Departamento de Ciência da COMPUTAÇÃO

UDESC Joinville

Rede RBF
(*Radial Basis Function*)

André Tavares da Silva

andre.silva@udesc.br

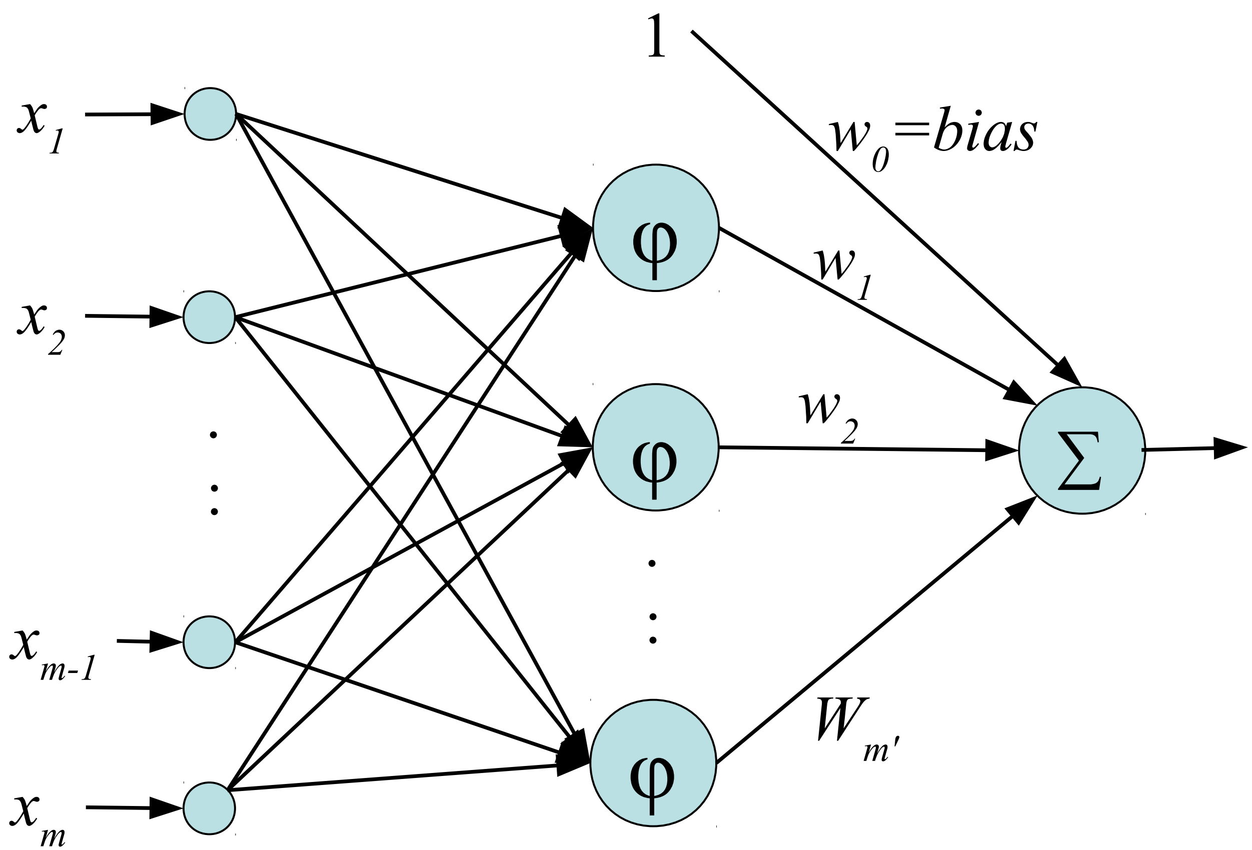
Roteiro

- Introdução à rede neural artificial RBF
- Teorema de Cover da separabilidade de padrões
- RBF x MLP
- RBF
  - Função de ativação
  - Definição da camada oculta
  - Treinamento da camada de saída
- Exemplos

RNA RBF

- Uma rede neural com Função de Ativação de Base Radial (RBF) consiste em um modelo neural multicamadas, capaz de aprender padrões complexos e resolver problemas não-linearmente separáveis.
- A arquitetura de uma rede RBF tem três camadas: camada de entrada, na qual os padrões são apresentados à rede; a camada intermediária (única) que aplica uma transformação não linear do espaço de entrada para o espaço escondido (alta dimensionalidade); e camada de saída que fornece a resposta da rede ao padrão apresentado.

RNA RBF



entradas

Camada escondida
(funções de base radial)

Camada de saída

Teorema de Cover

- Um problema de classificação de padrões que "cai" num espaço de alta dimensão é mais provável ser linearmente separável do que em espaço de baixa dimensão (teorema de Cover da separabilidade de padrões): razão porquê a dimensão do espaço escondido de uma rede RBF ser alta.
- Quando se tem padrões linearmente separáveis (em uma dimensão maior), o problema de classificação torna-se mais simples.

O problema do XOR (novamente)

- Construiremos um classificador de padrões que produza 0 (zero) para entradas (1,1) ou (0,0) e 1 para (0,1) e (1,0).
- Para camada oculta usaremos as funções gaussianas:

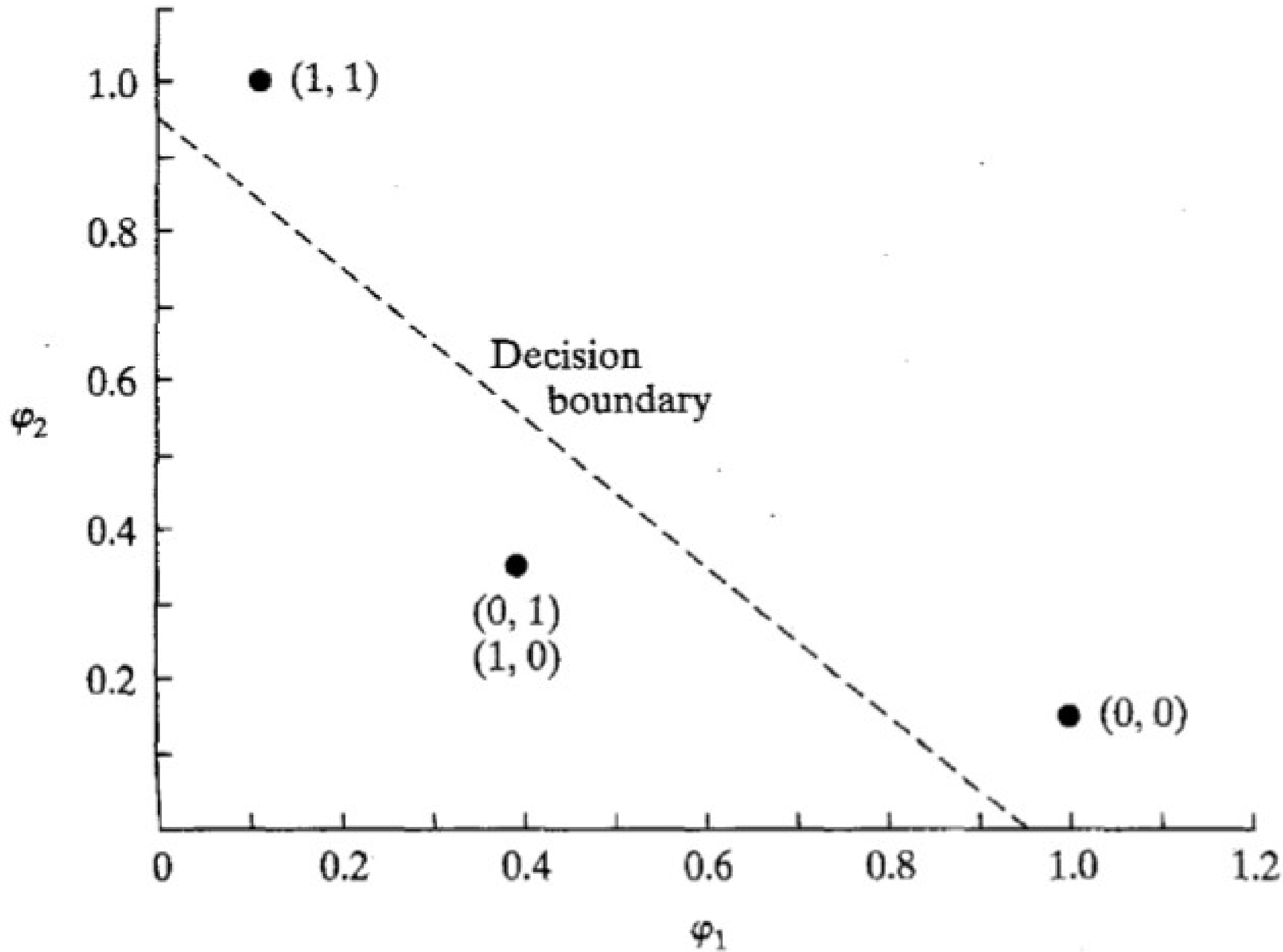
$$\varphi 1(x) = e^{-\|x-t_1\|^2}$$

$$\varphi 2(x) = e^{-\|x-t_2\|^2}$$

onde t_1=(1,1) e t_2=(0,0)

x	φ1	φ2
(0,0)	1	0.1353
(0,1)	0.3678	0.3678
(1,0)	0.3678	0.3678
(1,1)	0.1353	1

Diagrama de decisão do problema XOR



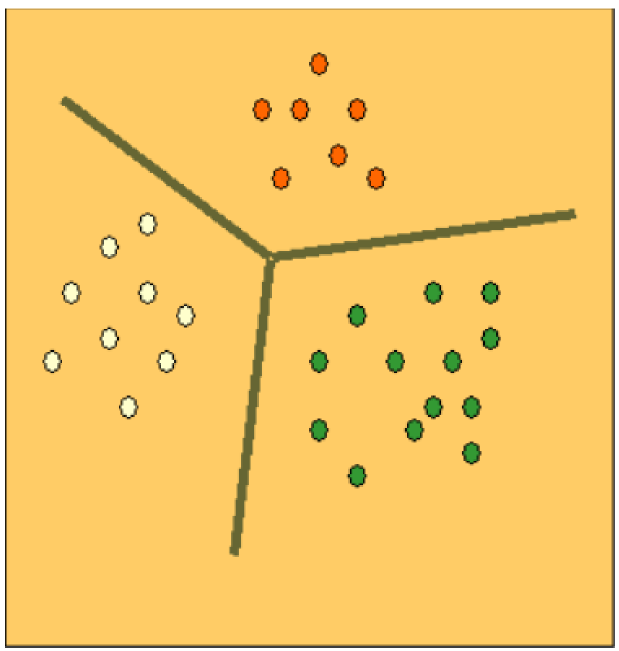
Rede RBF

- Nas redes de Função de Base Radial (RBF), a função de ativação de cada neurônio da camada escondida é função da distância entre seus vetores de peso e de entrada.
- É uma evolução da MLP
- Redes de duas camadas:
  - Primeira camada: Utiliza funções de ativação não lineares (funções de base radial).
  - Segunda camada: Utiliza funções de ativação lineares.

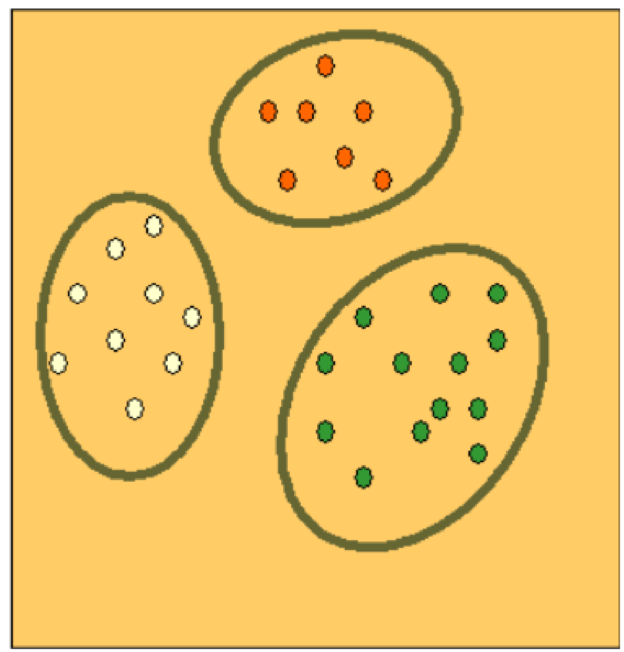
Rede RBF

- A diferença principal entre e RBF é que a primeira utiliza hiperplanos para particionar espaço de entradas enquanto a segunda utiliza hiperelipsóides para particionar o espaço de entradas (na camada escondida).

MLP



RBF



MLP x RBF

- Uma rede RBF (na maioria dos casos) tem uma única camada escondida.
- Os nós da MLP, localizados nas camadas escondidas e de saída, compartilham um modelo neuronal comum. Já na rede RBF, os nós da camada escondida são calculados diferentemente e têm um propósito diferente dos de saída.
- A camada escondida de uma rede RBF é não-linear e a saída é linear.

MLP x RBF

- O argumento da função de ativação de cada unidade escondida numa rede RBF calcula a distância (euclidiana ou não) entre o vetor de entrada e o centro da unidade. Na MLP é calculado produto interno do vetor de entrada e do vetor de pesos sinápticos da unidade.
- Redes RBF normalmente usam não-linearidades localizadas exponencialmente decrescentes (ex.: funções gaussianas) gerando aproximações locais.

MLP x RBF

- Uma rede MLP frequentemente tem muitas camadas de pesos e um complexo padrão de conectividade. Além disso, uma variedade de diferentes funções de ativação podem ser utilizadas na mesma rede. Uma rede RBF, no entanto, geralmente tem uma arquitetura simples.
- A performance de generalização de uma rede MLP é em geral mais robusta.

MLP x RBF

- Todos os parâmetros em uma rede MLP são usualmente determinados ao mesmo tempo, como parte de uma única estratégia global de treinamento, envolvendo treinamento supervisionado de alto custo computacional pela necessidade de retropropagação do erro. Já uma rede RBF é tipicamente treinada em dois estágios: as funções de base radial sendo determinadas primeiramente por meio de técnicas não-supervisionadas; e a segunda camada determinada por métodos lineares supervisionados de convergência rápida.

RBF

- Cada neurônio da camada oculta calcula uma função base radial
  - centro: protótipo de um *cluster*
  - largura: área de influência do protótipo
- Entrada total

$$u = \|x_i - t_i\| \qquad \text{(camada oculta)}$$

$$u = \sum w_i \varphi_i \|x_i - t_i\| \qquad \text{(camada de saída)}$$

- Medida de distância normalmente é a Euclidiana

RBF

- Estados de ativação:

  1 (+1) = ativo.

  0 (-1) = inativo.

- Função da primeira camada é transformar conjunto de dados não-linearmente separáveis em linearmente separáveis.
- Função de ativação das unidades escondidas:
  – Não linear.
  – Valor aumenta ou diminui com relação à distância a um ponto central.

Funções de Ativação

Lâmina *spline* fina	$\phi(\zeta) = \frac{\zeta}{\sigma^2} \log\left(\frac{\zeta}{\sigma}\right)$
Multi-Quadrática	$\phi(\zeta) = \sqrt{\zeta^2 + \sigma^2}$
Multi-quadrática inversa	$\phi(\zeta) = \frac{1}{\sqrt{\zeta^2 + \sigma^2}}$
Gaussiana	$\phi(\zeta) = \exp\left(-\frac{\zeta^2}{2\sigma^2}\right)$

Funções de Ativação

- Escolha da função depende de:
  - Nível de conhecimento sobre os dados: as 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.
- Treinamento é feito em dois estágios
  - Definição da camada oculta (não supervisionado)
  - Treinamento da camada de saída (supervisionado)

Definição da camada oculta

- Determinar os parâmetros das funções de base radial:
  - Número de bases,
  - Centros das bases,
  - Larguras das bases.
- Definições de centros:
  - Existem várias abordagens: seleção aleatória dos centros ou *clustering* (K-means, SOM, algoritmos genéticos,...).

Definição da camada oculta

- Número de funções base:
  - Geralmente definido por tentativa e erro.
  - Sejam m o número de funções base, n o tamanho de Z_1 e c a quantidade de classes: $c < m \ll n$ (m=n leva a overfitting e m=c não funciona se alguma classe tiver mais de uma região associada).
  - Deve ser determinado pela complexidade dos dados.
  - Número de funções base radial = número de *clusters*.

Definição dos centros

- K-*means clustering*:
  - Centros são colocados no meio de agrupamentos de vetores de entrada (*clusters*).
  - Utiliza aprendizado não-supervisionado.
  - Divide os vetores de entrada em K conjuntos disjuntos S_j: cada conjunto S_j tem N_j vetores.
  - Objetivo: minimizar distâncias entre vetores de S_j e seu centro.

Definição das larguras

- Heurísticas para definir larguras σ das funções radiais

  1) Atribuir a σj um valor constante (geralmente 1).

  2) 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.

$$\sigma = \frac{1}{m}\sum_{i=1}^{m} \|t_i - t_j\|$$

onde t_i é o centro mais próximo de t_j.

Definição das larguras

- Heurísticas para definir larguras σ das funções radiais

  3) Atribuir a cada unidade uma largura diferente baseada na distância do seu centro ao centro da unidade mais próxima $\sigma_i = \alpha\|t_i - t_j\|$ onde t_i é o centro mais próximo de t_j e $1.0 < \alpha < 1.5$

  4) Atribuir a cada σ_j a distância média de seu centro aos N vetores de entrada mais próximos.

Treinamento da camada de saída

- Determinar pesos da camada de saída:
  – Recebe vetores linearmente separáveis,
  – Supervisionado,
  – Classificação/regressão dos vetores de entrada.
- Métodos para ajustar pesos:
  – Decomposição em valores singulares,
  – Regra *delta*.

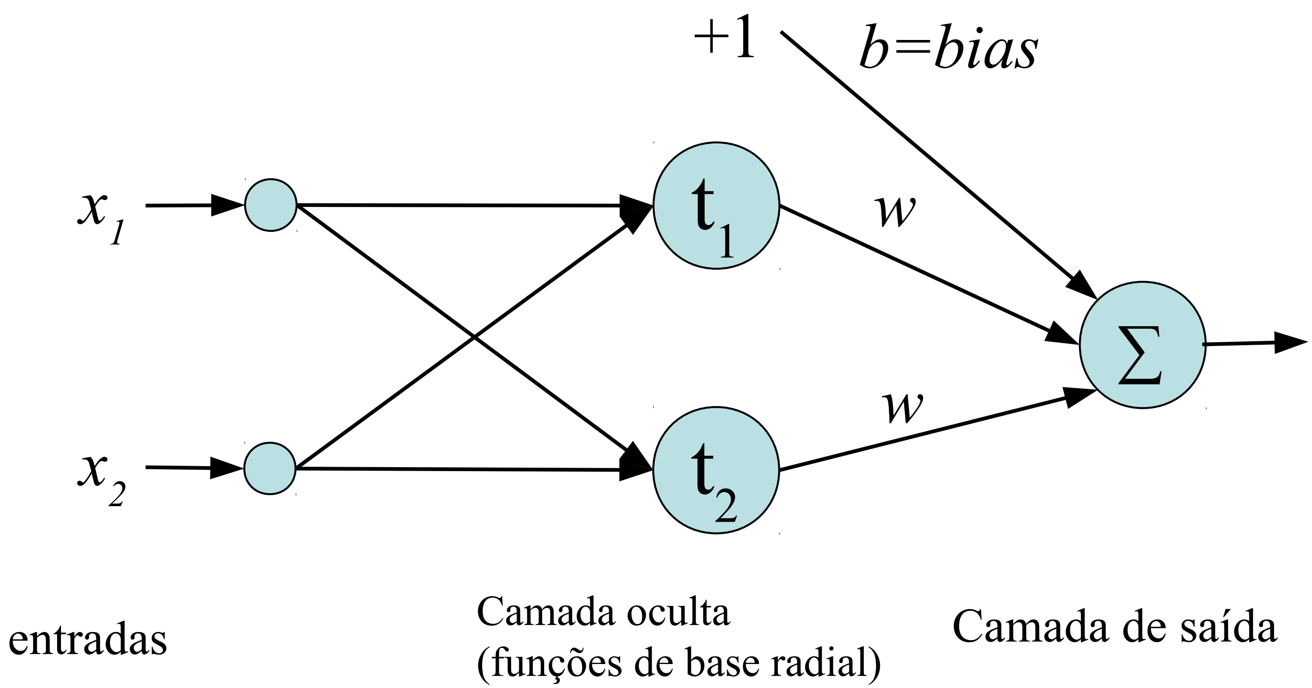
XOR usando uma rede RBF

- Os centros t_1 e t_2 são: $t_1 = [1, 1]^T$ e $t_2 = [0, 0]^T$
- Par de funções gaussianas: $G(\|x - t_i\|) = e^{-\|x - t_i\|^2}$, i=1,2
- Para a unidade de saída, assume-se:
  - usa compartilhamento de pesos, devido à simetria do problema, ou seja, ambas as unidades escondidas tem o mesmo peso *w;*
  - e a unidade de saída inclui um bias *b*.

Departamento
de Ciência da
COMPUTAÇÃO

UDESC
Joinville

XOR usando uma rede RBF



XOR usando uma rede RBF

- A relação entrada-saída da rede é definida por: $y(x)=\sum_{i=1}^{2} wG(\|x-t_i\|)+b$

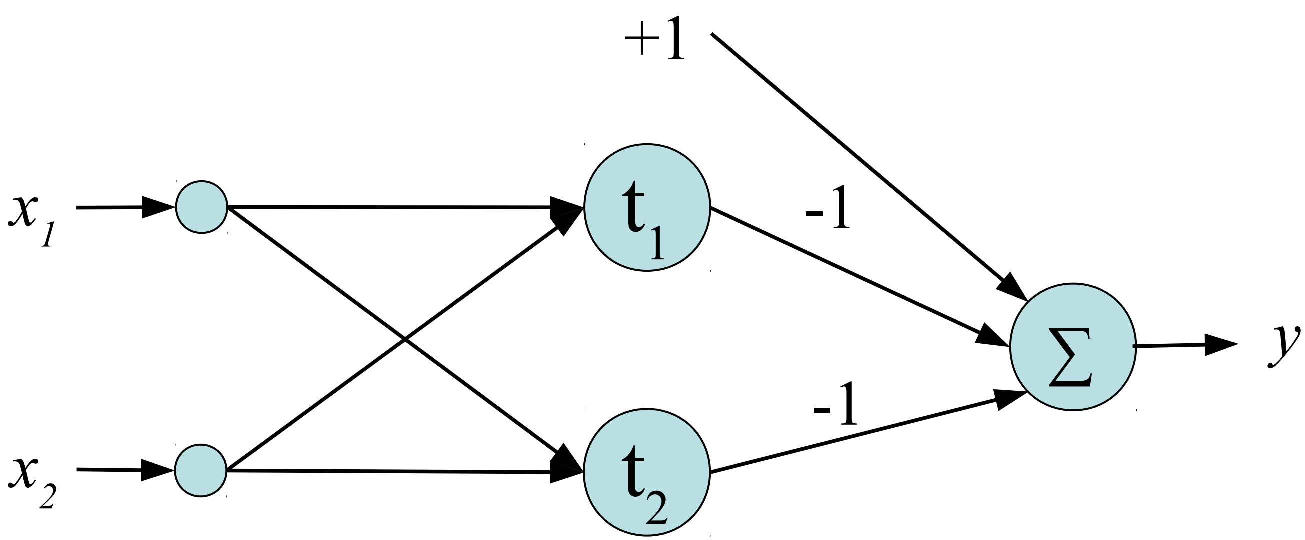
- Para ajustar os dados de treinamento é necessário que $\boldsymbol{y(x_j) = d_j}$, $\boldsymbol{j=1,2,3,4}$ onde $\boldsymbol{x_j}$ é um vetor de entrada e $\boldsymbol{d_j}$ é o valor desejado na saída.

j	x_j	d_j
1	(0,0)	0
2	(0,1)	1
3	(1,0)	1
4	(1,1)	0

XOR usando uma rede RBF

- Sendo t_1=(1,1), t_2=(0,0), w=–1 e b=+1, tem-se

  $$y = -e^{-\|x - t_1\|^2} - e^{-\|x - t_2\|^2} + 1$$



- Se $y > 0$ então é da "classe" 1, senão é 0 (zero).

Diagrama de decisão do problema XOR