

## **Redes Neurais Artificiais**

**Edielson Prevato Frigieri**

***edielson@inatel.br***

***Av. João de Camargo, 510***

***Santa Rita do Sapucaí - MG***

***Tel: (35) 3471-9279***

**Novembro 2013**

## Redes Neurais Artificiais

### Introdução

- Como é possível o cérebro processar tantas informações de imagem, sons e outros dados ao mesmo tempo e de forma eficiente?

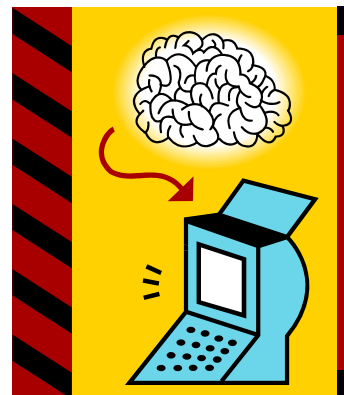


Computador altamente complexo,  
não-linear e paralelo

# O que é uma rede neural?

- Uma rede neural é um processador maciçamente paralelamente distribuído constituído de unidades de processamento simples, que tem a propensão natural de armazenar conhecimento experimental e torná-lo disponível para uso.

Haykin, 2001



## Principais características

- **Não-linearidade:** consegue mapear fenômenos não-lineares, por exemplo, sinal de voz.
- **Capacidade de aprendizado:** aprendizado a partir de uma sequência de treinamento (exemplos) com saídas desejadas associadas à entradas únicas;
- **Adaptabilidade:** capacidade de adaptar seus pesos sinápticos a partir de modificações do meio-ambiente (novo treinamento);
- **Habilidade de generalização:** após o processo de treinamento da rede, essa é capaz de generalizar o conhecimento adquirido (estimar soluções desconhecidas);
- **Organização de dados:** a rede é capaz de realizar a sua organização interna visando possibilitar o agrupamento de padrões;

- **Informação contextual:** cada neurônio da rede é potencialmente afetado pela atividade de todos os outros neurônios na rede.
- **Tolerância à falhas:** perdas de neurônios não representam degradação significativas na respostas de forma global (hardware);
- **Implementação em larga escala:** adequada para implementação utilizando tecnologia de integração em larga escala (natureza paralela);
- **Analogia neurobiológica:** motivado pela semelhança com o cérebro (prova viva de que o processamento paralelo é rápido e poderoso);
- **Facilidade de prototipagem:** pode ser implementada facilmente em software e hardware (processo de execução).

## Resumo histórico

1943

- Primeira publicação relacionada à neurocomputação (McCulloch & Pitts);
- Primeiro modelamento matemático inspirado em um neurônio biológico;

1949

- Primeiro método de treinamento para redes neurais denominado regra de aprendizado de Hebb;

1958

- Primeiro neurcomputador (*Mark I*) desenvolvido por Frank Rosenblatt;

1960

- Widrow & Hoff desenvolveram um tipo de rede denominada *Adaline* cujo aprendizado é fundamentado na chamada regra Delta;

1969

- Minsky & Papert apresentaram a limitação das redes neurais, como *Perceptron* e *Adaline*, com problemas não linearmente separáveis;
- Congelamento da área de RN;



## Resumo histórico, cont.

1982

- Hopsfield propôs a proposição de redes recorrentes baseadas em funções de energia; Retomada da área de RN;

1986

- Publicação do livro *Parallel distributed processing* [Rumelhart *et al*], propondo um algoritmo que permitia ajustar os pesos em uma rede com mais de uma camada (*backpropagation*); Resolveram problemas não linearmente separáveis;

1994

- Algoritmos de aprendizado baseados no método Levenberg-Marquardt que permite incrementar a eficiência do treinamento de redes neurais artificiais [Hagan & Menhaj];

1998

- Redes neurais artificiais baseadas em máquinas de vetores suporte (*support vector machines - SVM*) que podem ser utilizadas em classificação de padrões e regressão linear [Vapnik];

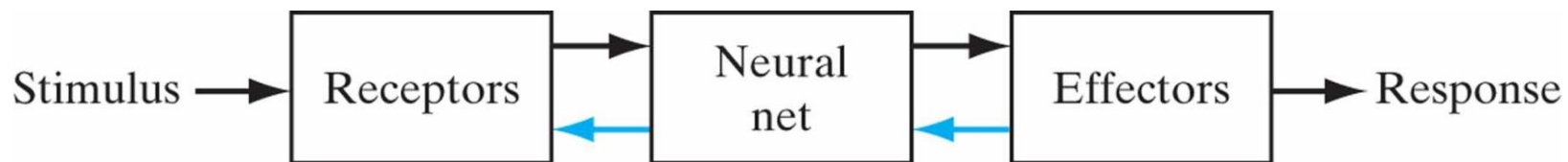
2003

- Implementação de circuitos integrados neurais com diversas configurações de tipologia [Beiu *et al*];



- **Aproximador universal de funções:** tem como objetivo mapear o relacionamento funcional entre as variáveis (reais) de um sistema a partir de um conjunto de valores conhecidos;
  - ✓ Mapeamento de processos diversos
- **Controle de processos:** tem como objetivo identificar ações de controle que permitam o alcance dos requisitos de um processo;
  - ✓ Robótica, aeronaves, elevadores, etc.
- **Reconhecimento / classificação de padrões:** tem como objetivo associar um padrão de entrada para uma das classes previamente definidas;
  - ✓ Reconhecimento de imagens, voz, escrita, etc.
- **Otimização de sistemas:** tem como objetivo minimizar ou maximizar uma função de custo (objetivo);
  - ✓ Otimização restrita, programação dinâmica, otimização combinatorial

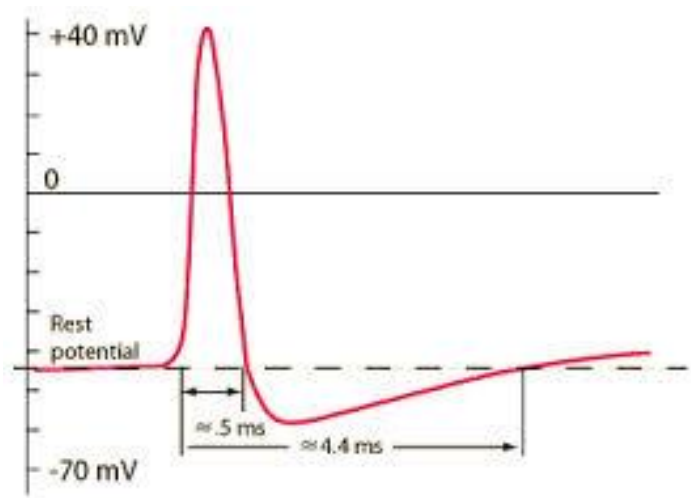
- **Agrupamento de dados (clusterização):** tem como objetivo identificar e detectar semelhanças e particularidades entre os diversos padrões de entrada para efetuar o agrupamento;
  - ✓ Identificação automática de classes
- **Sistemas de previsão:** tem como objetivo estimar valores futuros de um processo levando-se em consideração diversas medidas prévias observadas em seu domínio;
  - ✓ Previsão de séries temporais, mercados financeiros, previsões climáticas
- **Memórias associativas:** tem como objetivo recuperar padrões corretos mesmo se os seus elementos constituintes forem apresentados de forma incerta ou imprecisa;
  - ✓ Processamento de imagens, transmissão de sinais, etc.



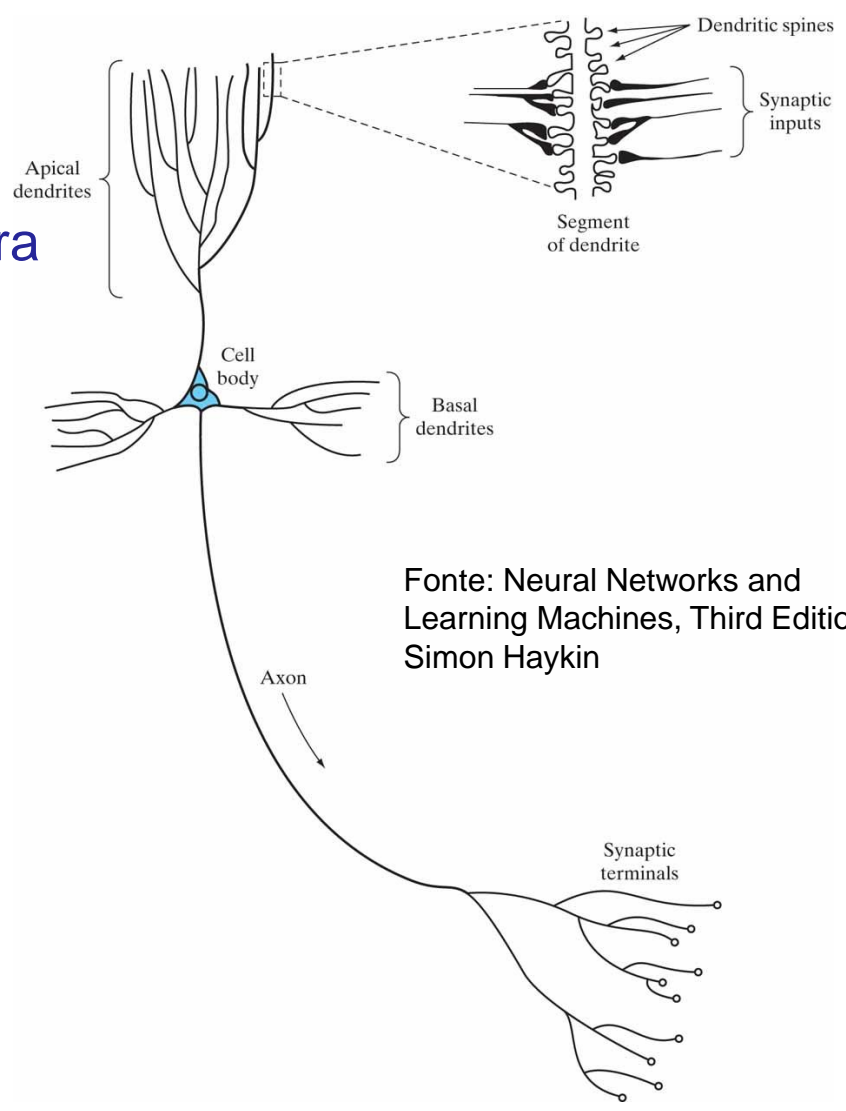
**Figura 1** Representação em diagrama em blocos do sistema nervoso.

## O neurônio

- Unidade de processamento de informação que é fundamental para a operação de uma rede neural.



Potencial de ativação



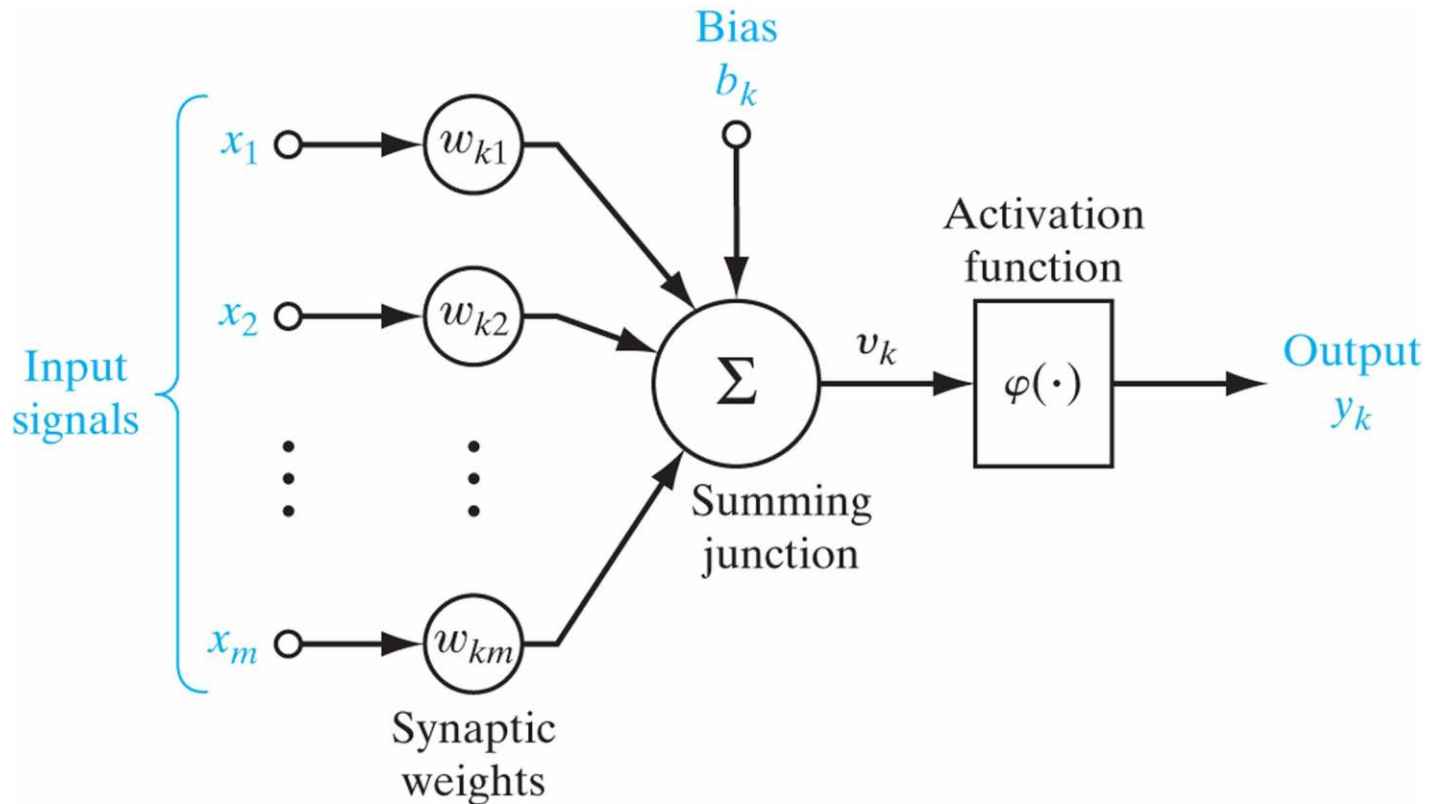
Fonte: Neural Networks and Learning Machines, Third Edition  
Simon Haykin

## Redes Neurais Artificiais

### Neurônio artificial

## Neurônio artificial

- Modelo matemático não-linear de um neurônio  $k$ . (McCulloch & Pitts)



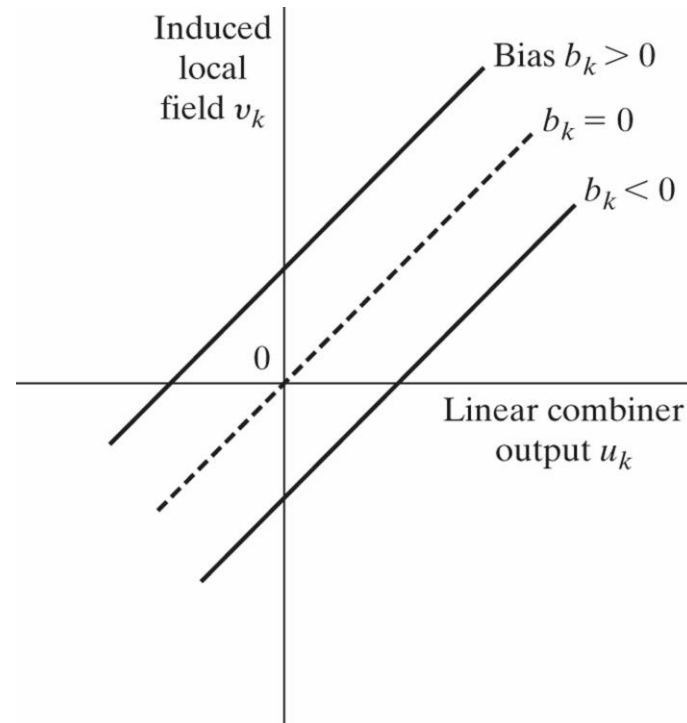
- **Sinais de entrada**  $\{x_1, x_2, \dots, x_m\}$ : sinais ou medidas do meio externo que representam os valores das variáveis;
- **Pesos sinápticos**  $\{w_{k1}, w_{k2}, \dots, w_{km}\}$ : valores que servirão para ponderar cada uma das variáveis de entrada (indicam a relevância de cada entrada na rede);
- **Combinador linear**  $\{\Sigma\}$ : sua função é agregar todos os sinais de entrada que foram ponderados pelos respectivos pesos sinápticos a fim de gerar um valor de potencial de ativação;



## Elementos básicos do modelo, cont.

- **Limiar de ativação  $\{\theta\}$  ou  $\{b_k\}$ :** variável que especifica qual será o patamar apropriado para que o resultado produzido pelo combinador linear possa gerar um valor de disparo em direção a saída do neurônio;

Campo local induzido ou potencial de ativação



- **Potencial de ativação  $\{v_k\}$ :** resultado produzido pela diferença do valor produzido entre o combinador linear e o limiar de ativação. Se tal valor é positivo, ou seja, se  $u_k + b_k \geq 0$  então o neurônio produz um potencial excitatório; caso contrário, o potencial será inibitório;

$$u_k = \sum_{j=1}^m w_{kj} \cdot x_j$$

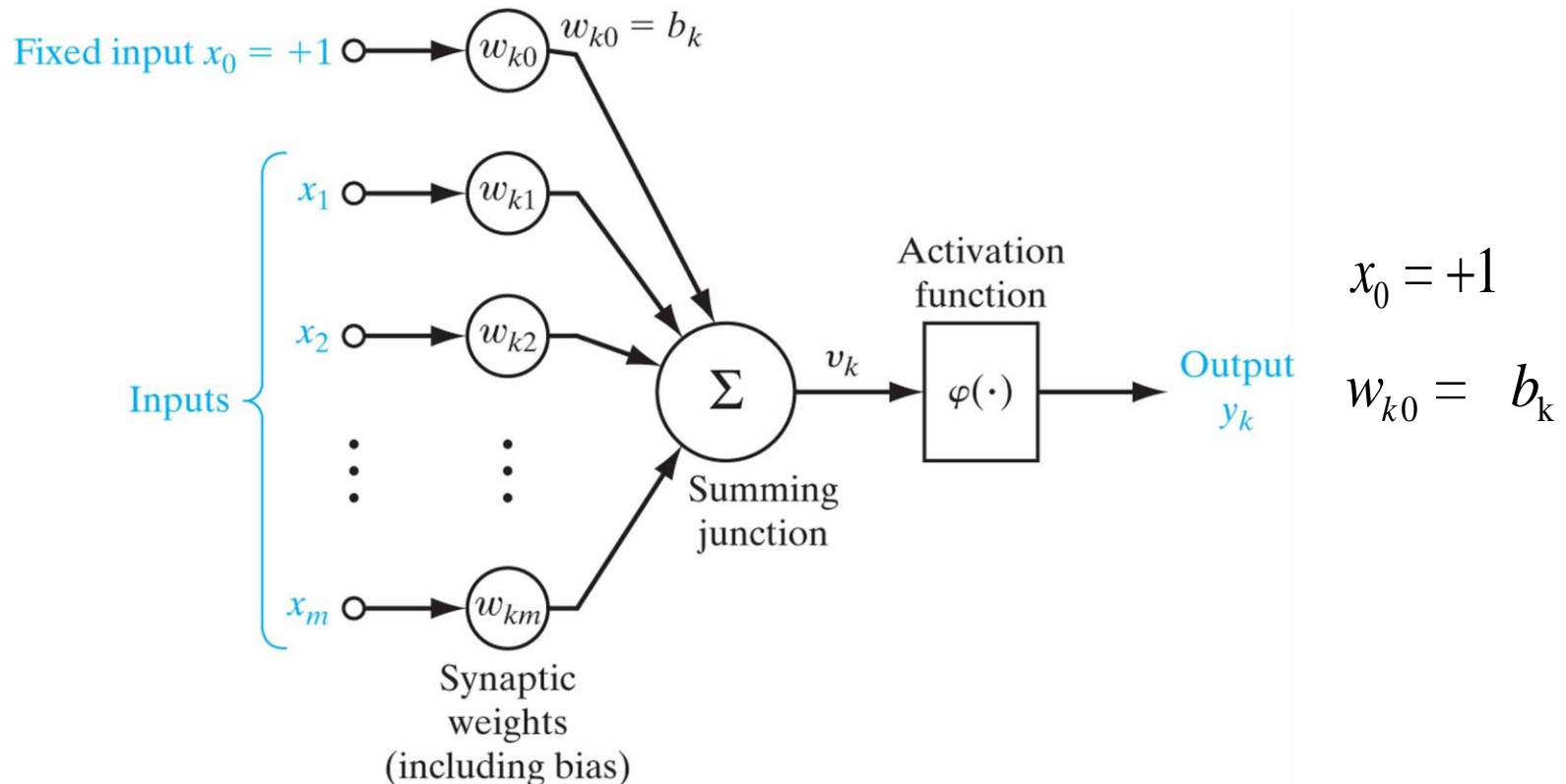
$$v_k = u_k + b_k$$

- **Função de ativação  $\{\varphi\}$ :** seu objetivo é limitar a saída do neurônio dentro de um intervalo de valores razoáveis a serem assumidos pela sua própria imagem funcional;
- **Sinal de saída  $\{y_k\}$ :** valor final produzido pelo neurônio em relação a um determinado conjunto de sinais de entrada, podendo ser também utilizado por outros neurônios que estão sequencialmente interligados.;

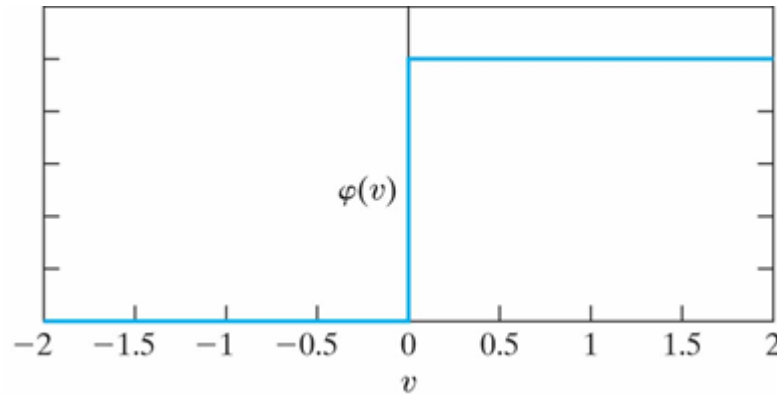
$$y_k = \varphi(u_k + b_k) = \varphi(v_k)$$

## Outro modelo de neurônio

- Outro modelo matemático não-linear de um neurônio

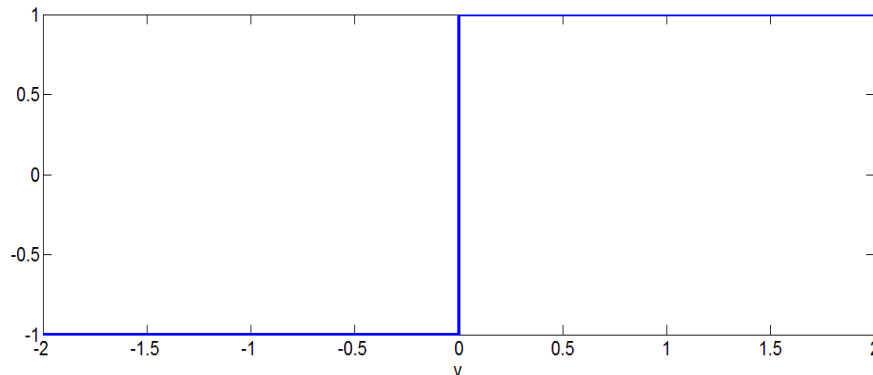


(a) Função de limiar ou função de *heaviside*.



$$\varphi(v) = \begin{cases} 1 & \text{se } v \geq 0 \\ 0 & \text{se } v < 0 \end{cases}$$

(b) Função de *heaviside simétrica*.



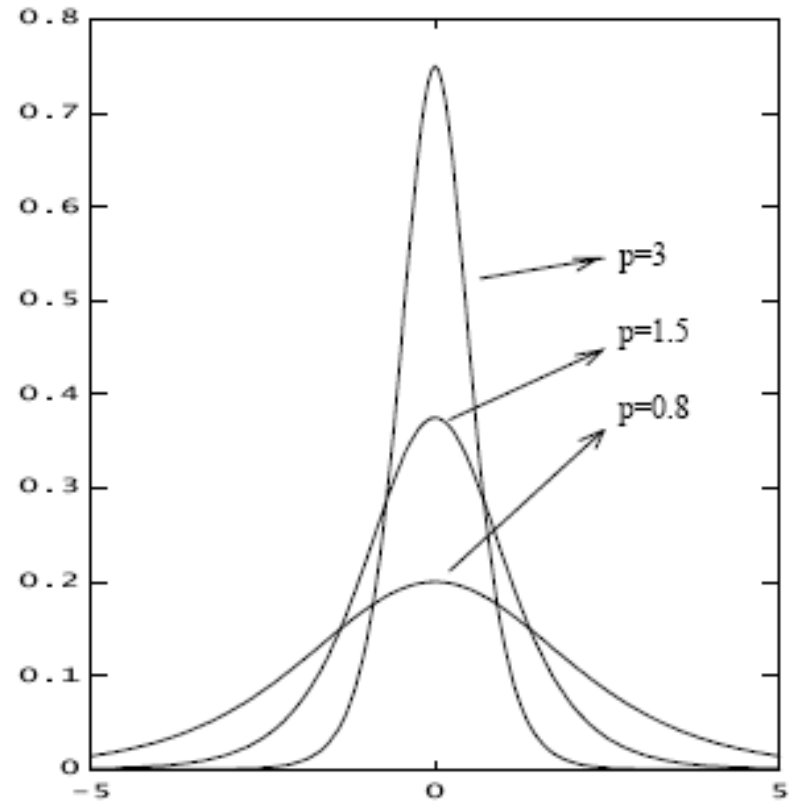
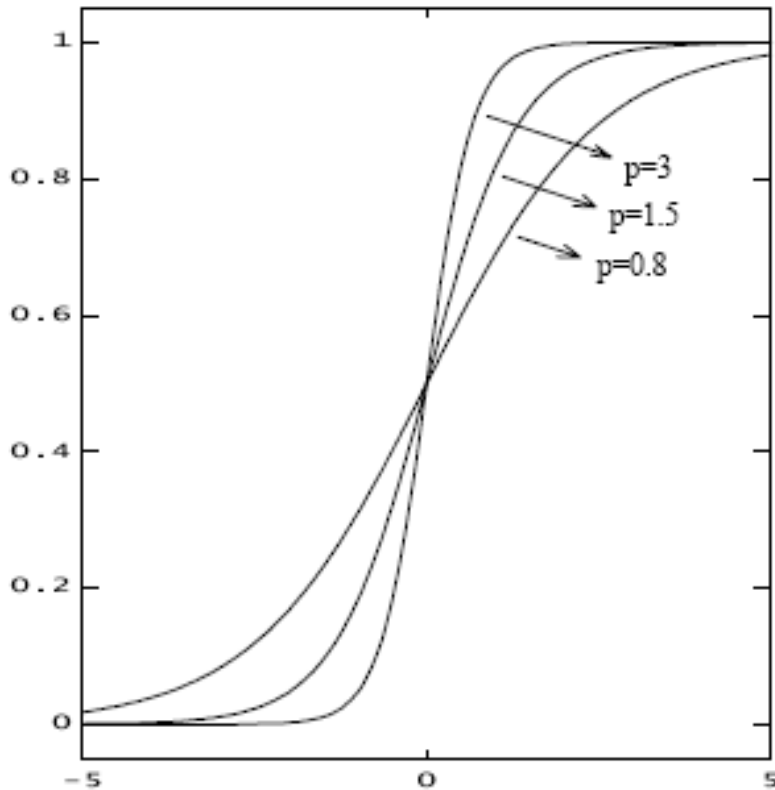
$$\varphi(v) = \begin{cases} +1 & \text{se } v \geq 0 \\ 0 & \text{se } v = 0 \\ -1 & \text{se } v < 0 \end{cases}$$

## Tipos de função de ativação

(c) Função de ativação logística (sigmóid).

$$\varphi(v) = \frac{e^{pv}}{e^{pv} + 1} = \frac{1}{1 + e^{-pv}}$$

$$\frac{\partial \varphi(v)}{\partial v} = pv(1-v) > 0$$

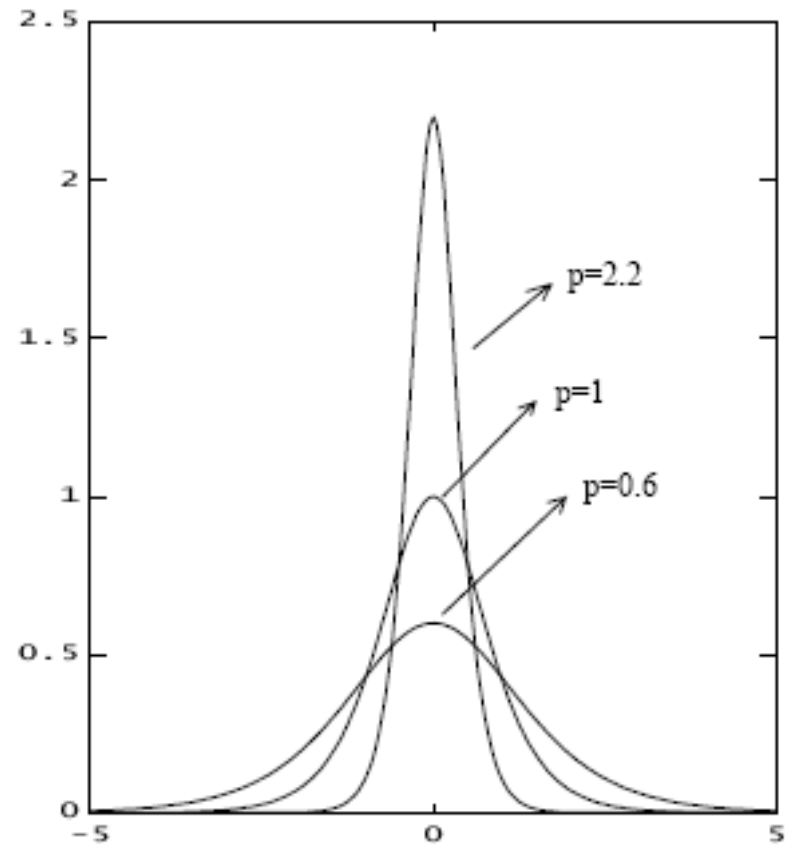
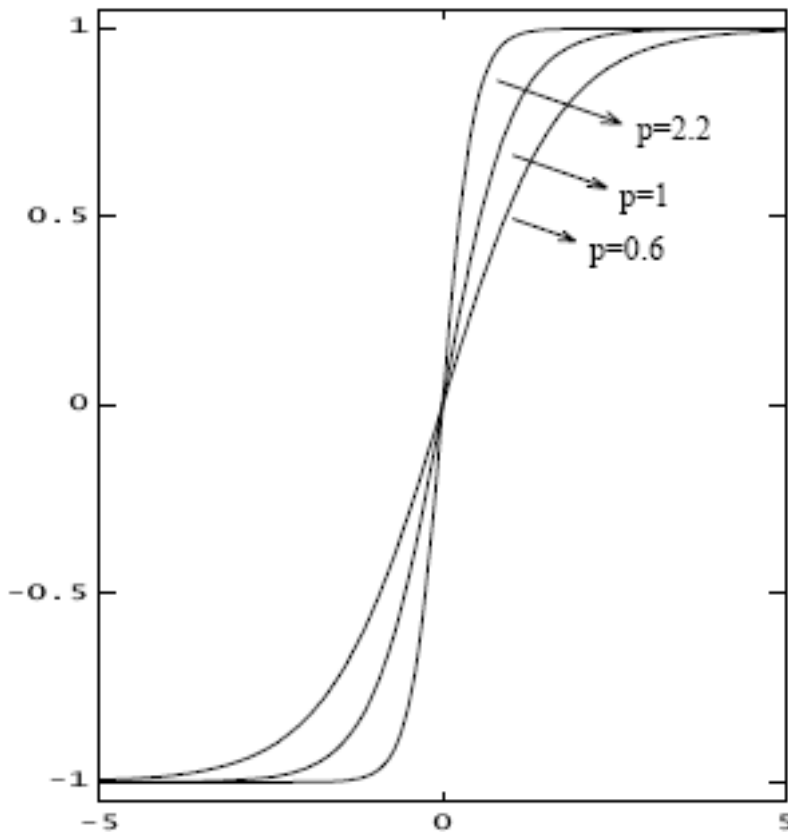


## Tipos de função de ativação

(d) Função de ativação tangente hiperbólica.

$$\varphi(v) = \tanh(pv) = \frac{e^{pv} - e^{-pv}}{e^{pv} + e^{-pv}}$$

$$\frac{\partial \varphi(v)}{\partial v} = p(1 - v^2) > 0$$



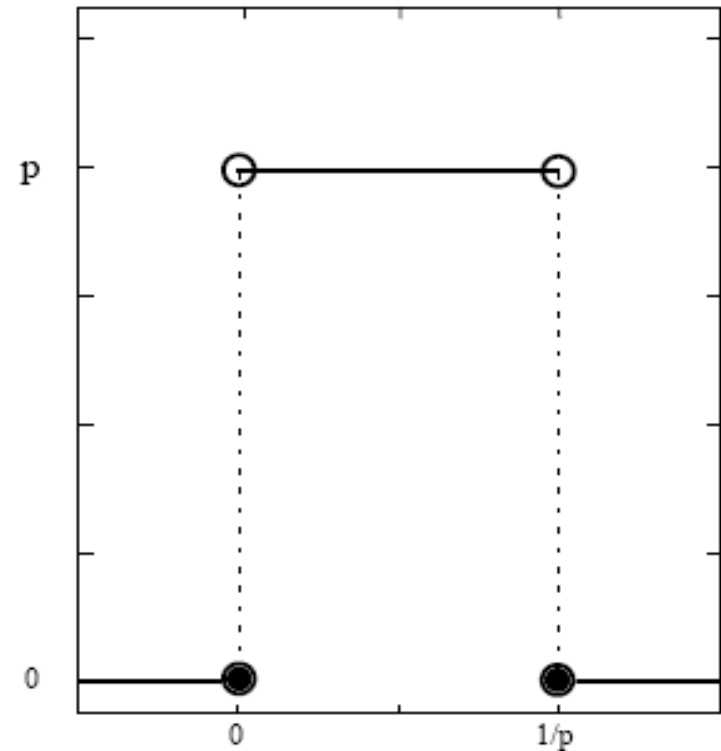
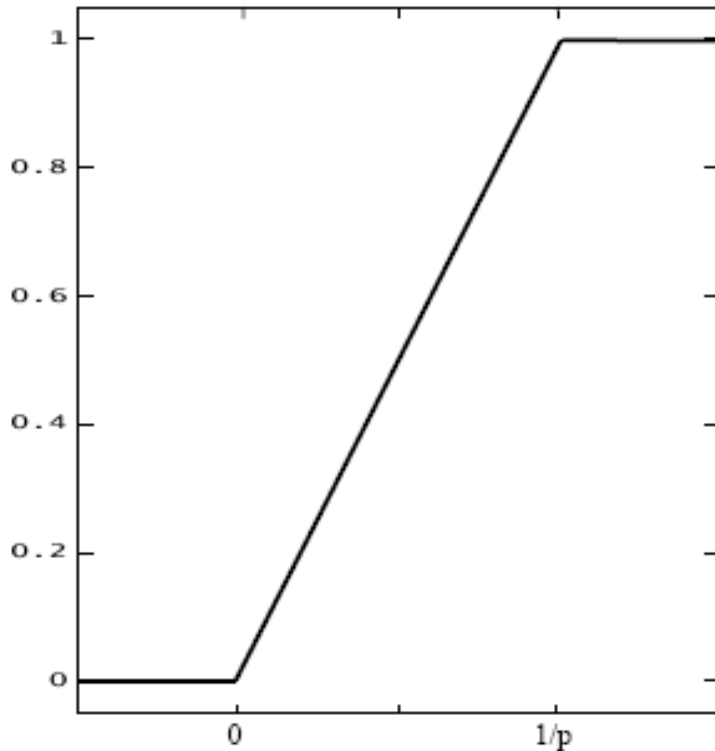


## Tipos de função de ativação

(e) Função de ativação semi-linear.

$$\varphi(v) = \begin{cases} +1 & \text{se } pv \geq 1 \\ pv & \text{se } 0 < pv < 1 \\ 0 & \text{se } pv \leq 0 \end{cases}$$

$$\frac{\partial \varphi(v)}{\partial v} = p$$

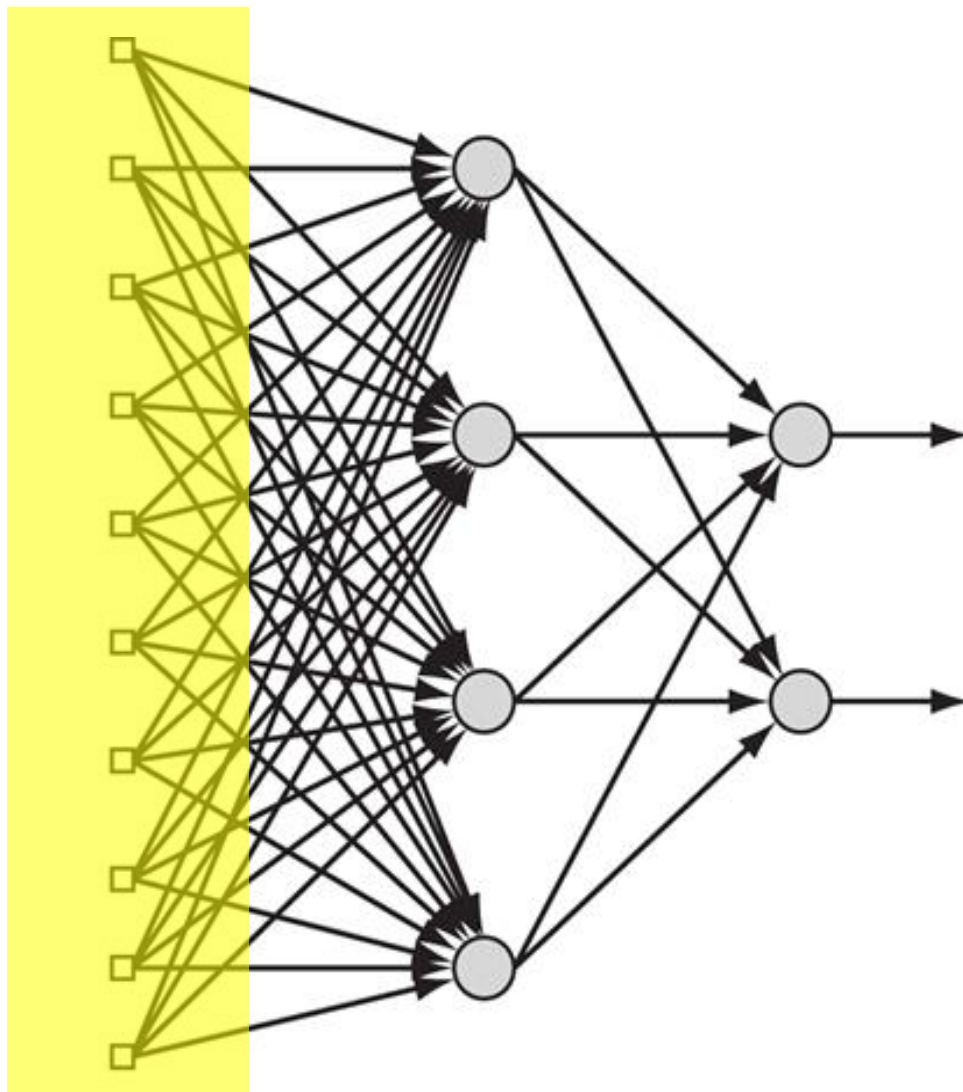


## Redes Neurais Artificiais

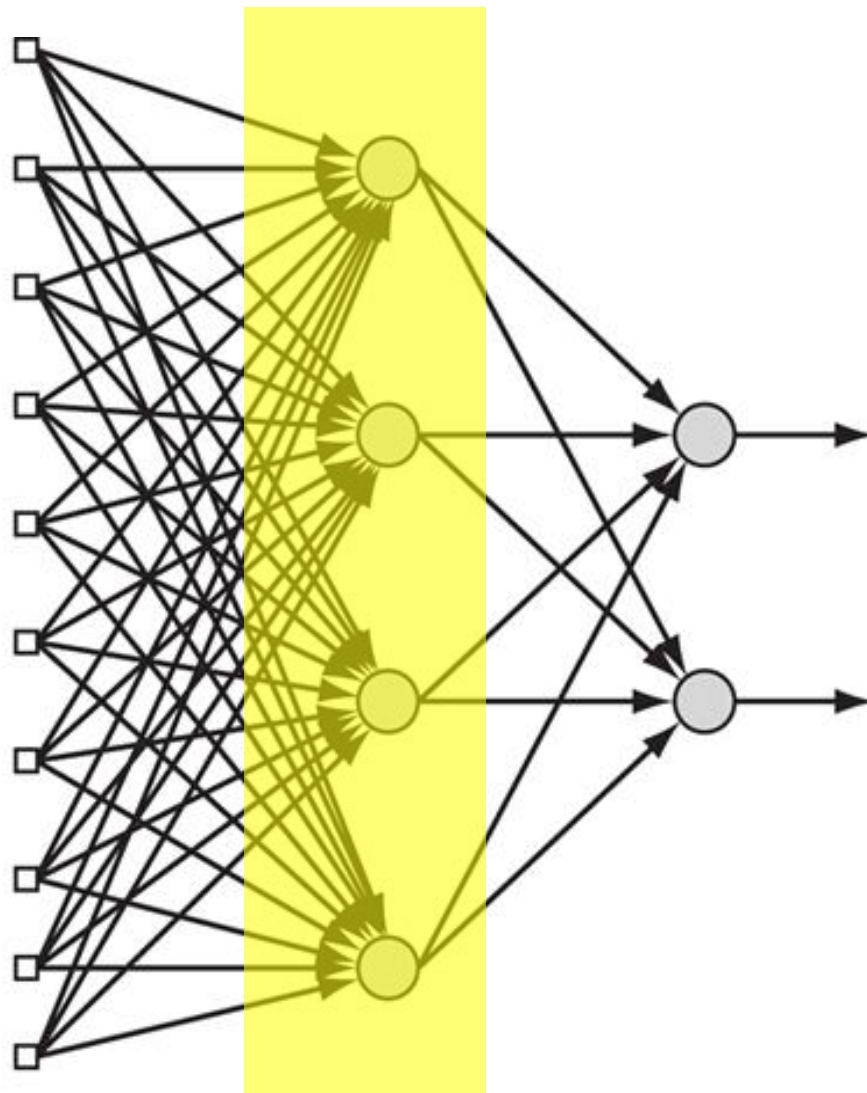
## Arquiteturas

- A arquitetura de uma rede neural define a forma como os seus diversos neurônios estão arranjados, ou dispostos, uns em relação aos outros;
- Uma rede neural pode ser dividida em três partes básicas:
  - a) Camada de entrada: responsável pelo recebimento de informações (dados) do meio externo;
  - b) Camadas escondidas: compostas por neurônios que possuem a responsabilidade de extrair as características associadas ao processo ou sistema a ser inferido;
  - c) Camada de saída: também constituída de neurônios sendo responsável pela produção e apresentação dos resultados finais da rede;

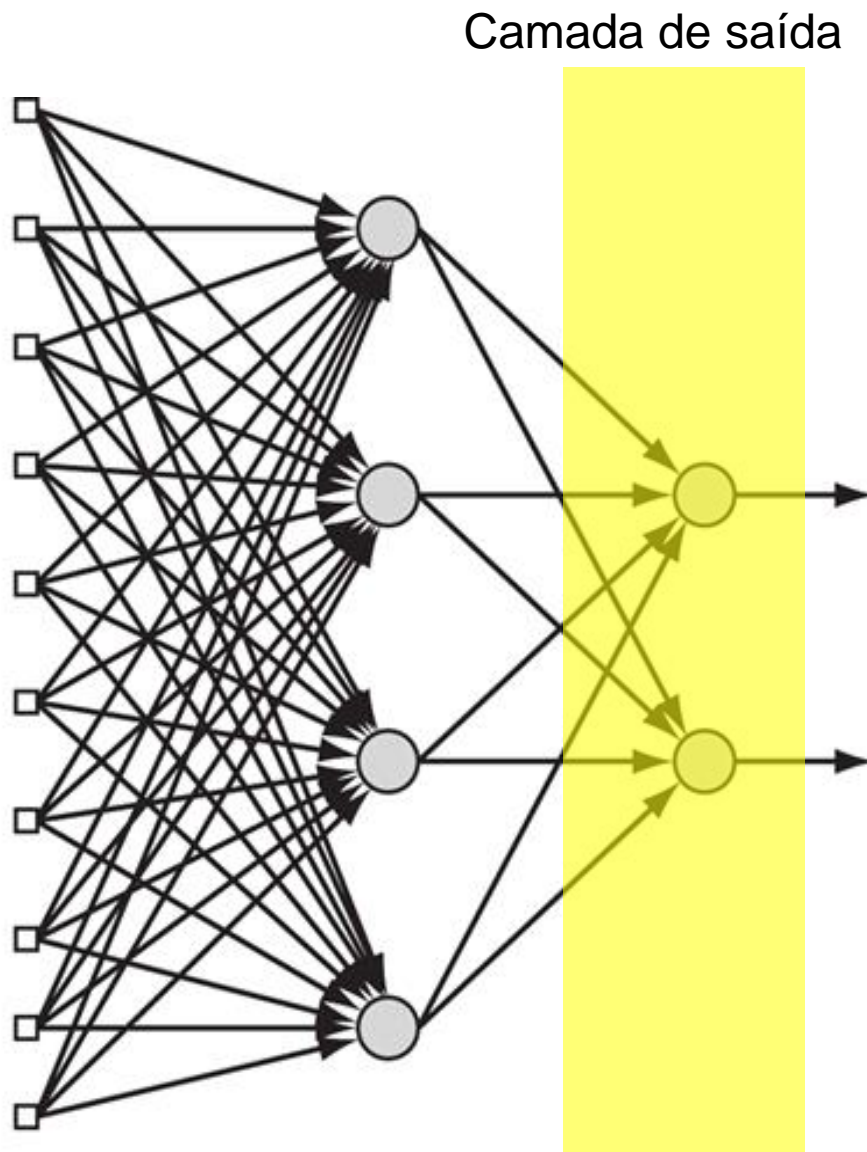
Camada de entrada



Camada escondida ou intermediária

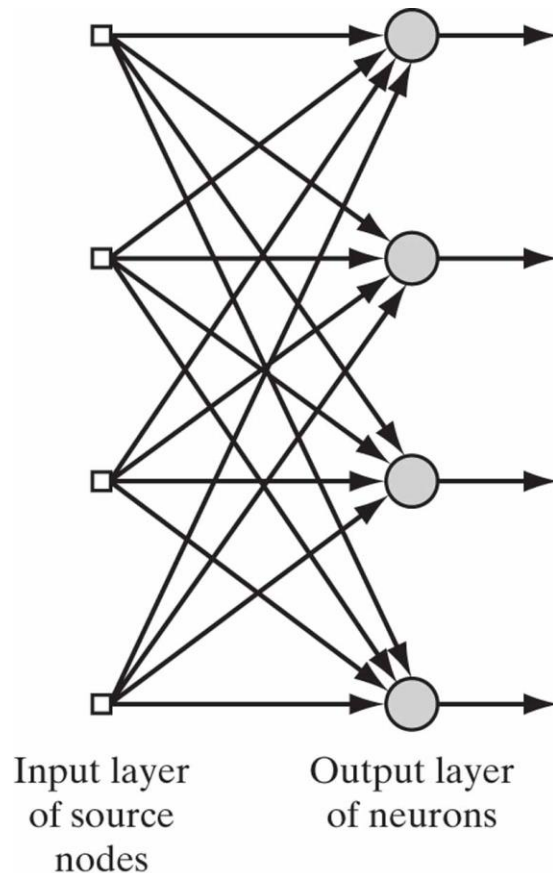


# Arquiteturas de redes neurais





- Redes alimentadas adiante (*feedforward*) com camada única

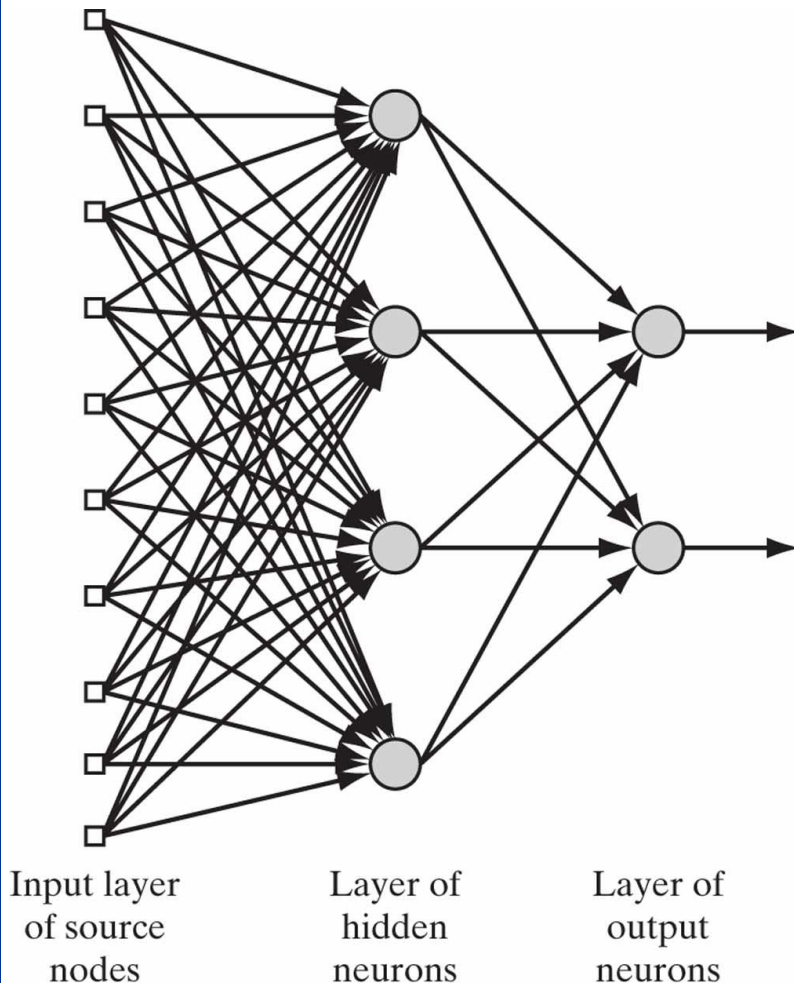


- Uma camada de entrada de nós de fonte que se projeta sobre uma camada de saída de neurônios.



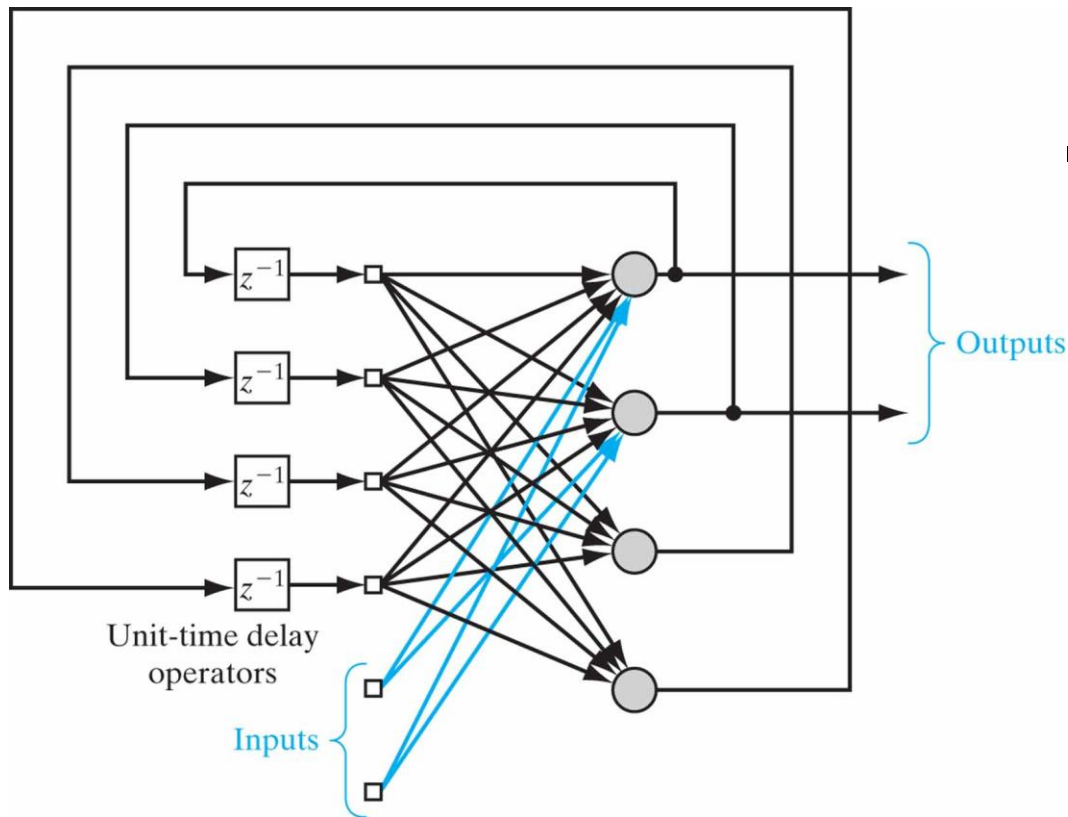
## Arquiteturas básicas, cont.

### ➤ Redes alimentadas adiante com múltiplas camadas



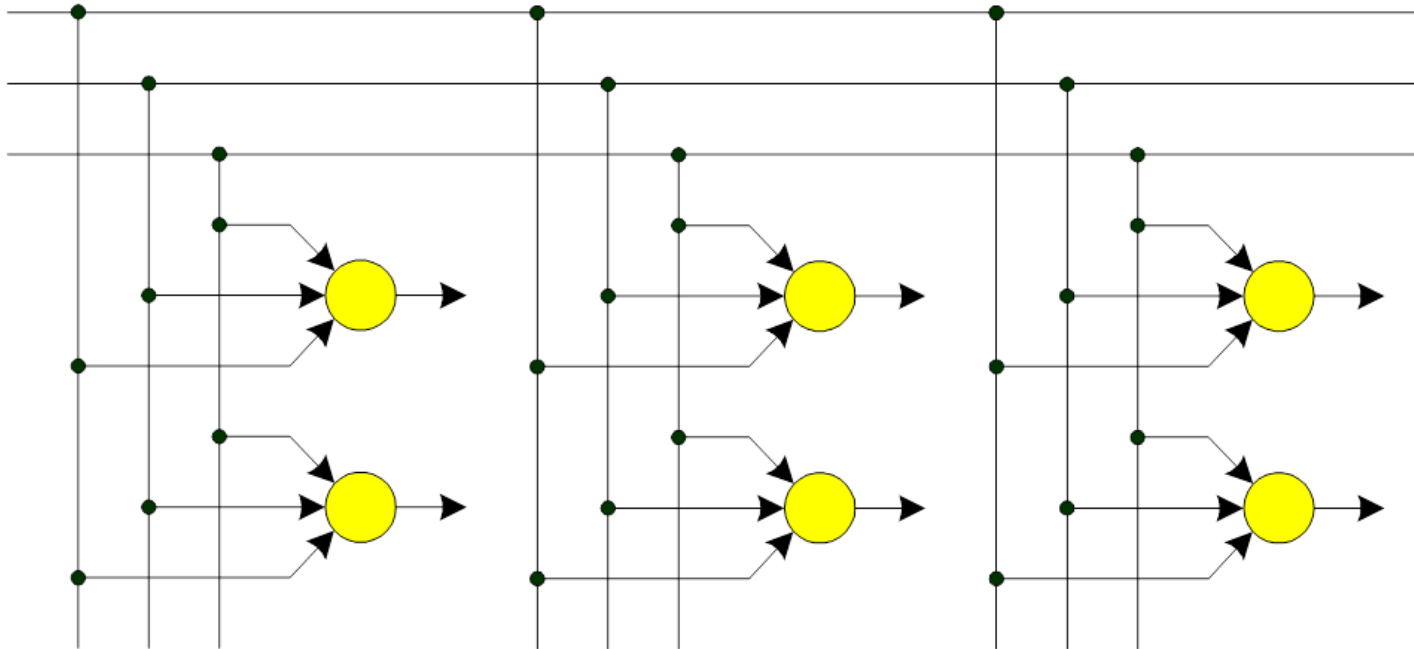
- Os nós de fonte da camada de entrada da rede fornecem os respectivos elementos do padrão de ativação, que constituem os sinais de entrada aplicados aos neurônios na segunda camada.

### ➤ Redes recorrentes ou realimentadas



- Difere da rede neural alimentada adiante por ter pelo menos um laço de realimentação.

## ➤ Redes em estrutura reticulada



- Tem como principal característica a disposição espacial dos neurônios. Muito aplicada em problemas de agrupamento (Kohonen)

## Redes Neurais Artificiais

### Processo de treinamento e aspectos de aprendizado

## Processo de aprendizado

- Um dos destaques mais relevantes das redes neurais artificiais é a capacidade de aprender a partir de amostras (exemplos) que exprimem o comportamento do sistema;
- Isso é feito através de passos ordenados a fim de sintonizar os pesos sinápticos e limiares de seus neurônios, etapa conhecida como **processo de treinamento**;
- O aprendizado ocorre quando a rede neural atinge uma solução generalizada para o sistema.

- Tipos de exemplos:
  - Rotulados: cada exemplo que representa um sinal de entrada é associado a uma resposta desejada.
  - Não-rotulados: ocorrências diferentes dos próprios sinais de entrada.
- Um conjunto de pares de entrada-saída é referido como um conjunto de dados de treinamento ou amostra de treinamento.
- O conjunto total de amostras disponíveis sobre o comportamento do sistema é dividido em dois subconjuntos: subconjunto de treinamento (60 a 90% do conjunto total) e subconjunto de teste (10 a 40%);
- Cada apresentação completa dos dados de treinamento é denominado **época**.

## Regras para o aprendizado

- Sinais de entrada similares provenientes de classes de eventos ou objetos similares devem produzir representações similares dentro da rede e devem ser classificados como pertencentes à mesma categoria;
- Itens que devem ser classificados em categorias separadas devem provocar representações bastante distintas dentro da rede;
- Se uma característica é importante, então deve haver um grande número de neurônios envolvidos na sua representação;
- Informações conhecidas a priori e invariância devem ser embutidas no projeto da rede.



## Similaridade entre entradas

$$X_i = [x_{i1}, x_{i2}, \dots, x_{iN}]^T$$

$$X_j = [x_{j1}, x_{j2}, \dots, x_{jN}]^T$$

➤ Distância euclidiana

$$d(X_i, X_j) = \|X_i - X_j\| = \left[ \sum_{n=1}^N (x_{in} - x_{jn})^2 \right]^{1/2}$$

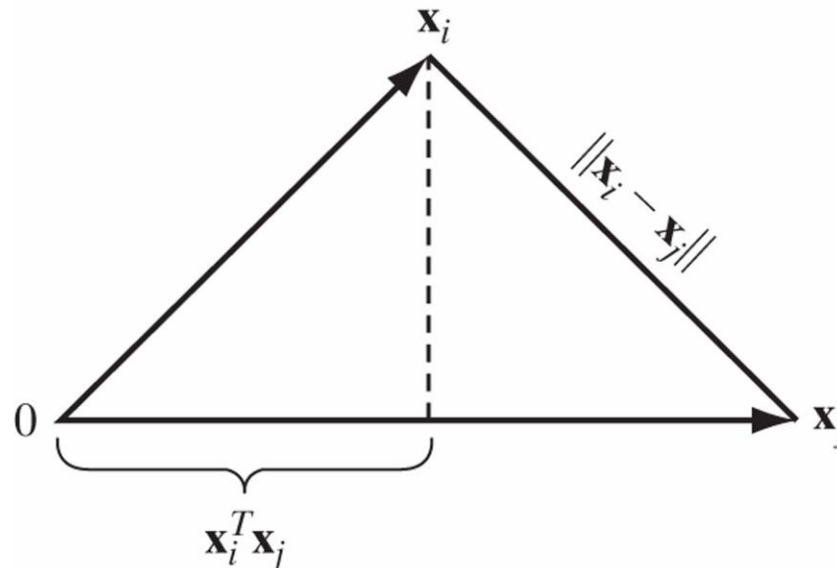
➤ Produto escalar ou produto interno

$$(X_i, X_j) = X_i^T X_j = \sum_{n=1}^N x_{in} x_{jn}$$

**Similaridade:  
inverso da  
distância**

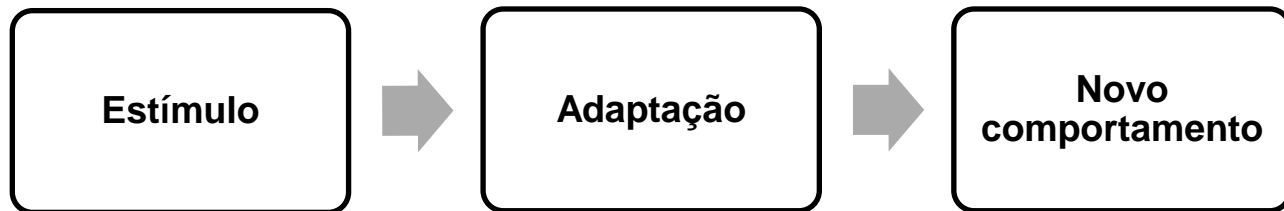
## Similaridade entre entradas

- Relação entre as medidas de similaridade
  - A distância euclidiana  $\|X_i - X_j\|$  entre os vetores  $X_i$  e  $X_j$  está relacionada com a “projeção” do vetor  $X_i$  sobre o vetor  $X_j$ .
  - Quanto mais similares forem  $X_i$  e  $X_j$ , maior será o produto interno  $X_i^T X_j$ .



## Definição de aprendizagem

- É o processo no qual os parâmetros livres de uma rede neural são alterados pela estimulação contínua causada pelo ambiente na qual a rede está inserida;
- O tipo do aprendizado é determinado pela maneira pela qual os parâmetros são alterados



## Aprendizado em redes neurais artificiais

### ✓ Algoritmos

- Correção de erro
- Máquina de Boltzman
- Lei de Hebb
- Competição

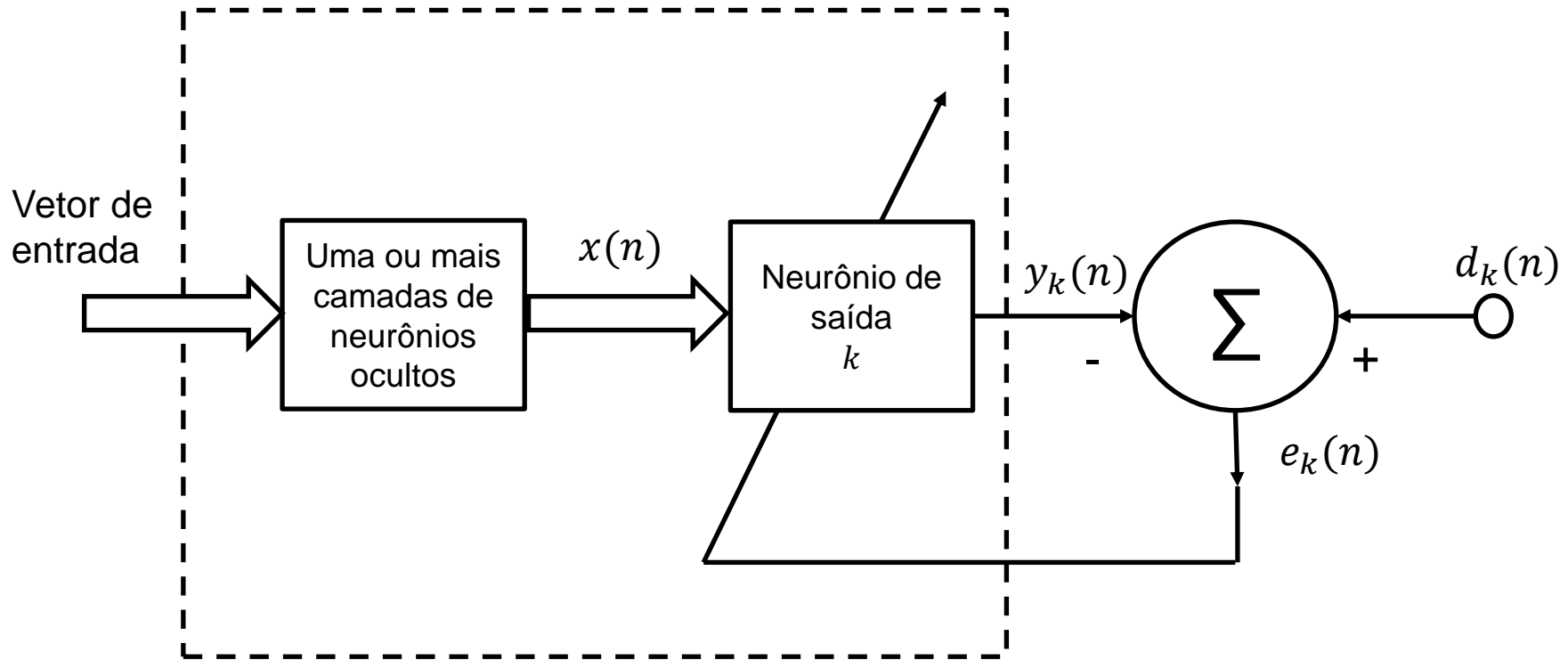
### ✓ Paradigmas

- Supervisionado
- Não supervisionado
- Reforço

Objetivo final do aprendizado: obtenção de um modelo implícito dos conhecimentos adquiridos.

## Aprendizado em redes neurais artificiais, cont.

- **Aprendizado supervisionado**, quando é utilizado um agente externo que indica à rede a resposta desejada para o padrão de entrada;
- **Aprendizado não supervisionado** (auto-organização), quando não existe um agente externo indicando a resposta desejada para os padrões de entrada;
- **Reforço**, quando um crítico externo avalia a resposta fornecida pela rede.



Rede de múltiplas camadas  
alimentadas adiante

- Elementos da rede:
- $d_k(n)$ : resposta desejada para o neurônio  $k$  no instante  $n$ ;
  - $y_k(n)$ : resposta observada e obtida através de um estímulo  $x(n)$  na entrada da rede;
  - $x(n)$  e  $d_k(n)$  constituem um exemplo de par “estímulo-resposta” apresentados à rede no instante  $n$ , retirados de um ambiente ruidoso, cujas distribuições de probabilidade são, em geral, desconhecidas;
  - $e_k(n) = d_k(n) - y_k(n)$ : sinal de erro no instante  $n$ .



## Aprendizagem por correção de erro, cont.

- ✓ **Objetivo:** minimizar um critério funcional baseado em  $e_k(n)$ , de modo que, para  $n$  suficientemente grande,  $y_k(n)$  esteja próximo de  $d_k(n)$  no sentido estatístico.
- Um critério muito adotado é o do erro quadrático médio

$$J = E \left[ \frac{1}{2} \sum_k e_k^2(n) \right]$$

## Aprendizagem por correção de erro, cont.

- A minimização de  $J$  leva ao chamado método do decréscimo do gradiente, o qual necessita do conhecimento das características estatísticas do sistema.
- Uma aproximação deste método utiliza o valor instantâneo da energia do erro.

$$E(n) = \frac{1}{2} e_k^2(n)$$

- e a minimização de  $E(n)$  é realizada em relação aos pesos sinápticos da rede.

### ➤ Regra Delta:

- Regra delta, ou regra Widrow-Hoff, é a minimização da função  $E(n)$ .
- Supondo que  $w_{kj}(n)$  represente o valor do peso sináptico  $w_{kj}$  do neurônio  $k$  excitado por um elemento  $x_j(n)$  do vetor do sinal  $x(n)$  no passo de tempo  $n$ .

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

- Onde  $\eta$  é uma constante positiva que determina a taxa de aprendizagem.

### ✓ Regras:

- Se dois neurônios em ambos os lados de uma sinapse 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 aquela sinapse é seletivamente enfraquecida ou eliminada.

$$\Delta\omega_{kj}(n) = \eta y_k(n) x_j(n)$$

- Os neurônios de saída de uma rede neural competem para se tornar ativos.
- Na aprendizagem competitiva, apenas um neurônio pode estar ativo em um determinado instante.
- Adequado para descobrir características estatisticamente salientes que podem ser utilizadas para classificar um conjunto de padrões de entrada.

- Elementos básicos da aprendizagem competitiva:
  - Um conjunto de neurônios que são todos iguais entre si, exceto por alguns pesos sinápticos distribuídos aleatoriamente;
  - Um limite imposto sobre a “força” de cada neurônio;
  - Um mecanismo que permite que o neurônio compita pelo direito de responder a um dado subconjunto de entradas, de forma que somente um neurônio de saída, ou somente um neurônio por grupo, esteja ativo.

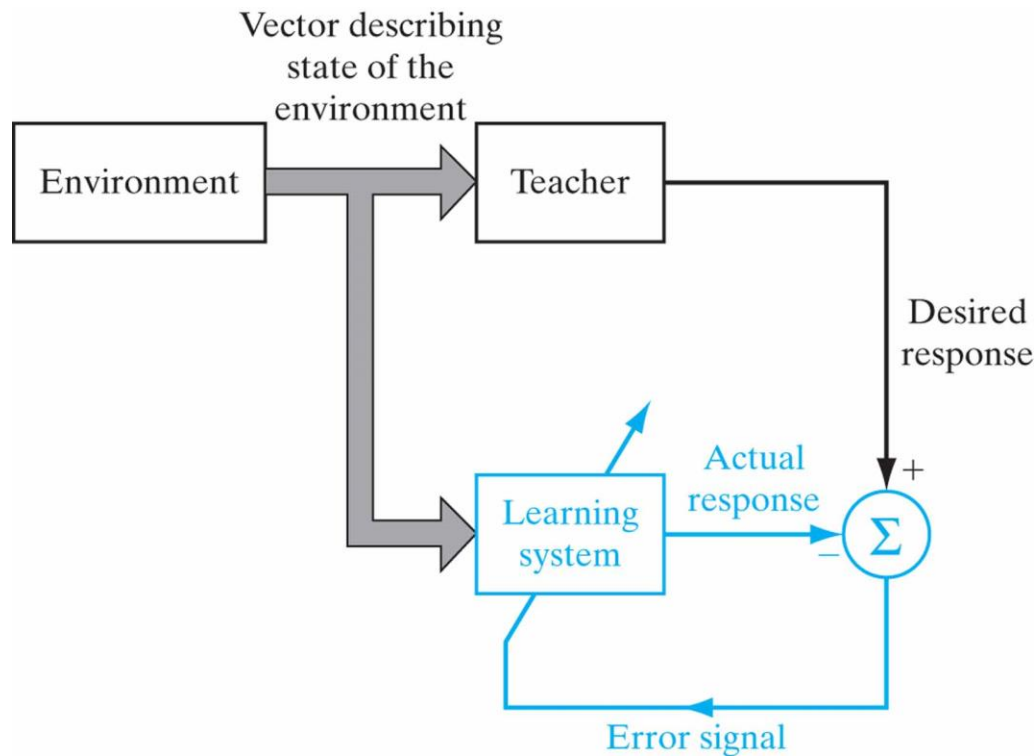
- Para um neurônio  $k$  ser o neurônio vencedor, seu campo local induzido  $u_k$  para um padrão especificado  $X$  deve ser o maior entre todos os neurônios da rede.
- A saída  $y_k$  do neurônio vencedor é colocada em 1 e a saída dos demais neurônios são colocadas em 0.

$$\Delta w_{kj} = \begin{cases} \eta(x_j - w_{kj}) & \text{se vencedor} \\ 0 & \text{se perdedor} \end{cases}$$



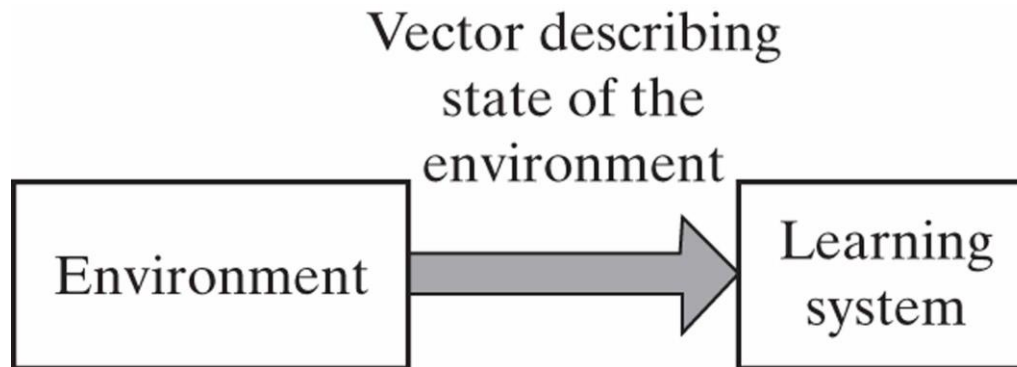
## Aprendizagem com professor

- ✓ Conhecida como aprendizagem supervisionada.



- Ambiente desconhecido pela rede neural.
- O conhecimento é apresentado por um conjunto de exemplos entrada-saída.
- A partir de um conhecimento prévio, o professor é capaz de fornecer uma resposta desejada para o vetor de treinamento;
- Os parâmetros da rede são ajustados sob a influência combinada do vetor de treinamento e do sinal de erro.
- Este ajuste é realizado passo a passo, iterativamente com o objetivo de fazer a rede emular o professor.

- ✓ Conhecida como aprendizagem auto-organizada.



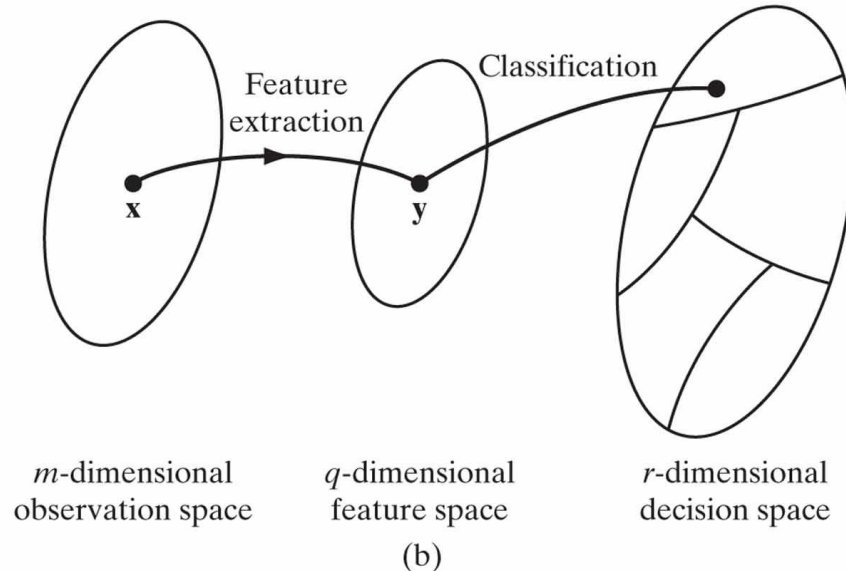
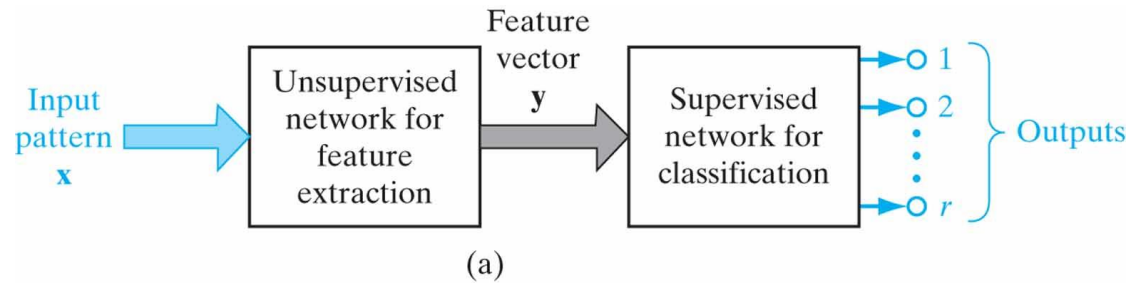
- Não há professor para supervisionar o processo de aprendizagem.
- Não há exemplos rotulados da função a ser aprendida pela rede.
- São dadas condições para realizar um **medida independente da tarefa** da qualidade da representação que a rede deve aprender.
- Pode-se utilizar a regra de aprendizagem competitiva.

- ✓ Associação de padrões
  - A rede neural armazena um conjunto de padrões (vetores), que são apresentados repetidamente à rede.
  - Posteriormente, apresenta-se à rede uma descrição parcial ou distorcida (ruidosa) de um padrão original armazenado pela rede e a tarefa é recuperar aquele padrão.



- ✓ Reconhecimento de padrões
  - A rede neural passa por um processo de treinamento onde são apresentados repetidamente um conjunto de padrões de entrada junto com a categoria ao qual cada padrão pertence.
  - Posteriormente é apresentado à rede um padrão ainda não visto antes mas que pertence à mesma população de padrões utilizados no treinamento e a mesma é capaz de identificar a classe daquele padrão.

### ✓ Reconhecimento de padrões

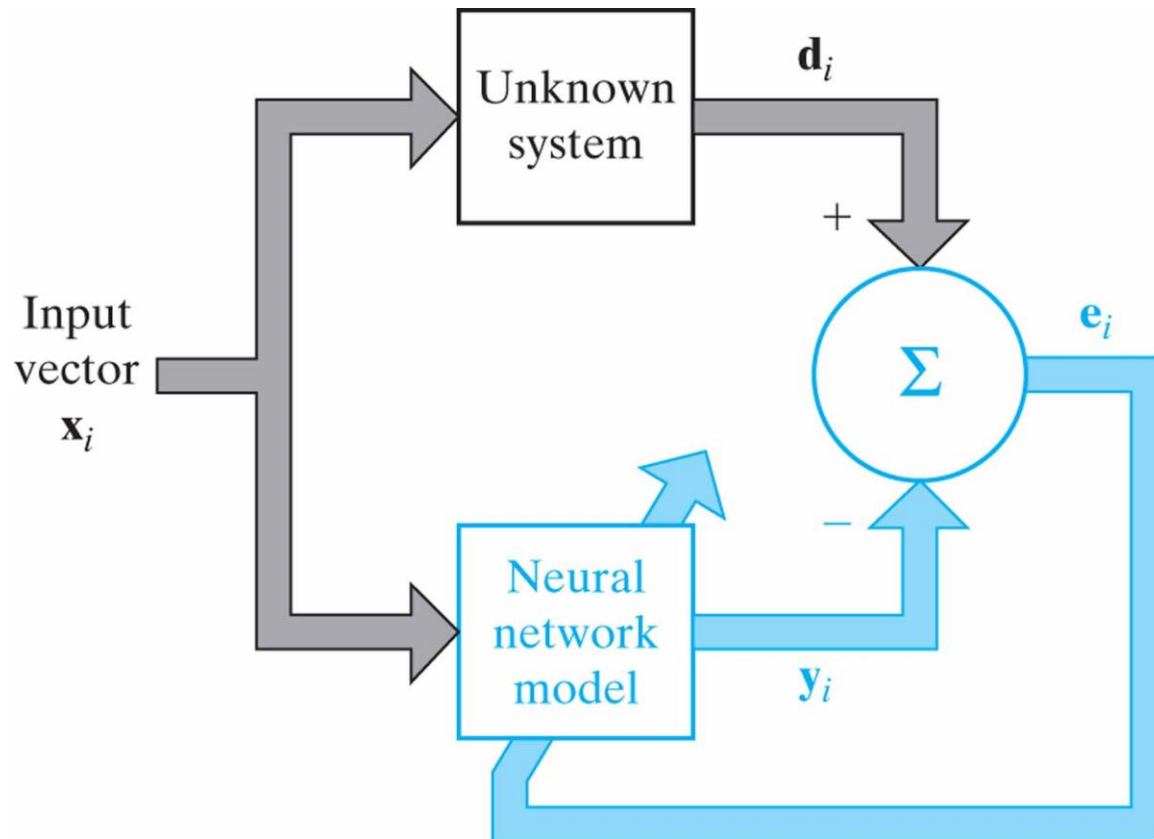




- ✓ Aproximação de funções
  - A rede neural pode ser utilizada para a aproximação de funções desconhecidas.
  - Utilização de uma aprendizagem supervisionada.
  - A habilidade de aproximar o mapeamento entrada-saída desconhecido pode ser aplicado para a identificação de sistemas desconhecidos ou para obter o sistema inverso.

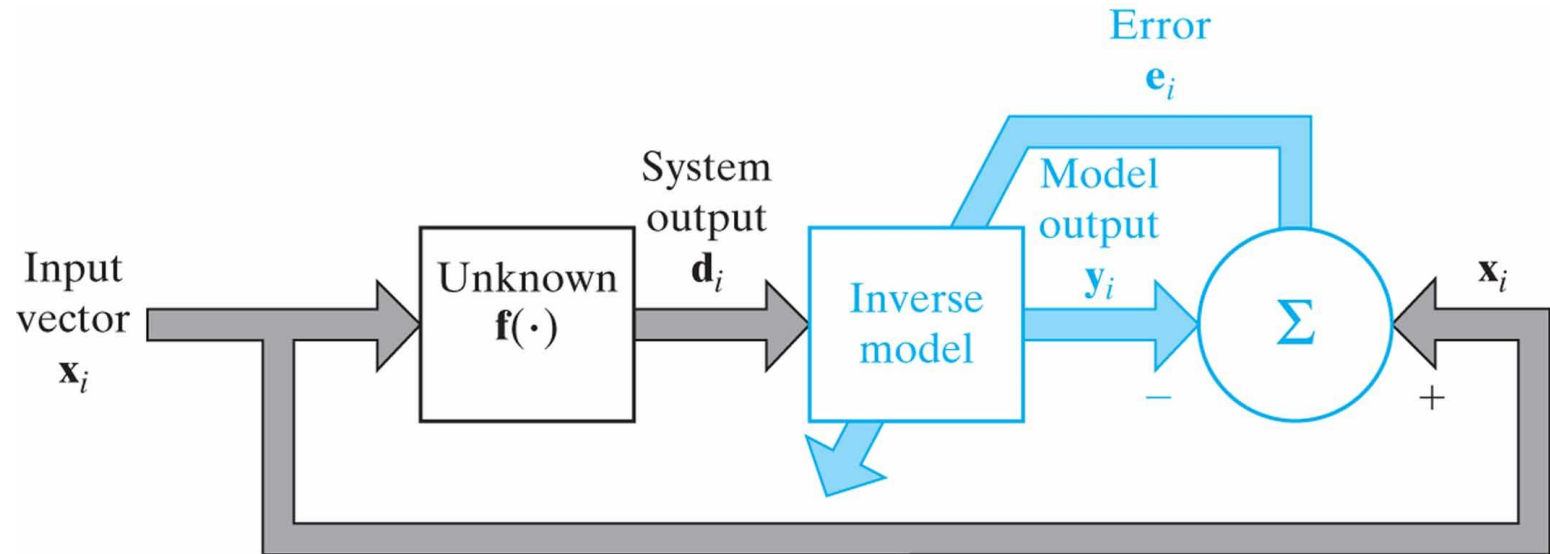
## Tarefas de aprendizagem, cont.

- ✓ Aproximação de funções: identificação de sistemas



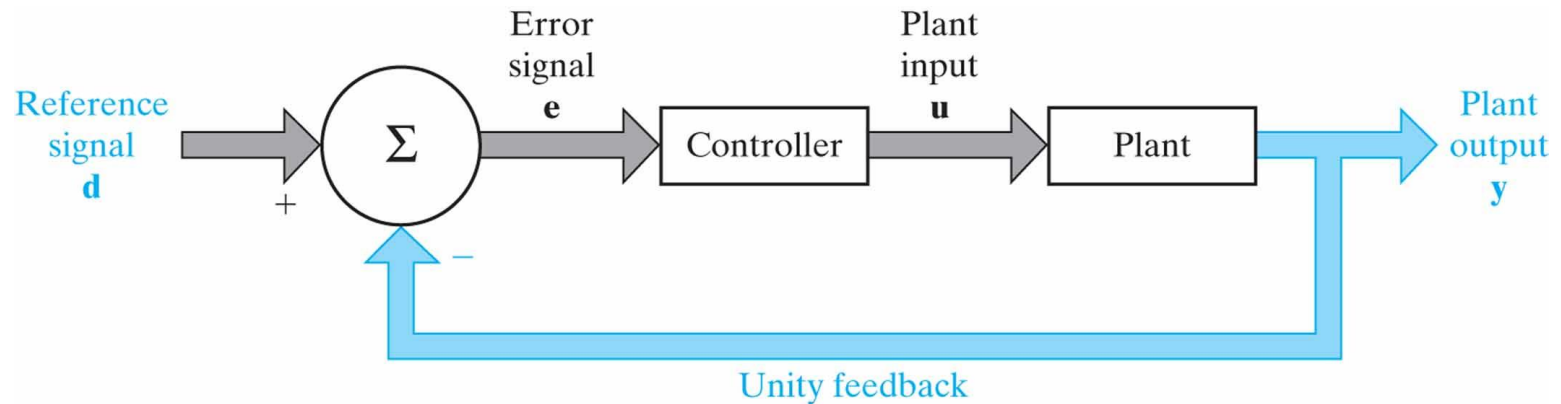
## Tarefas de aprendizagem, cont.

- ✓ Aproximação de funções: obtenção de sistemas inversos



### ✓ Controle

- O controle de um processo ou parte crítica de um sistema é uma tarefa de aprendizagem que pode ser feita por uma rede neural.



### ✓ Filtragem

- A rede neural pode ser utilizada para filtragem, ou seja, extrair informações sobre uma determinada grandeza de interesse a partir de um conjunto de dados ruidosos.
- O filtro pode realizar: filtragem, suavização ou previsão.



**Inatel**  
Instituto Nacional de Telecomunicações

Edielson Prevato Frigieri

edielson@inatel.br

