

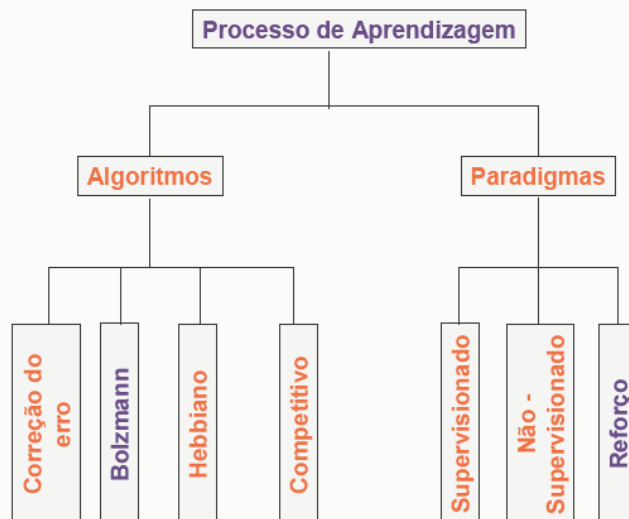
## APRENDIZADO EM MODELOS CONEXIONISTAS

- Aprendizado é o processo pelo qual os parâmetros (os pesos das conexões entre os neurônios) de uma rede neural são ajustados através de uma forma continuada de estímulo pelo ambiente no qual a rede está operando.
- Aprendizado também é chamado como *treinamento*.

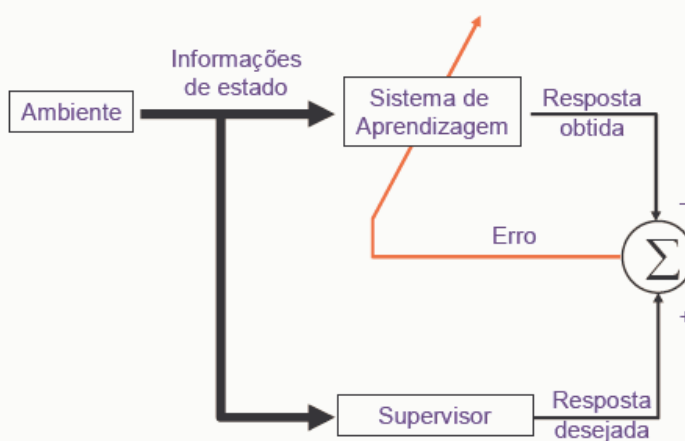
## APRENDIZADO EM MODELOS CONEXIONISTAS

- O processo de aprendizado implica a seguinte seqüência de eventos:
  1. A rede neural é estimulado por um ambiente.
  2. A rede neural sofre modificações nos seus parâmetros livres como resultado desta estimulo.
  3. A rede neural responde de uma maneira nova ao ambiente, devido às modificações ocorridas na sua estrutura interna.
- Um conjunto de regras bem-definidas para a solução de um problema de aprendizagem é denominado um *algoritmo de aprendizagem* (ou *algoritmo de treinamento*)

### Taxonomia do Processo de Aprendizagem (Haykin, 1994)

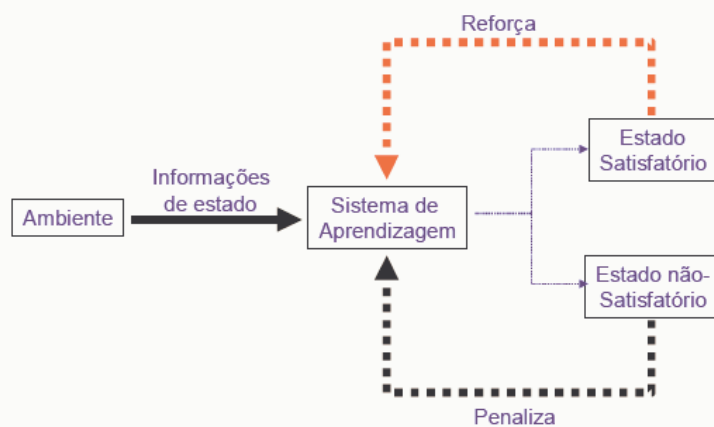


### Aprendizagem Supervisionada



### Aprendizagem por Reforço (on - line)

- Idéia Básica (Baseada na Lei do efeito de Thorndike - Haykin, 1994)



### Aprendizagem Não-Supervisionada



- Sem exemplos específicos
- Os pesos são otimizados segundo uma medida independente da tarefa que indica a qualidade da representação que a rede deve aprender.
- Condição de parada: aprendizagem pára quando a rede está sintonizada com as regularidades estatísticas dos dados de entrada.

## PARADIGMAS PRINCIPAIS

Aprendizado supervisionado

Perceptron, MLP, RBF, etc.

- Aprendizado não supervisionado

SOM, ICA, ART1, Fuzzy ART, Modelo de Hopfield, etc.

- Aprendizado por reforço

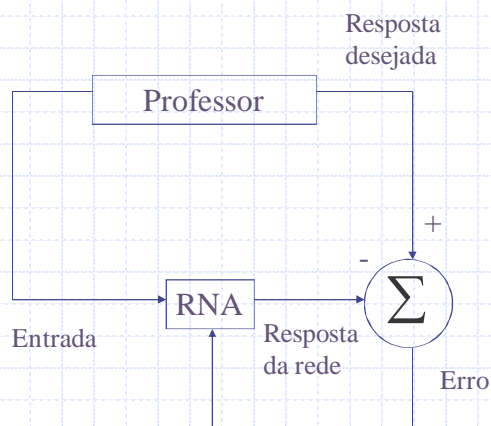
Programação Dinâmica por Redes Neurais, etc.

## APRENDIZADO SUPERVISIONADO

- A entrada e saída desejadas para a rede são fornecidas por um supervisor. Objetivo é ajustar os parâmetros da rede, de forma a encontrar uma ligação entre os pares de entrada e saídas fornecidas.
- O professor indica explicitamente um comportamento bom ou ruim para a rede, visando a direcionar o processo de treinamento.
- A técnica mais utilizada é chamada *correção de erros*.

## APRENDIZADO SUPERVISIONADO

### Correção de Erros



## APRENDIZADO SUPERVISIONADO

### Correção de Erros

- Processo de minimização do erro quadrático pode ser realizado pelo método do *Gradiente Descendente*

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}}$$

- Cada peso sináptico  $i$  do elemento processador  $j$  é atualizado proporcionalmente ao *negativo da derivada parcial do erro* deste processador com relação ao peso.

## APRENDIZADO SUPERVISIONADO

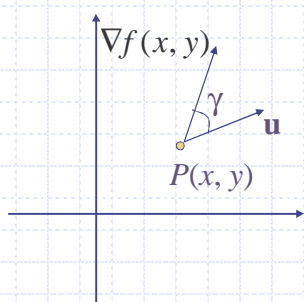
### Relembra um pouco conhecimento de Cálculo

Gradiente:  $\nabla f(x, y) = \left( \frac{\partial}{\partial x} f(x, y), \frac{\partial}{\partial y} f(x, y) \right)$

Derivada direcional:  $D_{\mathbf{u}} f(x, y) = \nabla f(x, y) \cdot \mathbf{u}$

$$= \|\nabla f(x, y)\| \|\mathbf{u}\| \cos \gamma$$

$$= \|\nabla f(x, y)\| \cos \gamma$$



$D_{\mathbf{u}} f(x, y)$  é a taxa de variação de  $f(x, y)$  na direção definida por  $\mathbf{u}$ .

## APRENDIZADO SUPERVISIONADO

### Relembra um pouco conhecimento de Cálculo (cont.)

**Teorema do gradiente:** Seja  $f$  uma função de duas variáveis, diferenciáveis no ponto  $P(x, y)$ .

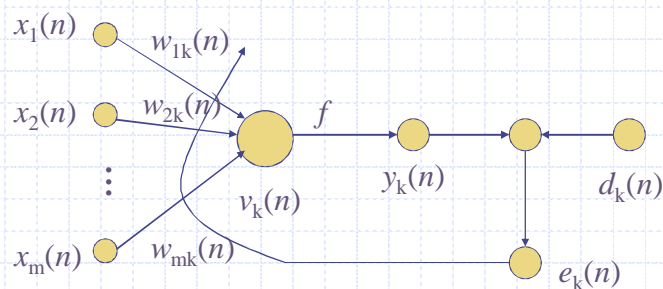
- i) O máximo de  $D_{\mathbf{u}} f(x, y)$  em  $P(x, y)$  é  $\|\nabla f(x, y)\|$ .
- ii) O máximo da taxa de crescimento de  $f(x, y)$  em  $P(x, y)$  ocorre na direção de  $\nabla f(x, y)$ .

**Corolário:** Seja  $f$  uma função de duas variáveis, diferenciáveis no ponto  $P(x, y)$ .

- i) O mínimo de  $D_{\mathbf{u}} f(x, y)$  em  $P(x, y)$  é  $-\|\nabla f(x, y)\|$ .
- ii) O máximo da taxa de decréscimo de  $f(x, y)$  em  $P(x, y)$  ocorre na direção de  $-\nabla f(x, y)$ .

## APRENDIZADO SUPERVISIONADO

### Correção de Erros



Valores de entrada e saída: binário

$$y_j = f(v_j) = f\left(\sum_i x_i w_{ij} + \theta_i\right) = \begin{cases} 1 & v_j > 0 \\ 0 & v_j \leq 0 \end{cases}$$

## APRENDIZADO SUPERVISIONADO

### Calcula $\Delta w_{ij}$

$$\Delta w_{ij} = -\eta \frac{\partial E_p}{\partial w_{ij}} = -\eta \frac{\partial E_p}{\partial y_j} \frac{\partial y_j}{\partial w_{ij}}$$

$$E_p = \frac{1}{2} \sum_j (d_j - y_j)^2$$

$$y_j = \sum x_i w_{ij} + \theta_j$$

$$2 \times \frac{1}{2} \times (d_j - y_j) \times (-1)$$

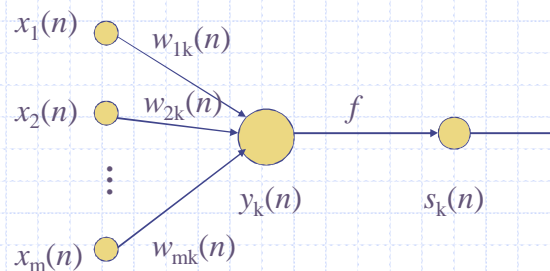
 $x_i$ 

$$\Delta w_{ij} = \eta (d_j - y_j) x_i = \eta e_j x_i$$

## APRENDIZADO HEBBIANO

- Se dois neurônios em ambos os lados de uma sinapse (conexão) são ativados simultaneamente (sincronamente), então a força daquela sinapse é seletivamente aumentada;
- Se dois neurônios em ambos os lados de uma sinapse são ativados assincronamente, então a força daquela sinapse é seletivamente enfraquecida ou eliminada

## APRENDIZADO HEBBIANO





## APRENDIZADO HEBBIANO

Forma geral:

$$\Delta w_{kj} = F(s_k(n), x_j(n))$$

Forma correlativa

$$\Delta w_{kj} = \eta s_k(n) x_j(n)$$

Forma covariância

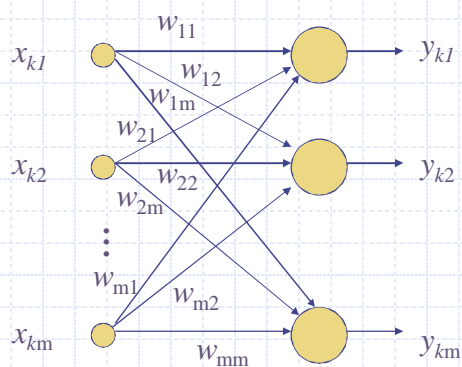
$$\Delta w_{kj} = \eta \left( x_j - \bar{x} \right) \left( s_k - \bar{s} \right)$$

$\eta$ : taxa de aprendizado

$\bar{x}$ ,  $\bar{s}$ : valor médios, que constituem os limiares pré-sináptico e pós-sináptico

## APRENDIZADO HEBBIANO

### Aprendizado Hebbiano em Associador Linear



## APRENDIZADO HEBBIANO

### Aprendizado Hebbiano em Associador Linear

$$\mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{km}]^T$$

$$\mathbf{y}_k = [y_{k1}, y_{k2}, \dots, y_{km}]^T$$

$$\mathbf{y}_k = \mathbf{W}(k)\mathbf{x}_k$$

$$\mathbf{W}(k) = \begin{bmatrix} w_{11}(k) & w_{12}(k) & \dots & w_{1m}(k) \\ w_{21}(k) & w_{22}(k) & \dots & w_{2m}(k) \\ \dots & \dots & \dots & \dots \\ w_{m1}(k) & w_{m2}(k) & \dots & w_{mm}(k) \end{bmatrix}$$

$$\mathbf{M} = \sum_{k=1}^q \mathbf{W}(k) \quad \mathbf{M}_k = \mathbf{M}_{k-1} + \mathbf{W}(k)$$

## APRENDIZADO HEBBIANO

### Aprendizado Hebbiano em Associador Linear

Regra de Aprendizado:  $\hat{\mathbf{M}} = \sum_{k=1}^q \mathbf{y}_k \mathbf{x}_k^T$  Estimativa da matriz de memória M

$$\hat{\mathbf{M}} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m] \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \dots \\ \mathbf{x}_m^T \end{bmatrix}$$

$$\hat{\mathbf{M}}_k = \hat{\mathbf{M}}_{k-1} + \mathbf{y}_k \mathbf{x}_k^T$$

## APRENDIZADO HEBBIANO

### Aprendizado Hebbiano em Associador Linear

Recordação para entrada  $\mathbf{x}_j$ :

$$\begin{aligned}\hat{\mathbf{y}} &= \mathbf{M} \mathbf{x}_j \\ &= \sum_{k=1}^m \mathbf{y}_k \mathbf{x}_k^T \mathbf{x}_j \\ &= \sum_{k=1}^m (\mathbf{y}_k \mathbf{x}_j^T) \mathbf{x}_k \\ &= (\mathbf{x}_j^T \mathbf{x}_j) \mathbf{y}_j + \sum_{\substack{k=1 \\ k \neq j}}^m (\mathbf{y}_k \mathbf{x}_j^T) \mathbf{x}_k\end{aligned}$$

## APRENDIZADO HEBBIANO

### Aprendizado Hebbiano em Associador Linear

$$\begin{aligned}\mathbf{y} &= \mathbf{y}_j + \mathbf{v}_j \\ \mathbf{v}_j &= \sum_{\substack{k=1 \\ k \neq j}}^m (\mathbf{y}_k \mathbf{x}_j^T) \mathbf{x}_k \leftarrow \text{Inferência cruzada}\end{aligned}$$

Assume  $\mathbf{x}_j^T \mathbf{x}_j = 1$

## APRENDIZADO HEBBIANO

### Aprendizado Hebbiano em Associador Linear

$$\begin{aligned}\cos(\mathbf{x}_k, \mathbf{x}_j) &= \frac{\mathbf{x}_k^T \mathbf{x}_j}{\|\mathbf{x}_k\| \|\mathbf{x}_j\|} \\ &= \mathbf{x}_k^T \mathbf{x}_j \\ \mathbf{v}_j &= \sum_{\substack{k=1 \\ k \neq j}}^m \cos(\mathbf{x}_k, \mathbf{x}_j) \mathbf{y}_k\end{aligned}$$

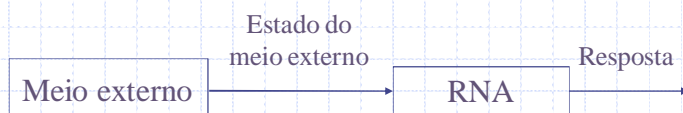
A memória é perfeita (a resposta  $\mathbf{y}$  iguala  $\mathbf{y}_j$ ) se os vetores-chave pertencerem a um conjunto ortonormal.

$$\mathbf{x}_k^T \mathbf{x}_j = \begin{cases} 1, & k = j \\ 0, & k \neq j \end{cases}$$

## APRENDIZADO NÃO SUPERVISIONADO

- Não há um professor ou supervisor para acompanhar o processo de aprendizado;
- A partir do momento em que a rede estabelece uma harmonia com as regularidades estatísticas da entrada de dados, desenvolve-se nela uma habilidade de formar representações internas para codificar características de entrada;
- Este tipo de aprendizado só se torna possível quando existe redundância nos dados de entrada.

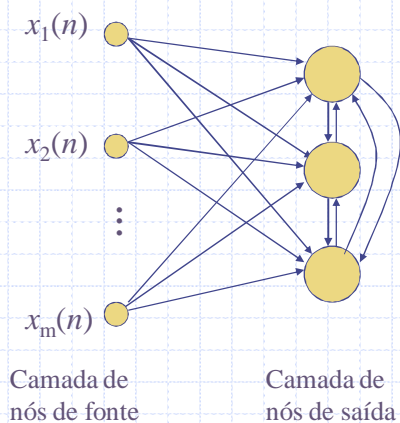
## APRENDIZADO NÃO SUPERVISIONADO



A técnica mais utilizada é chamada *aprendizado competitiva*

## APRENDIZADO NÃO SUPERVISIONADO

### Aprendizado Competitiva



## APRENDIZADO NÃO SUPERVISIONADO

### Aprendizado Competitiva

Existe três elementos básicos em uma regra de aprendizagem competitiva

- 1) Um conjunto de neurônios que são todos iguais entre si, exceto por alguns pesos sinápticos distribuídos aleatoriamente, e por isso respondem diferentemente a um conjunto de padrões de entrada;
- 2) Um limite imposto sobre a “força” de cada neurônio;

## APRENDIZADO NÃO SUPERVISIONADO

### Aprendizado Competitiva

- 3) Um mecanismo que permite que pelo direito de responder a um dado subconjunto de dados, de forma que somente um neurônio de saída, ou somente um neurônio por grupo, esteja ativo em um determinada instante. O neurônio que vence a competição é denominado um neurônio vencedor leva tudo

## APRENDIZADO NÃO SUPERVISIONADO

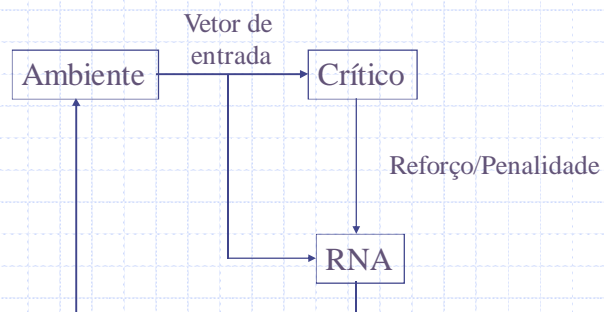
$$y_k = \begin{cases} 1 & \text{se } v_k > v_j \text{ para todos } j, j \neq k \\ 0 & \text{caso contrário} \end{cases}$$

O neurônio  $k$  é denominado como neurônio *vencedor*.

Regra de aprendizagem competitiva:

$$\Delta w_{kj} = \begin{cases} \eta(x_j - w_{kj}) & y_k = 1 \\ 0 & y_k = 0 \end{cases}$$

## APRENDIZADO POR REFORÇO



## APRENDIZADO NÃO SUPERVISIONADO

Se uma ação tomada pelo sistema de aprendizagem é seguida de estados satisfatórios, então a tendência do sistema de produzir esta ação particular é reforçada.

Se não for seguida de estados satisfatórios, a tendência do sistema de produzir esta ação é enfraquecida.

## APRENDIZADO NÃO SUPERVISIONADO

A aprendizagem por reforço é difícil ser realizada por duas razões básicas:

- Não existe um professor para fornecer uma resposta desejada exata em cada passo do processo de aprendizagem;
- O atraso incorrido na geração do sinal de reforço implica que a RNA deve resolver um problema de atribuição de crédito temporal.