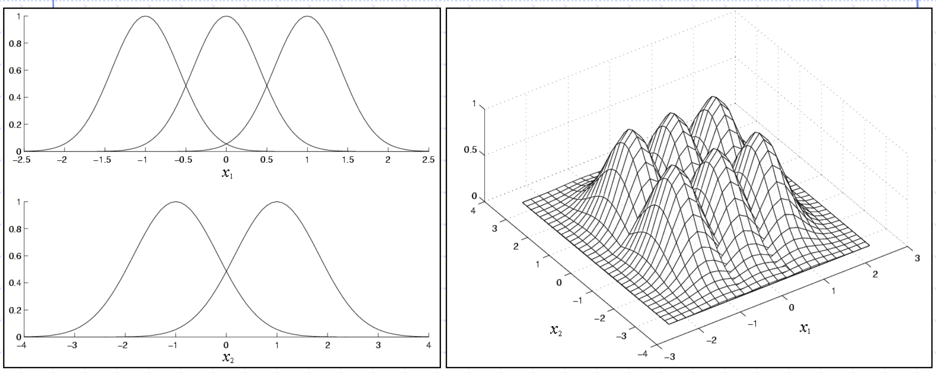
Determinação das Funções Radiais

- **Método 1** (cont.):
  - Os centros das funções com relação a cada variável (dimensão) podem ser distribuídos uniformemente (eqüidistantes) ao longo do domínio daquela var.
  - As aberturas das funções (desvios padrão) com relação a cada variável podem ser feitas iguais à distância entre dois centros consecutivos.

9

Determinação das Funções Radiais

- **Método 1** (exemplo):



10

Determinação das Funções Radiais

- **Método 2** (back-propagation):
  - Análogo ao treinamento de MLPs
  - A cada iteração ajusta-se o conjunto de parâmetros (centros e aberturas das funções) no sentido de minimizar o erro entre a saída da rede e a saída desejada para um conjunto de padrões
  - Usualmente tenta-se minimizar: $J = \frac{1}{N}\sum_{k=1}^{N}\left(y(\mathbf{x}_k) - y_{\mathrm{RBF}}(\mathbf{x}_k)\right)^2$
  - Demanda o cálculo do gradiente (derivadas) de J com relação aos parâmetros...
  - **Usualmente aplica-se para refinar o resultado inicial obtido com o método 1 ou método 3 (visto depois)**

11

Exemplo

- **Problema**:
  - Aproximar a função f(x) = sen(2x) / exp(x/5)
- N = 21 padrões de entrada e saída (k = 1, ..., 21):

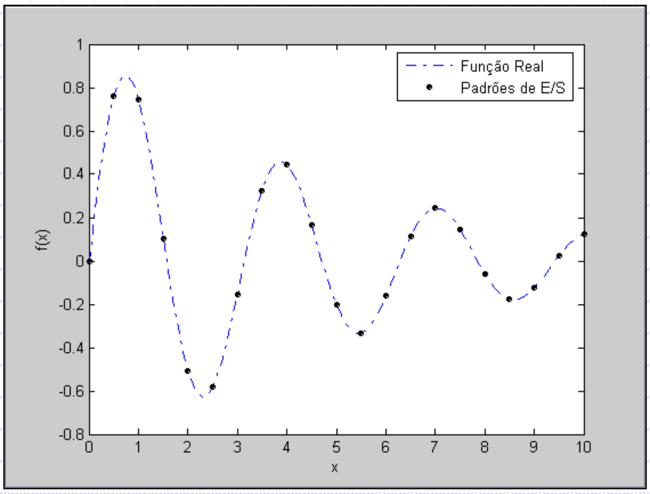
Padrão E/S	x	f(x)
1	0	0
2	0.5	0.7614
3	1.0	0.7445
4	1.5	0.1045
5	2.0	-0.5073
6	2.5	-0.5816
7	3.0	-0.1533
8	3.5	0.3262
9	4.0	0.4445
10	4.5	0.1676

11	5.0	-0.2001
12	5.5	-0.3329
13	6.0	-0.1616
14	6.5	0.1145
15	7.0	0.2443
16	7.5	0.1451
17	8.0	-0.0581
18	8.5	-0.1756
19	9.0	-0.1241
20	9.5	0.0224
21	10.0	0.1236

12

Exemplo

- **Problema**: Aproximar a função f(x) = sen(2x) / exp(x/5)



13

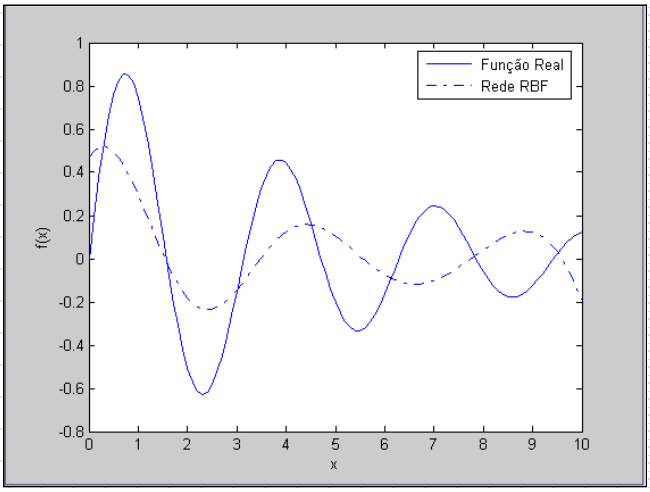
Exemplo

- **Solução 1 (Método 1)**:
  - Rede RBF com 6 neurônios (funções Gaussianas)
  - Centros distribuídos de maneira uniforme
    - $c_1 = 0, c_2 = 2, c_3 = 4, c_4 = 6, c_5 = 8, c_6 = 10$
  - Desvios padrão iguais à distância entre dois centros consecutivos
    - $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = \sigma_6 = 2$
  - Apenas os pesos otimizados via Mínimos Quadrados (MQ):
    - 7 pesos (w_0 ... w_6) obtidos via MQ: $\mathbf{w} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}$

14

Exemplo

- **Resultado Solução 1**:



15

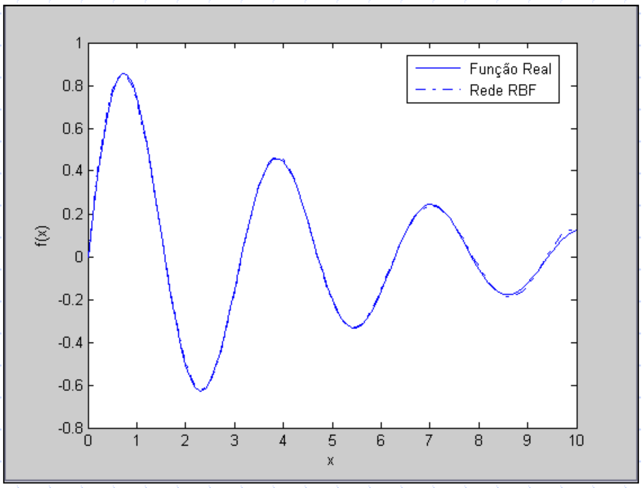
Exemplo

- **Solução 2 (Método 1)**:
  - Rede RBF com 11 neurônios
  - Centros distribuídos de maneira uniforme
    - $c_1 = 0$, $c_2 = 1$, $c_3 = 2$, ..., $c_{11} = 10$
  - Desvios padrão iguais à distância entre dois centros consecutivos
    - $\sigma_1 = \sigma_2 = \ldots = \sigma_{11} = 1$
  - Apenas os pesos otimizados via Mínimos Quadrados (MQ):
    - 12 pesos (w_0 ... w_{11}) obtidos via MQ: $\mathbf{w} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}$

16

Exemplo

◆ **Resultado Solução 2:**



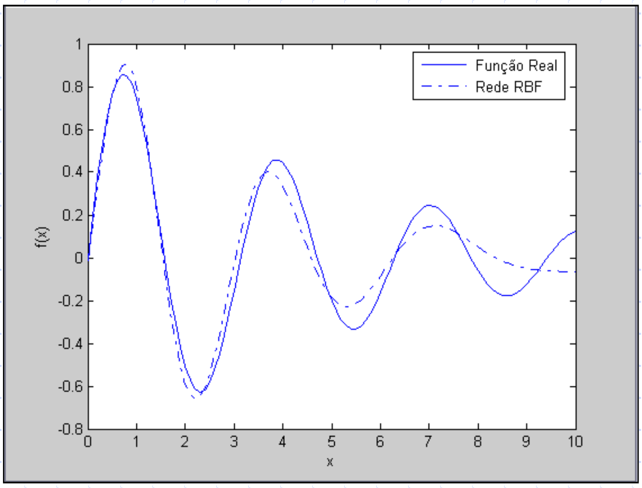
Exemplo

◆ **Solução 3 (Método 1 + 2):**

- Rede RBF com 6 neurônios
- Rede inicialmente configurada conforme Solução 1
- Sintonia fina realizada com **back-propagation**

Exemplo

- **Resultado Solução 3:**

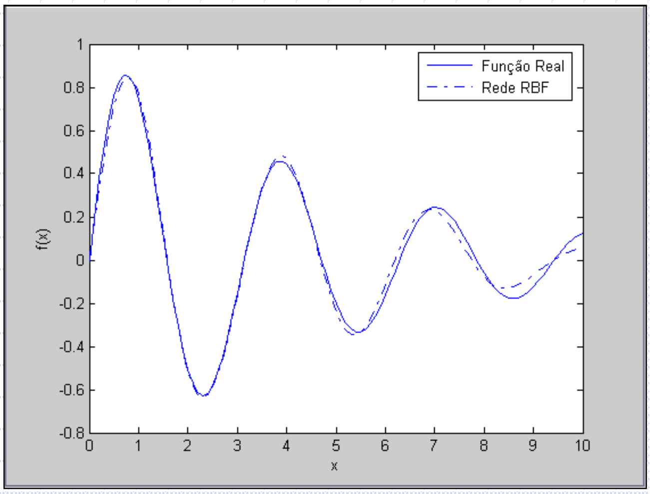


Exemplo

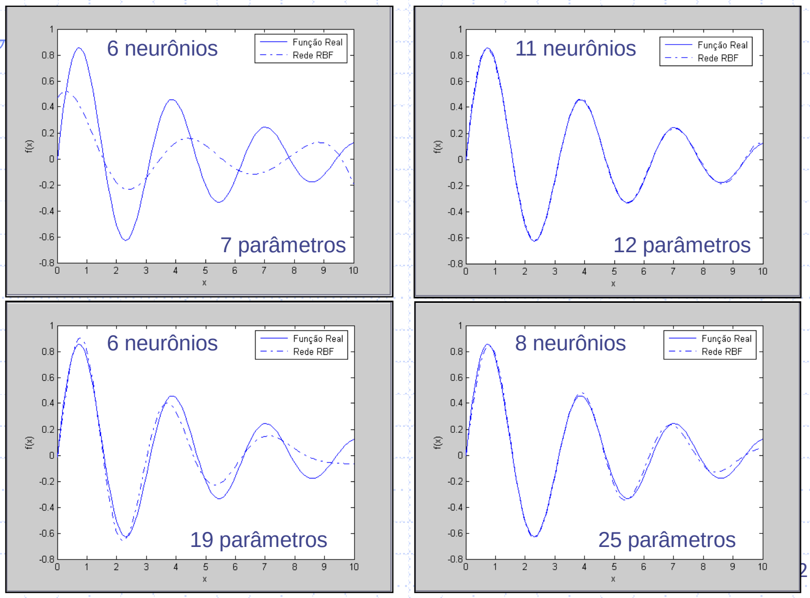
- **Solução 4 (Método 1 + 2):**
  - Rede RBF com 8 neurônios
  - Rede inicialmente configurada de forma análoga à Solução 1
    - Porém com 8 neurônios (Gaussianas)
  - Sintonia fina realizada com **back-propagation**

Exemplo

- **Resultado Solução 4:**



Exemplo (comparação)



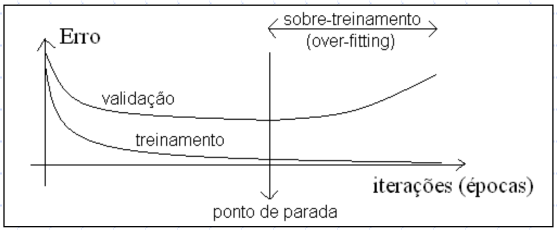
Exemplo (comparação)

- **Notas**:
  - A heurística de distribuição homogênea das funções radiais (método 1) foi muito eficaz nesse exemplo agindo sozinha, mas não necessariamente é sempre assim
  - Embora a rede com 8 neurônios refinados com back-propagation tenha uma quantidade de parâmetros maior durante a fase de **treinamento**, é um modelo mais simples e compacto do que aquele com 11 neurônios (para **utilização**).
  - A rede com 11 neurônios poderia ficar ainda mais precisa se seus parâmetros também fossem refinados com back-propagation

23

Dicas Básicas

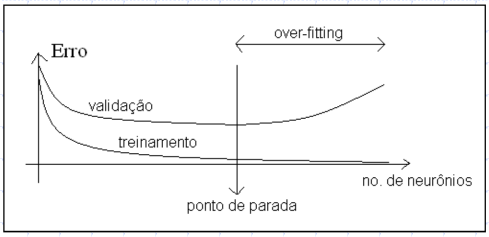
- Usualmente utiliza-se apenas uma parcela dos padrões disponíveis para treinar a rede e reserva-se uma outra parcela para teste ou **validação**
- Uma boa heurística para saber quando interromper o treinamento (sintonia fina) dos parâmetros via back-propagation é observar os erros de ambas as parcelas:



24

Dicas Básicas

- A mesma idéia pode ser usada para determinar quando parar de acrescentar neurônios à rede...



- Uma outra boa dica é normalizar os padrões de E/S de forma que cada variável tenha valores entre –1 e +1
  - Isso minimiza problemas numéricos durante o treinamento

25

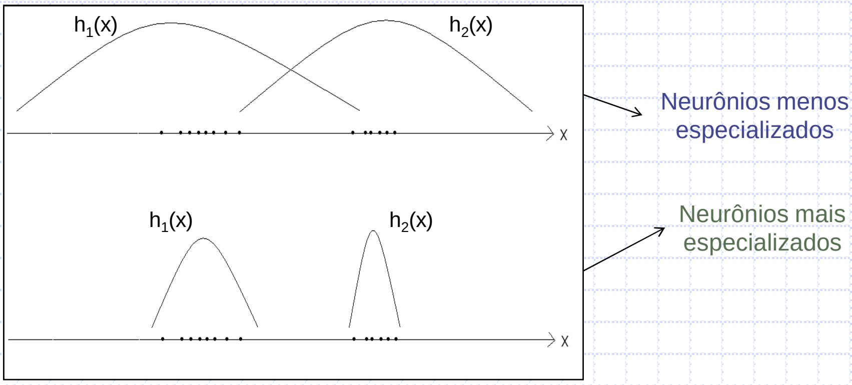
Determinação das Funções Radiais

- **Método 3** (Clustering):
  - Agrupar os padrões em grupos (clusters) de padrões mais similares entre si do que aos demais padrões.
  - Associar um neurônio (função radial) para cada grupo de padrões, otimizando a representatividade de cada neurônio / função.
  - Idéia é que cada neurônio responda de forma apropriada e similar a um determinado conjunto de padrões similares
    - Tipicamente (sub)classes em **problemas de classificação**

26

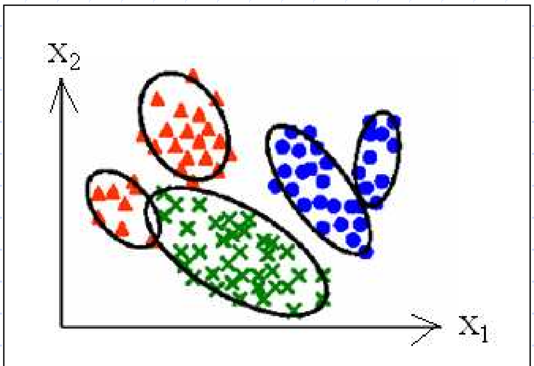
Determinação das Funções Radiais

- **Método 3** (cont.):
  - Idéia intuitiva (1 var. de entrada):



Determinação das Funções Radiais

- **Método 3** (cont.):
  - Idéia intuitiva (2 vars. de entrada):



http://people.sabanciuniv.edu/berrin/cs512/lectures/8-nn6-rbf.ppt (09/2008)

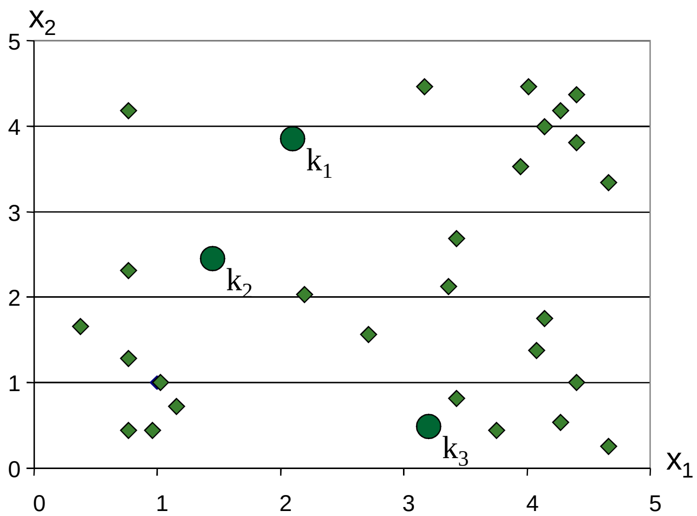
Determinação das Funções Radiais

- **Método 3** (cont.):
  - Um dos algoritmos de agrupamento mais populares é o **K-means**, que agrupa N padrões em K grupos:

1. Escolher K protótipos iniciais dos grupos
   - padrões ou pontos quaisquer do domínio de entrada
2. Atribuir cada um dos N padrões a um dos K grupos, de acordo com a maior proximidade aos protótipos dos grupos
3. Recalcular os protótipos como **centróides** (ponto médio) dos padrões pertencentes àquele grupo
4. Parar se não houver mais mudanças, ou retornar ao passo 2

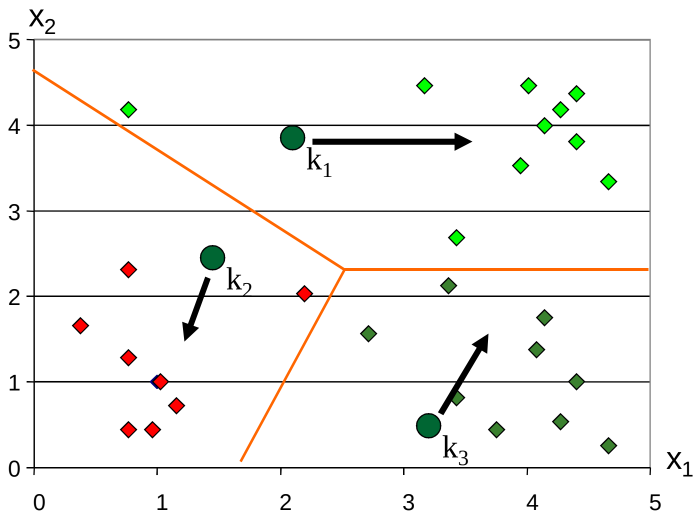
29

Exemplo de K-means: Iteração 1 / Passo 1



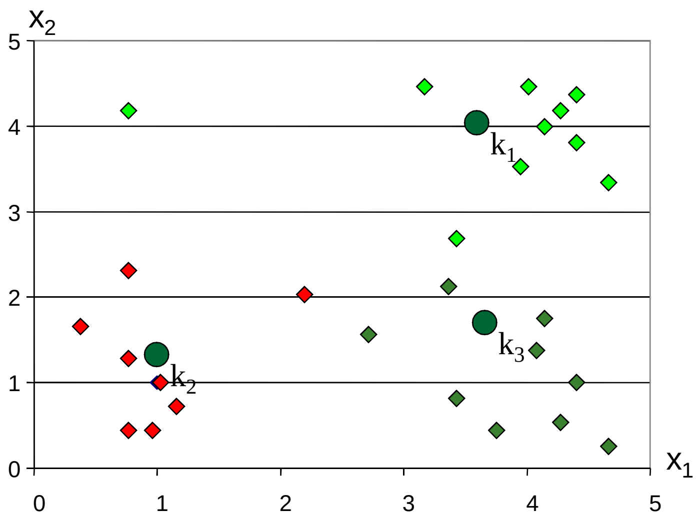
Fonte: Keogh, E. A Gentle Introduction to Machine Learning and Data Mining for the Database Community, SBBD 2003, Manaus.

K-means: Iteração 1 / Passo 2



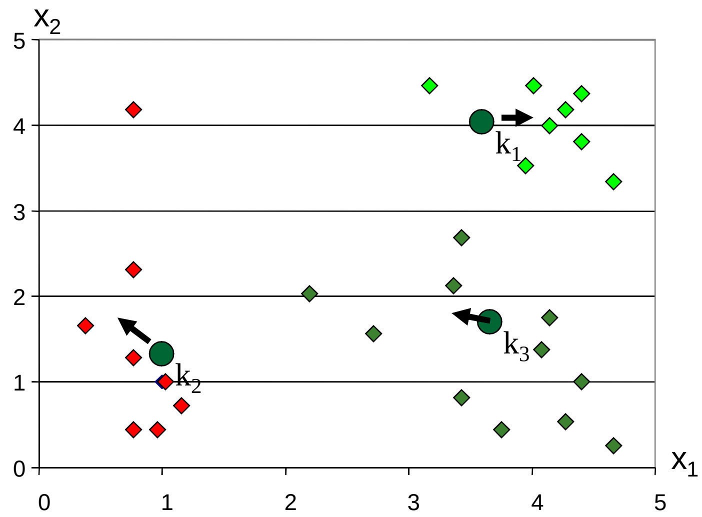
Fonte: Keogh, E. A Gentle Introduction to Machine Learning and Data Mining for the Database Community, SBBD 2003, Manaus.

K-means: Iteração 1 / Passo 3



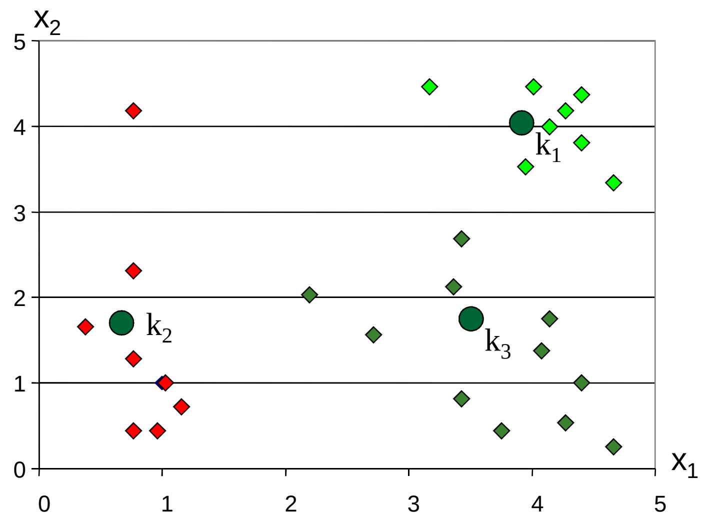
Fonte: Keogh, E. A Gentle Introduction to Machine Learning and Data Mining for the Database Community, SBBD 2003, Manaus.

K-means: Iteração 2 / Passo 2



Fonte: Keogh, E. A Gentle Introduction to Machine Learning and Data Mining for the Database Community, SBBD 2003, Manaus.

K-means: Iteração 2 / Passo 3 e Final



Fonte: Keogh, E. A Gentle Introduction to Machine Learning and Data Mining for the Database Community, SBBD 2003, Manaus.

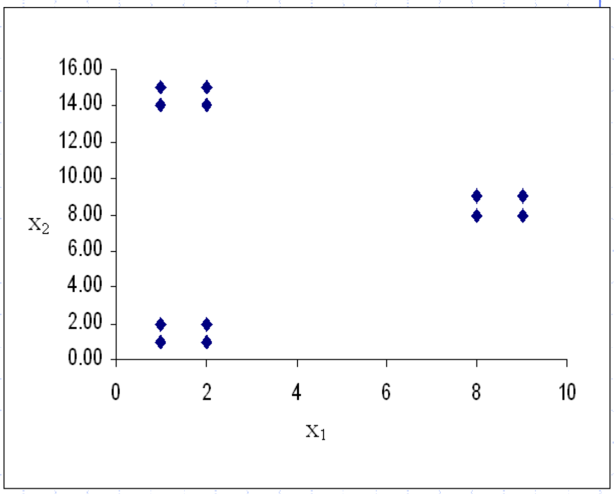
Determinação das Funções Radiais

- **Método 3** (cont.):
  - Executa-se K-means com **K** igual ao número **m** de neurônios (funções radiais) desejado na rede RBF
  - Toma-se o protótipo (centróide) de cada grupo resultante como centro de uma função radial
    - cada componente do centróide é o centro da função radial correspondente na respectiva variável
  - Toma-se os desvios padrão de cada grupo, em cada variável, como o desvio padrão da função radial correspondente na respectiva variável
    - forma simplificada...

35

Método 3 (Exemplo)

padrão (entrada)	x_1	x_2
1	1	2
2	2	1
3	1	1
4	2	2
5	**8**	**9**
6	**9**	**8**
7	**9**	**9**
8	**8**	**8**
9	1	15
10	2	15
11	1	14
12	2	14



Slide cedido pelo Prof. Eduardo Raul Hruschka

Método 3 (Exemplo)

- K-means com K = 3 produziria tipicamente os protótipos (1.5 , 1.5), (8.5 , 8.5), (1.5 , 14.5)
- O terceiro grupo, por exemplo, possui desvios padrão em x_1 e em x_2 respectivamente iguais a:

$$\sigma_1 = \sqrt{\frac{(1-1.5)^2 + (2-1.5)^2 + (1-1.5)^2 + (2-1.5)^2}{4}} = 0.5$$

$$\sigma_2 = \sqrt{\frac{(15-14.5)^2 + (15-1.5)^2 + (14-14.5)^2 + (14-14.5)^2}{4}} = 0.5$$

- Logo, a função radial (Gaussiana) correspondente pode ser definida como:

$$\boxed{h_3(x_1, x_2) = \exp\left(-\frac{(x_1 - 1.5)^2}{0.5^2}\right) \times \exp\left(-\frac{(x_2 - 14.5)^2}{0.5^2}\right)}$$

Bibliografia

- Braga, et al., "Redes Neurais Artificiais: Teoria e Aplicações", LTC, 2ª Edição, 2007
- Haykin, "Neural Networks", Prentice Hall, 2nd Edition, 1999
- Kovács, "Redes Neurais Artificiais: Fundamentos e Aplicações", Collegium Cognitio, 2ª Edição, 1996