RBF (RADIAL BASIS FUNCTION)

Prof. Valmir Macario Filho

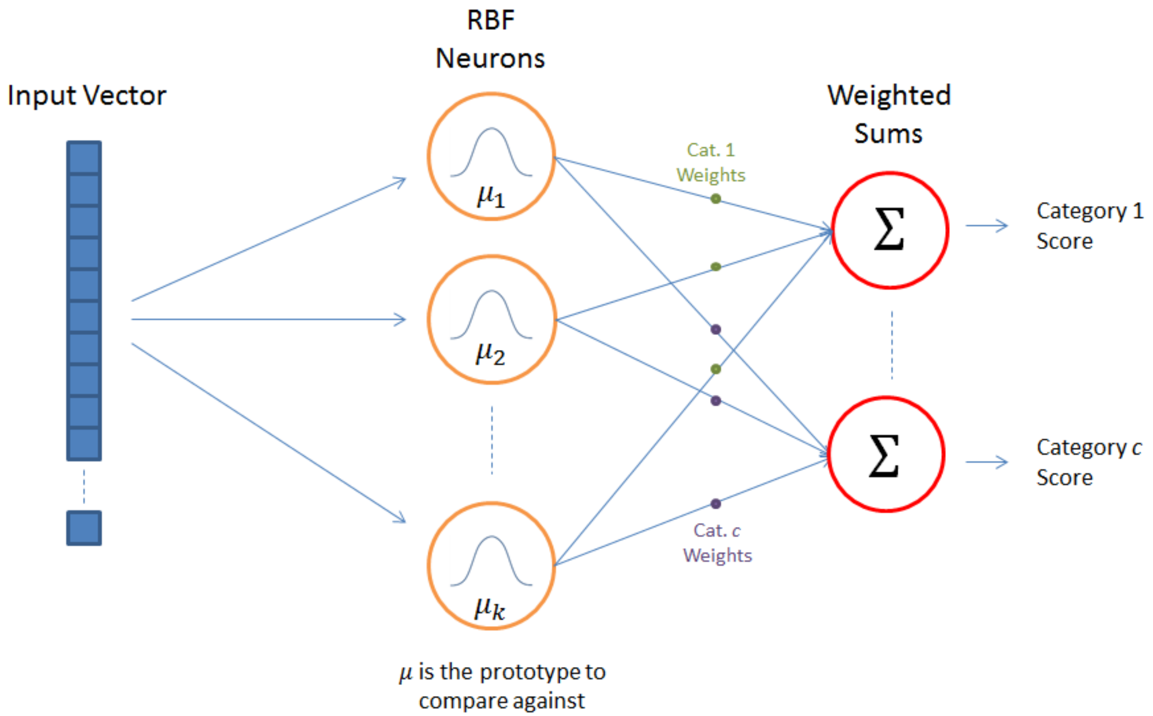
DC-UFRPE

FUNÇÃO DE BASE RADIAL (RBF)

- É uma abordagem inspirada na propriedade de alguns neurônios biológicos de sistemas nervosos, chamada de resposta localmente sintonizada (*locally tuned response*).
- Tais células nervosas respondem seletivamente a um intervalo finito do espaço de sinais de entrada.
- As células do sistema auditivo, por exemplo, têm uma resposta localmente ajustada à frequência, que é uma consequência das suas propriedades biofísicas.
- Redes RBF foram independentemente propostas por Broomhead e Lowe (1988), Lee e Kil (1988), Niranjan e Fallside (1988) e Moody e Darken (1989a, 1989b).

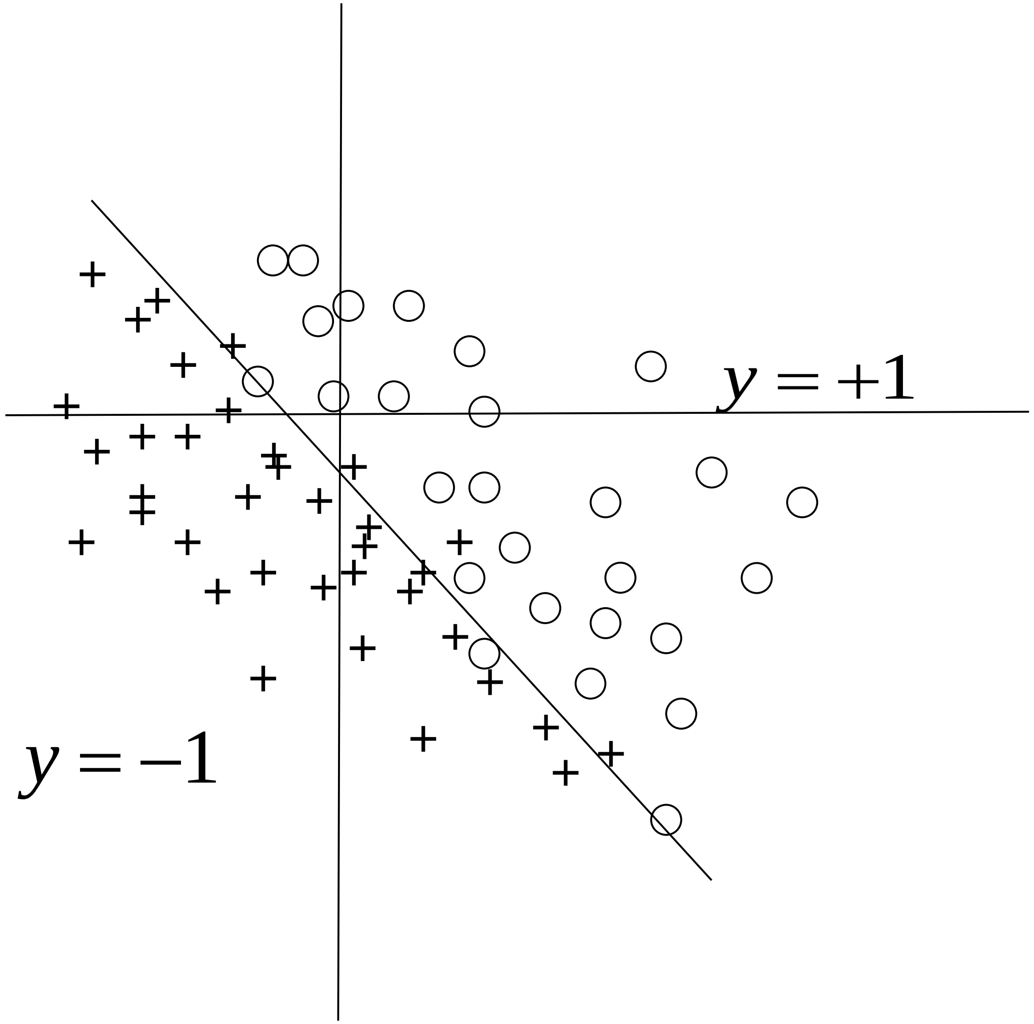
FUNÇÃO DE BASE RADIAL (RBF)

- 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.
  - Mapeamento não linear
- **Camada de saída:** unidades de processamento lineares.



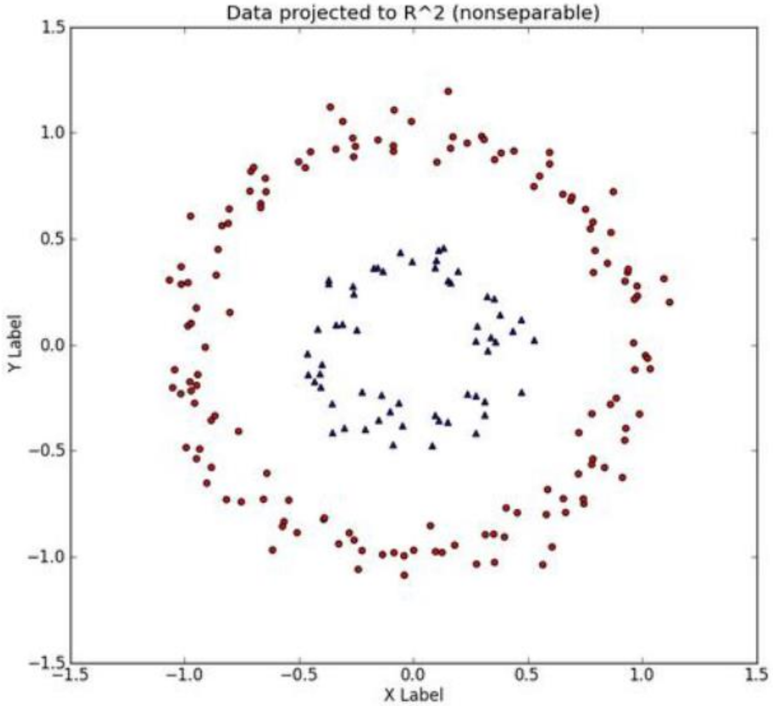
TEOREMA DE COVER

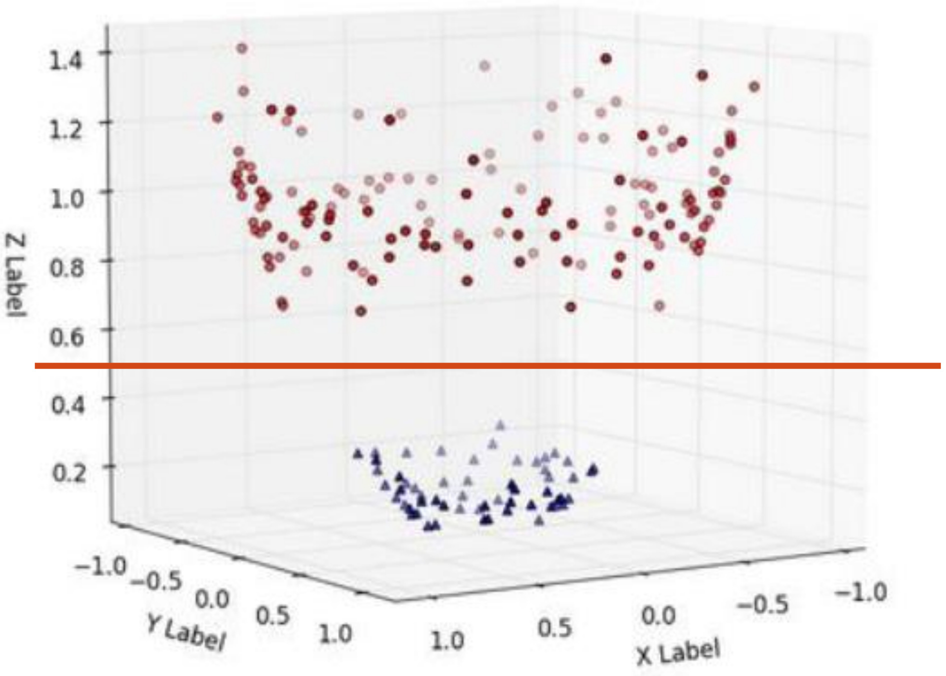
- "Um problema complexo de classificação de padrões disposto não linearmente em um espaço de alta dimensão tem maior probabilidade de ser linearmente separável do que em um espaço de baixa dimensionalidade."



TEOREMA DE COVER

- Um determinado problema Não Linearmente Separável pode, de forma probabilística, ser transformado em um problema Linearmente Separável através de uma transformação não linear que mapeia o espaço para outro espaço de ordem maior.

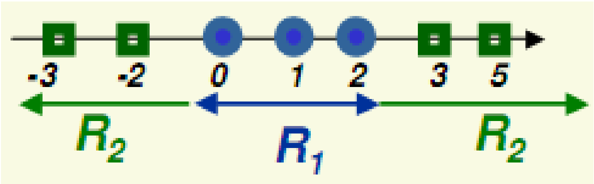




MAPEAMENTO NÃO LINEAR

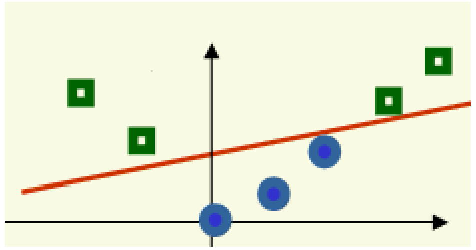
- Projetar os dados em outra dimensão usando uma função de kernel (kernel trick).
- Encontrar um hiperplano que separe os dados nesse espaço.

Em qual dimensão esses dados seriam linearmente separáveis?



1D

$$\varphi(x) = (x, x^2)$$



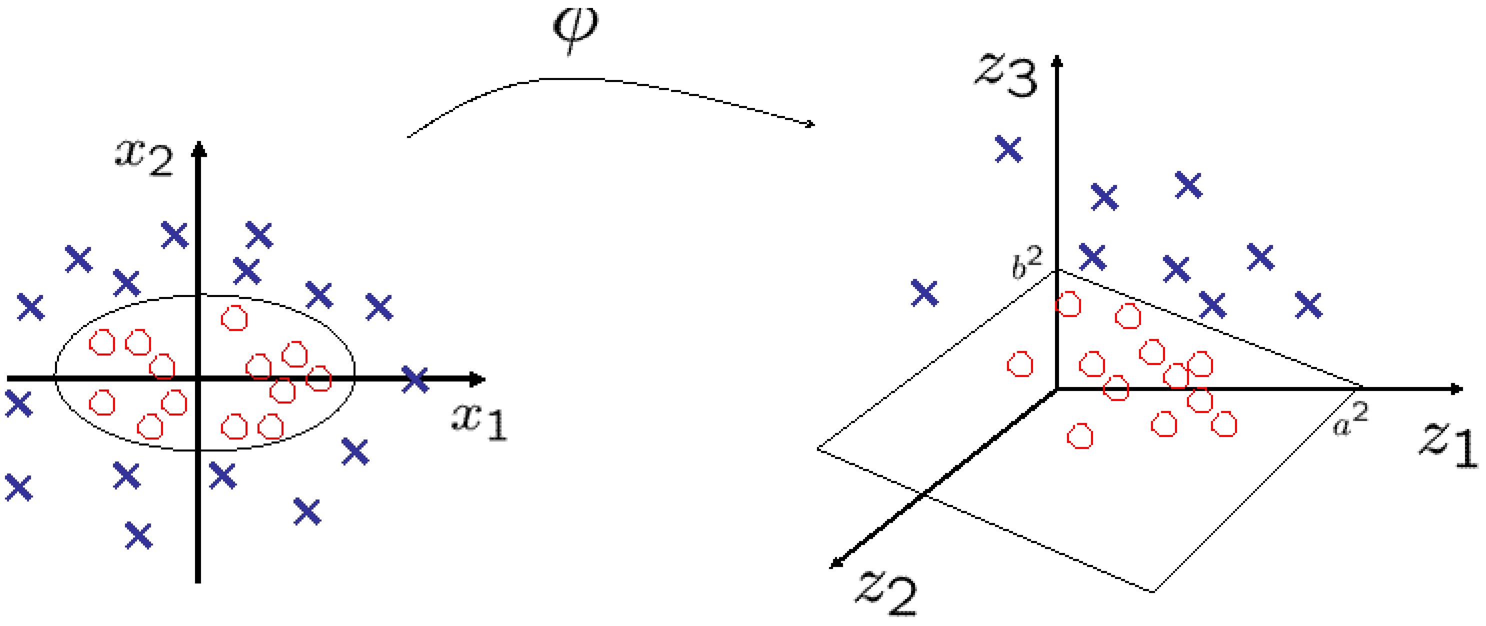
2D

KERNEL TRICK

- A função que projeta o espaço de entrada no espaço de características é conhecida como *Kernel*
- Baseado no teorema de Cover
  - Dados no espaço de entrada são transformados (transf. não linear) para o espaço de características, onde são linearmente separáveis.
- O vetor $\varphi(x_i)$ representa a "imagem" induzida no espaço de características pelo vetor de entrada

EXEMPLO DE KERNEL TRICK

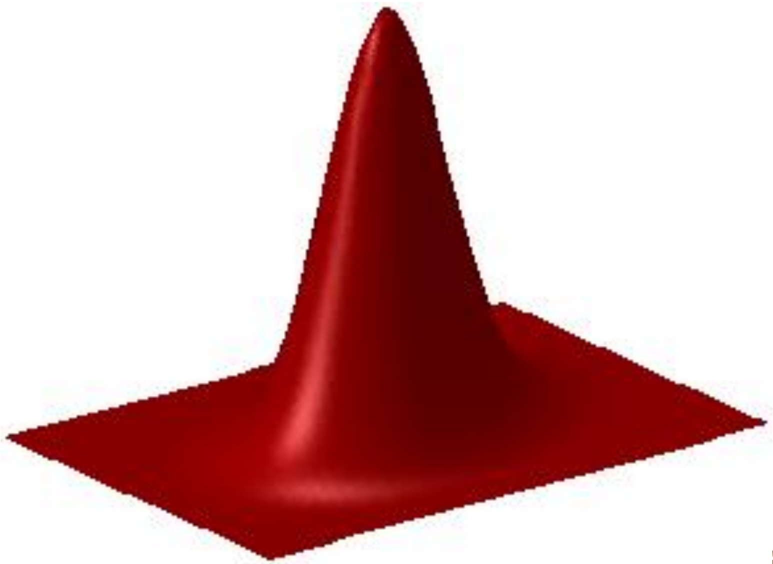
$$\begin{aligned} \phi : \quad \Re^2 &\longrightarrow \Re^3 \\ (x_1, x_2) &\longmapsto (z_1, z_2, z_3) = (x_1^2, \sqrt{2}x_1x_2, x_2^2) \end{aligned}$$



FUNÇÃO DE BASE RADIAL

$$f(\mathbf{x}) = \sum_{i=1}^{m} w_i \varphi_i(\mathbf{x})$$

- Funções radiais são uma classe especial de funções em que sua resposta decresce (ou cresce) monotonicamente com o distanciamento de um ponto central (média).
- Função radial: $$\phi_i(\mathbf{x}) = \phi\left(\|\mathbf{x} - \mathbf{x}_i\|\right)$$
- Centro: $\mathbf{x}_i$
- Distância: $u = \|\mathbf{x} - \mathbf{x}_i\|$
- Forma: ϕ



FUNÇÃO DE BASE RADIAL COMUNS

- Gaussiana

$$\varphi(u) = e^{-\frac{u^2}{2\sigma^2}} \quad \sigma > 0 \ \text{ e } \ u \in \Re$$

- Multiquadrática

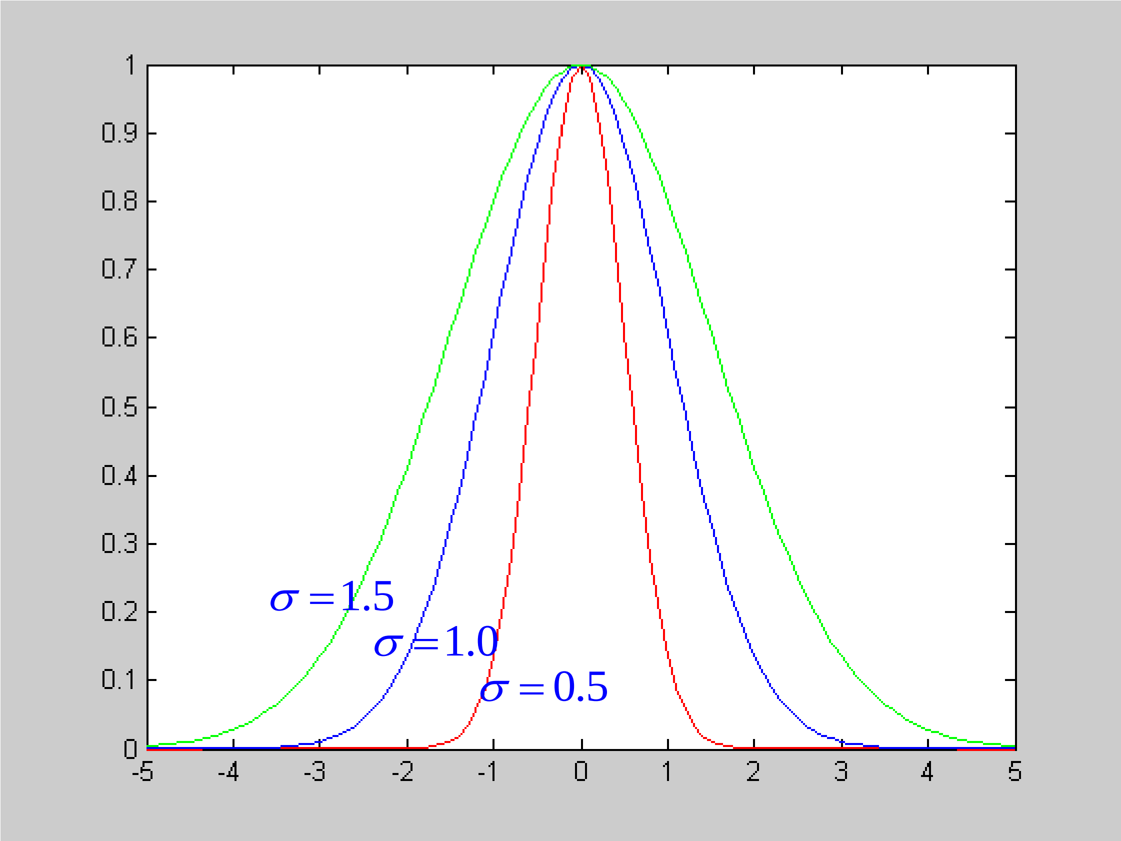
$$\varphi(u) = \frac{\sqrt{u^2 + c^2}}{c} \quad c > 0 \ \text{ e } \ u \in \Re$$

- Multiquadrática Inversa

$$\varphi(u) = \frac{c}{\sqrt{u^2 + c^2}} \quad c > 0 \ \text{ e } \ u \in \Re$$

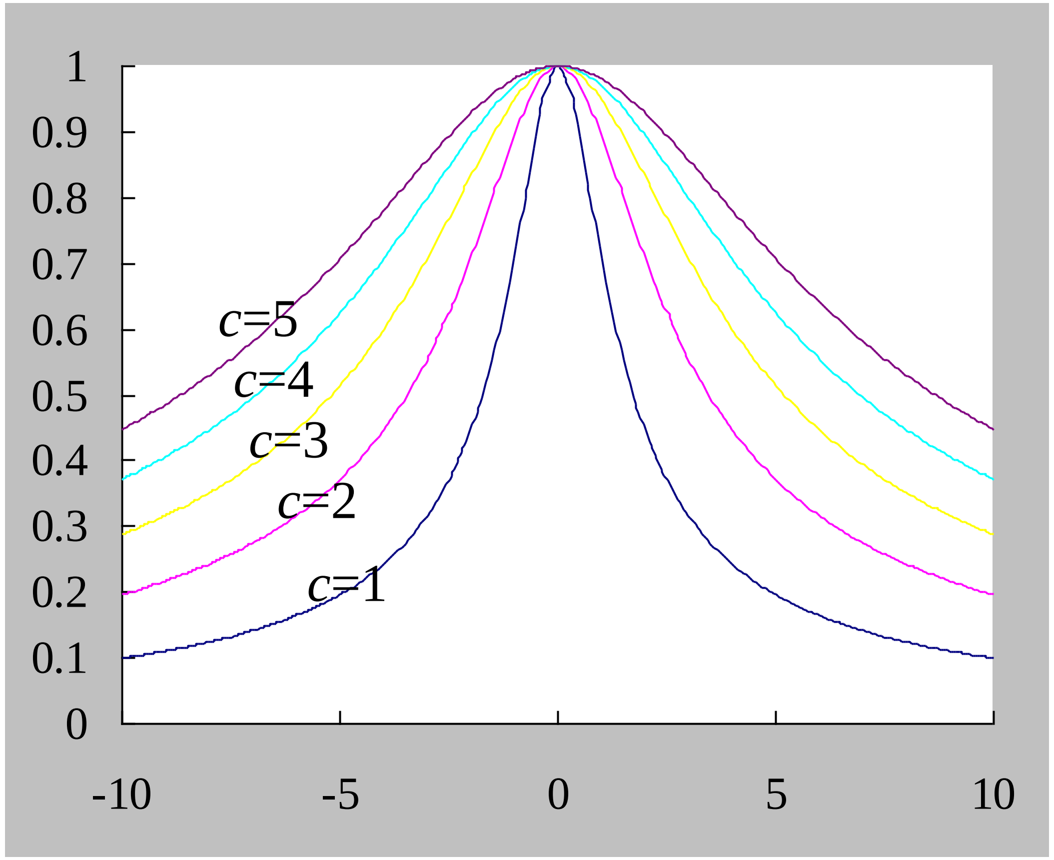
$$\varphi(u) = e^{-\frac{u^2}{2\sigma^2}} \quad \sigma > 0 \;\; \mathrm{e} \;\; u \in \Re$$

FUNÇÃO GAUSSIANA (σ=0.5,1.0,1.5)



$$\varphi(u) = \frac{c}{\sqrt{u^2 + c^2}} \quad c > 0 \ \text{ e } \ u \in \Re$$

MULTIQUADRÁTICA INVERSA

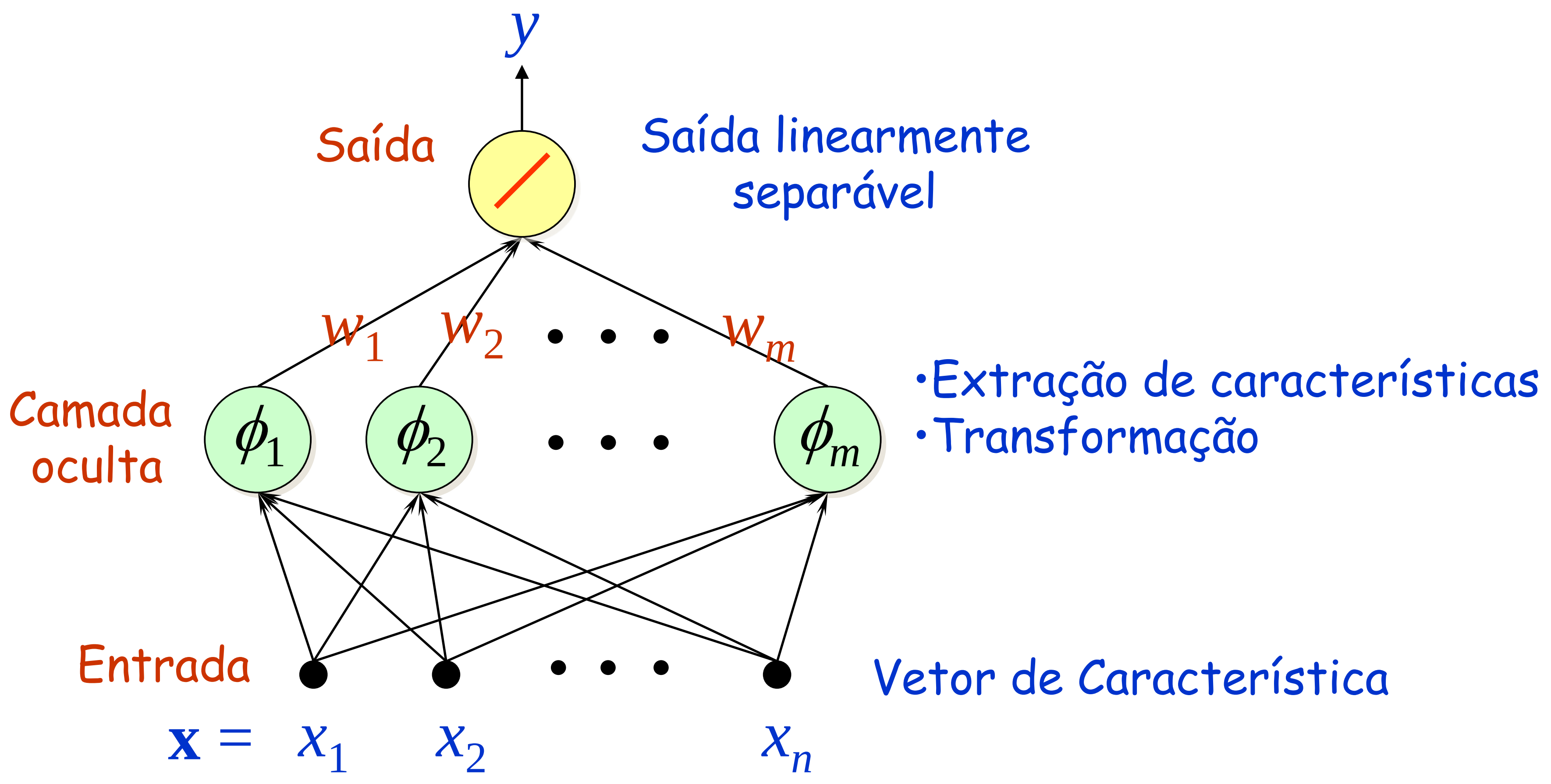


MODELO LINEAR

$$f(\mathbf{x}) = \sum_{i=1}^{m} \overbrace{w_i}^{\text{Pesos}} \underbrace{\phi_i(\mathbf{x})}_{\substack{\text{Funções} \\ \text{de Base Radial}}}$$

$$f(\mathbf{x}) = \sum_{i=1}^{m} w_i \phi_i(\mathbf{x})$$

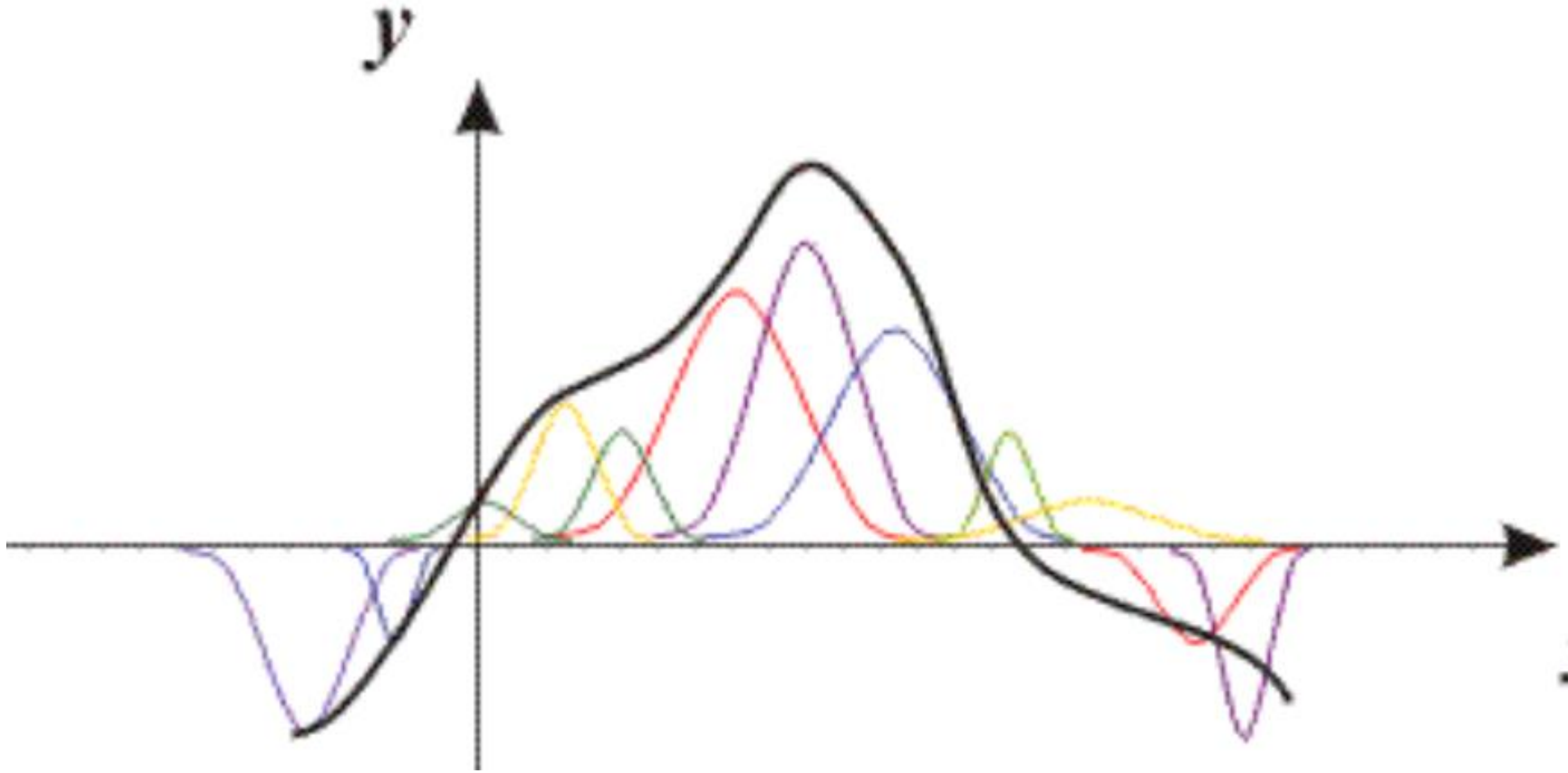
REDE DE FUNÇÃO DE BASE RADIAL

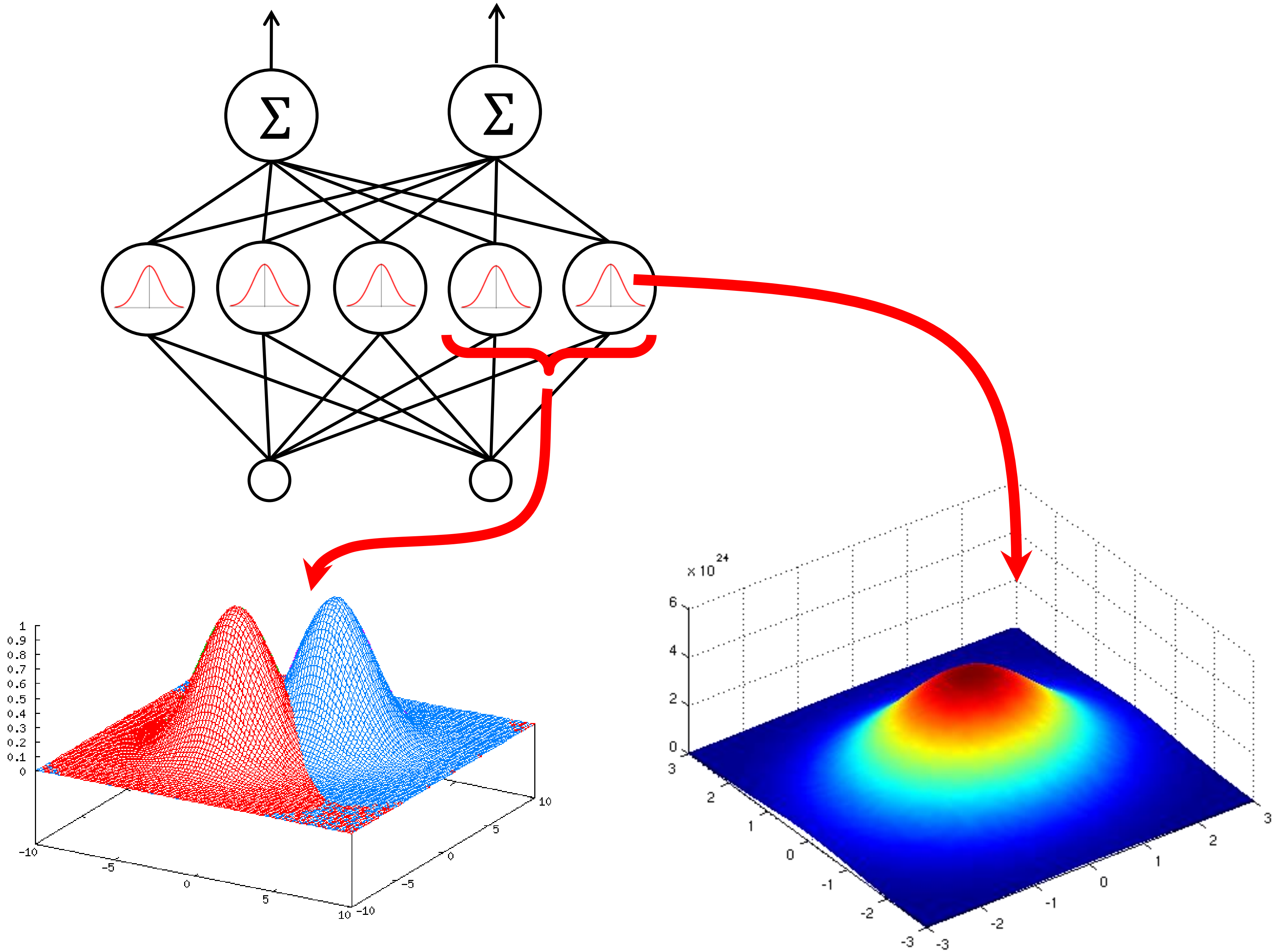


$$f(\mathbf{x}) = \sum_{i=1}^{m} w_i \phi_i(\mathbf{x})$$

REDE DE FUNÇÃO DE BASE RADIAL

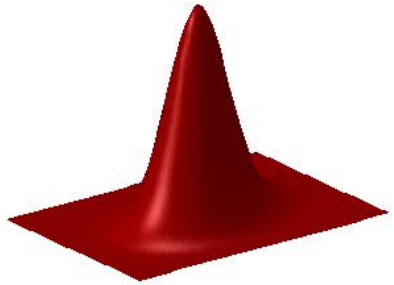
- Aproximador universal de funções através da soma ponderada de funções de base radial.




Σ
Σ
x 10^24

PROCESSAMENTO DA REDE

- Unidades escondidas recebem o vetor de entrada
- Camada de entrada: propaga os estímulos.
- Camada escondida: unidades de processamento localmente sintonizáveis:
- $z_i(t) = \varphi_i(t, c_i, \sigma_i)$
  - onde $\varphi_i(.)$ é uma função de base radial
  - c_i é o j-ésimo centro e
  - σ_i é a largura do campo receptivo para o centro.



PROCESSAMENTO DA REDE

- Camada de saída: unidades de processamento lineares.
- $y_j(t) = \sum_{i=1}^{m} w_{ji}\varphi_i\,(t)$

- 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.

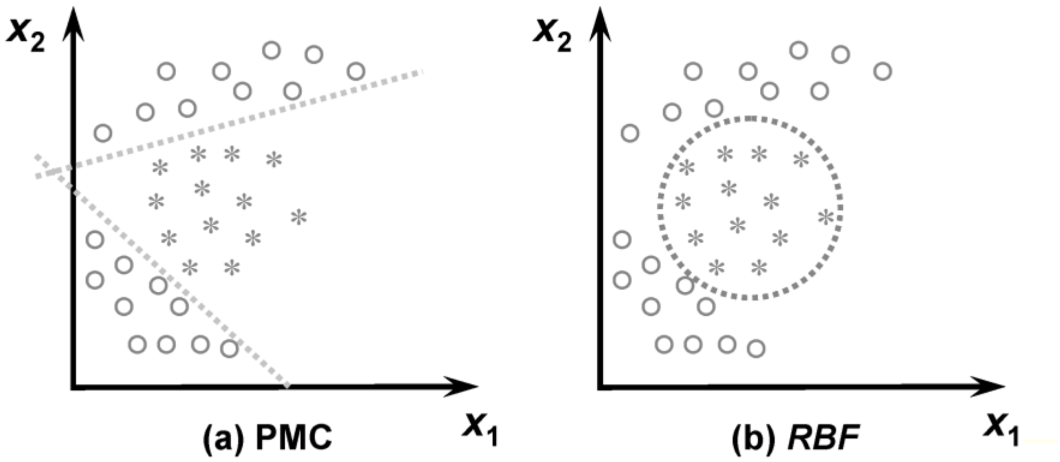
APRENDIZAGEM EM REDES RBF

- O método de treinamento, como os demais modelos supervisionados, deve reduzir o erro na saída da rede a valores aceitáveis através da adaptação dos parâmetros livres na rede RBF: centros, larguras dos campos receptivos e pesos entre a camada escondida e a de saída. A aprendizagem pode ser supervisionada, não-supervisionada ou híbrida.
- O treinamento híbrido combina aprendizagem não supervisionada (ANS) com supervisionada (AS). Este é o tipo de treinamento mais empregado pois, em geral, não se sabe que saídas se desejam para a camada escondida
  - ANS: treina a camada escondida definindo seus parâmetros livres.
  - AS: determina os valores dos pesos entre as camadas escondida e de saída, considerando constantes os parâmetros já definidos.

COMPUTAÇÃO DA REDE RBF

- O aprendizado em uma Rede RBF consiste em determinar os valores para:
  - centro das funções de base radial,
  - largura das funções, e
  - pesos da camada de saída.

- Cada neurônio da camada escondida computa uma
função base radial
  - Centro
    - Protótipo para um *cluster*.
  - Largura
    - Área de influência de um *cluster*.



TREINAMENTO

- É dividido comumente em 3 etapas:
  - na primeira etapa, utiliza-se um algoritmo não-supervisionado para encontrar os centros das funções de base radial;
  - na segunda etapa utilizam-se métodos heurísticos para determinar a largura de cada função;
  - terceira etapa, utiliza-se um algoritmo supervisionado para a determinação dos pesos da camada de saída da rede.

APRENDIZAGEM EM REDES RBF

1. Determine os valores dos centros por:
   - (a) seleção aleatória;
   - (b) distribuição sobre um intervalo pré-definido
   - (c) técnica de agrupamento (*clustering*);
   - (d) ou outro algoritmo.

APRENDIZAGEM EM REDES RBF

2. Determine a largura do campo receptivo através de uma heurística:

   a) As funções utilizam um único valor de largura que é utilizado por todos os neurônios:

      a) $\sigma = \frac{dist_{max}(c_i,c_j)}{\sqrt{2m}}, \forall\, i \neq j$

      Onde m é o número de centros.

   a) Cada neurônio possui uma largura própria:

      a) $\sigma_i = \frac{dist_{min}(c_i,c_j)}{2}, \forall\, i \neq j$

      b) $\sigma_i = \frac{\sum_k^m dist(c_i,c_k)}{m}, \forall\, i \neq k \; e \; k = 1, \ll k \ll m$

      Onde k é o número de centros mais próximos.

3. Define-se o grau de ativação de cada neurônio da camada oculta utilizando distância euclidiana (u):

   a) $u(t) = \|x(t) - c_i(t)\|, \; i = 1,..,H$

APRENDIZAGEM EM REDES RBF

4. Mapeamento do espaço não-linear (alguma função de base radial):
   - Por exemplo: $\varphi_i(t) = e^{-\frac{{u_i}^2}{2{\sigma_i}^2}}$

5. Cálculo da saída: $y_j(t) = \sum_{i=1}^{m} w_{ji}\varphi_i\ (t)$

6. Cálculo das saídas (O):
   - $o_k(t) = \begin{Bmatrix} 1, & y_j(t) \geq 0 \\ \hline 0, & y_j(t) < 0 \end{Bmatrix}$

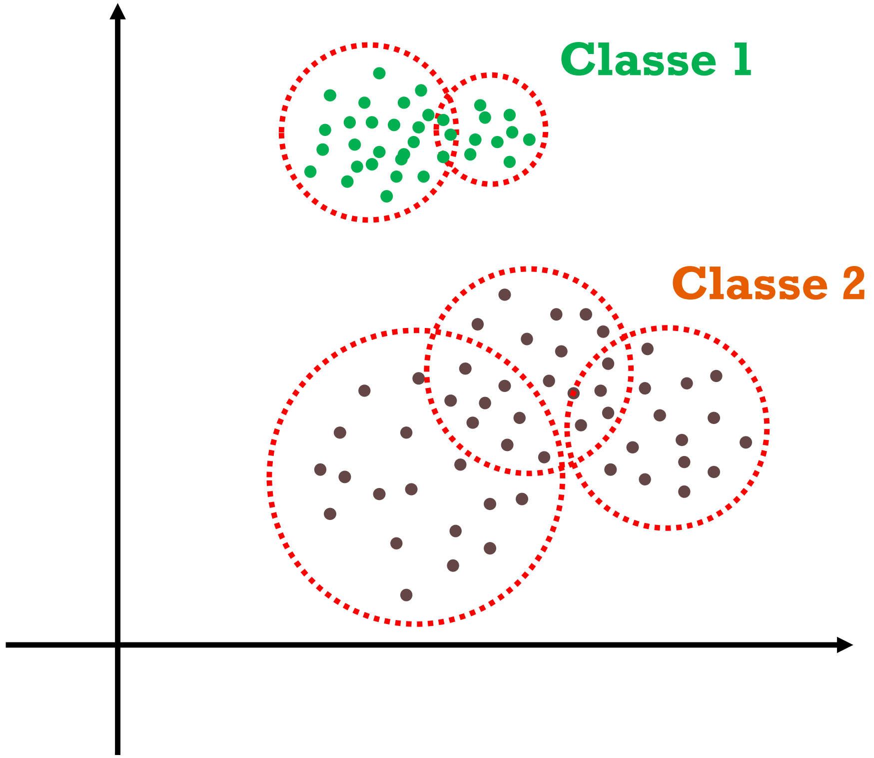
APRENDIZAGEM EM REDES RBF

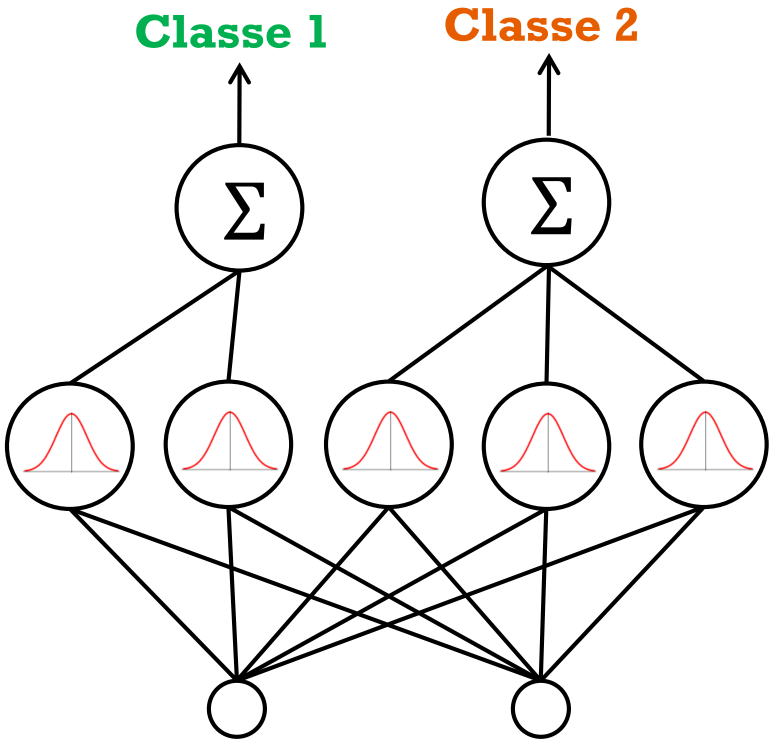
7. Cálculo de erro (e): $e_k(t) = d_k(t)$- $o_k(t)$

8. Ajuste dos pesos (w): $w_{ki}(t+1) = w_{ki}(t) + \eta e_k(t)\varphi_i(t)$

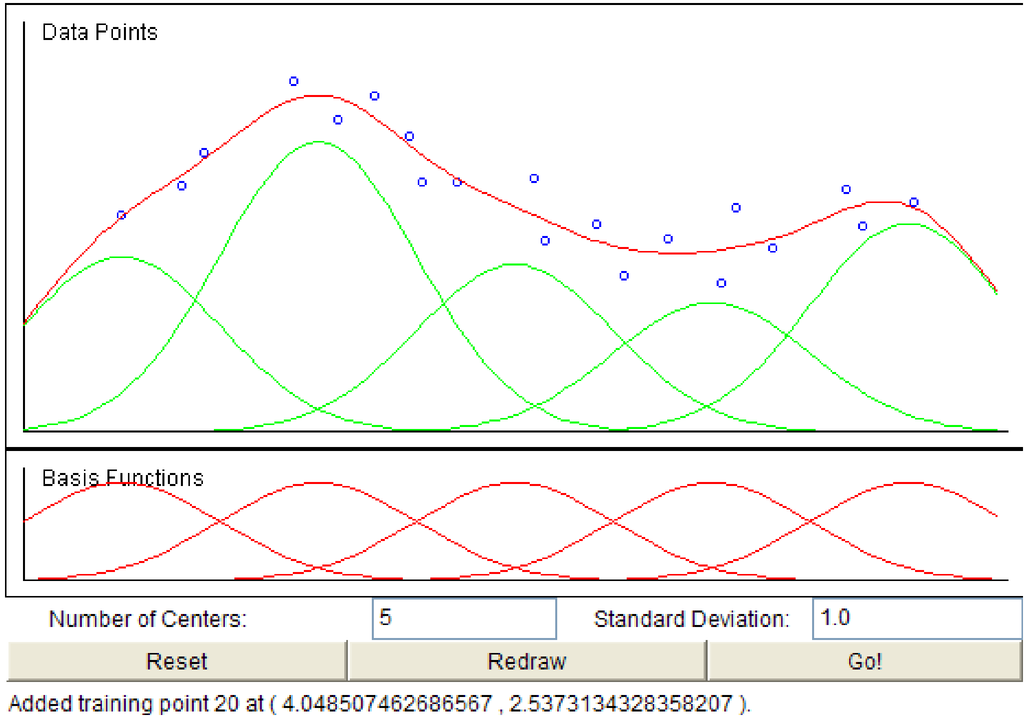
9. Condição de parada (E):
   - Atinge a convergência quando não houver mudanças significativas nos pesos
   - E $= \frac{1}{N}\sum_1^N (e_k\ (t))^2$

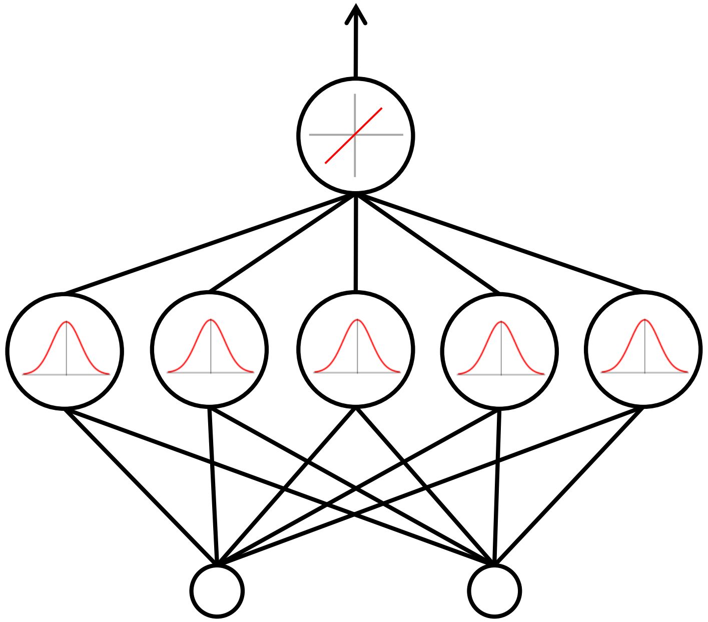
RBF PRA CLASSIFICAÇÃO





RBF PRA REGRESSÃO





http://diwww.epfl.ch/mantra/tutorial/english/rbf/html/

PERGUNTAS?