

C210 – Inteligência Computacional

Introdução às Redes Neurais Artificiais

Prof. Me. Marcelo Vinícius Cysneiros Aragão
marcelovca90@inatel.br

1º semestre / 2022

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 neurocomputador ([Perceptron 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 (em "[Perceptrons](#)") 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 as 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;

1992

- Riedmiller & Braun propuseram o método Resilient Propagation (Rprop), permitindo a atuação individual em pesos sinápticos e agilizando a convergência. Igel & Hüsken nomearam e propuseram variações.

1994

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

1998

- Apresentação das máquinas de vetores suporte (*support vector machines - SVM*), modelos que podem ser utilizados em classificação de padrões, regressão e agrupamento [Vapnik];

Resumo histórico, cont.

2002

- Ken Stanley (Universidade do Texas) propôs o NEAT ([NeuroEvolution of Augmenting Topologies](#)), um algoritmo genético (GA) para a geração de redes neurais artificiais evolucionárias (“complexificação” de topologias).

2003

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

2012

- Krizhevsky et. al. usaram redes convolutivas praticamente reduzir a metade a taxa de erro no reconhecimento de objetos, precipitando a rápida adoção do aprendizado profundo pela comunidade de visão computacional.

2014

- Os pesquisadores do Facebook publicam seu trabalho [DeepFace](#), um sistema que usa redes neurais capaz de identificar rostos com precisão de 97,35%.

2016

- O programa [AlphaGo](#) do Google se torna o primeiro a vencer um jogador humano profissional balanceado usando uma combinação de técnicas de aprendizado de máquina e busca de árvores.

- **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.

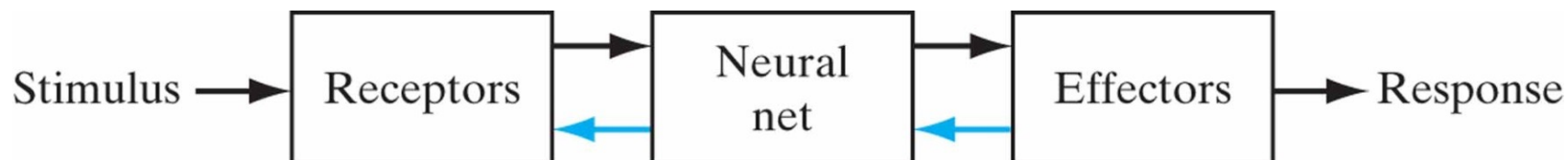
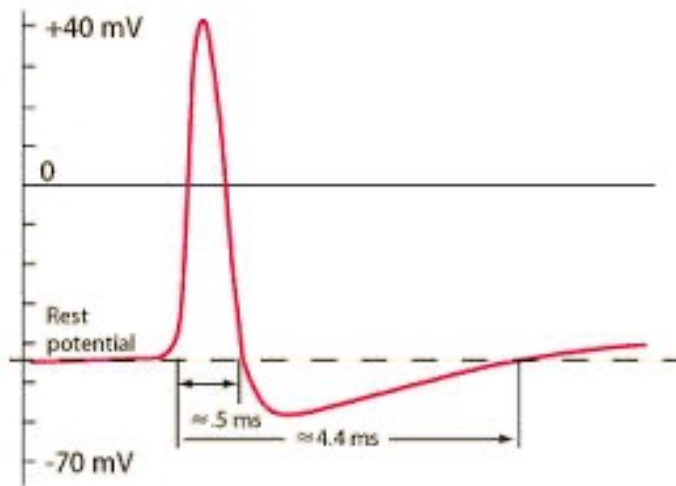


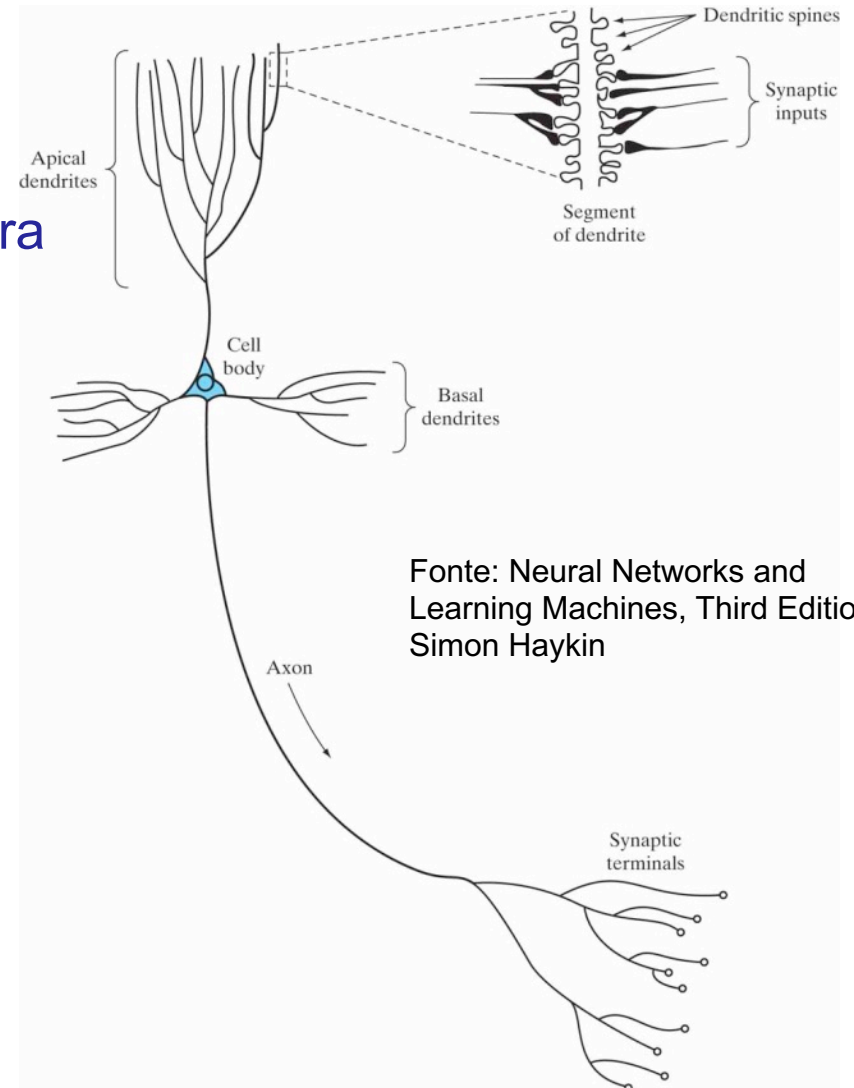
Figura 1 Representação em diagrama em blocos do sistema nervoso.

Neurônio biológico

- 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

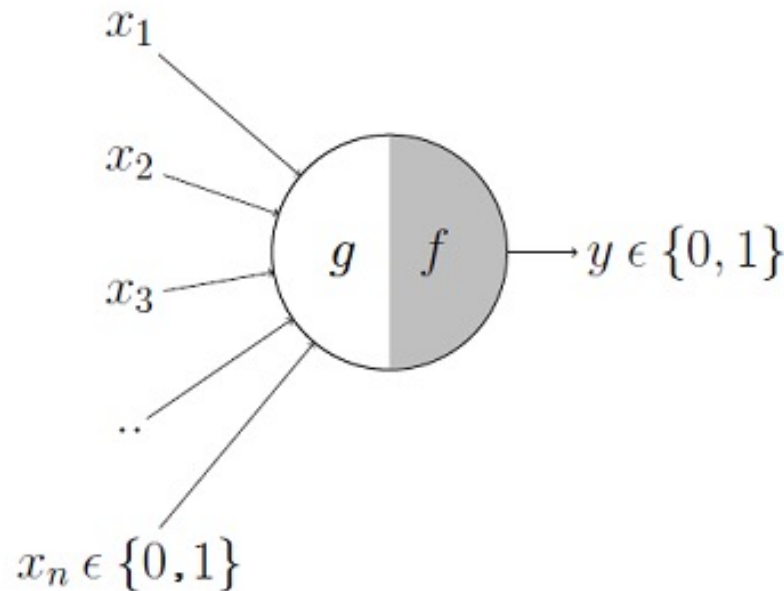
A célula piramidal.

Redes Neurais Artificiais

Arquiteturas

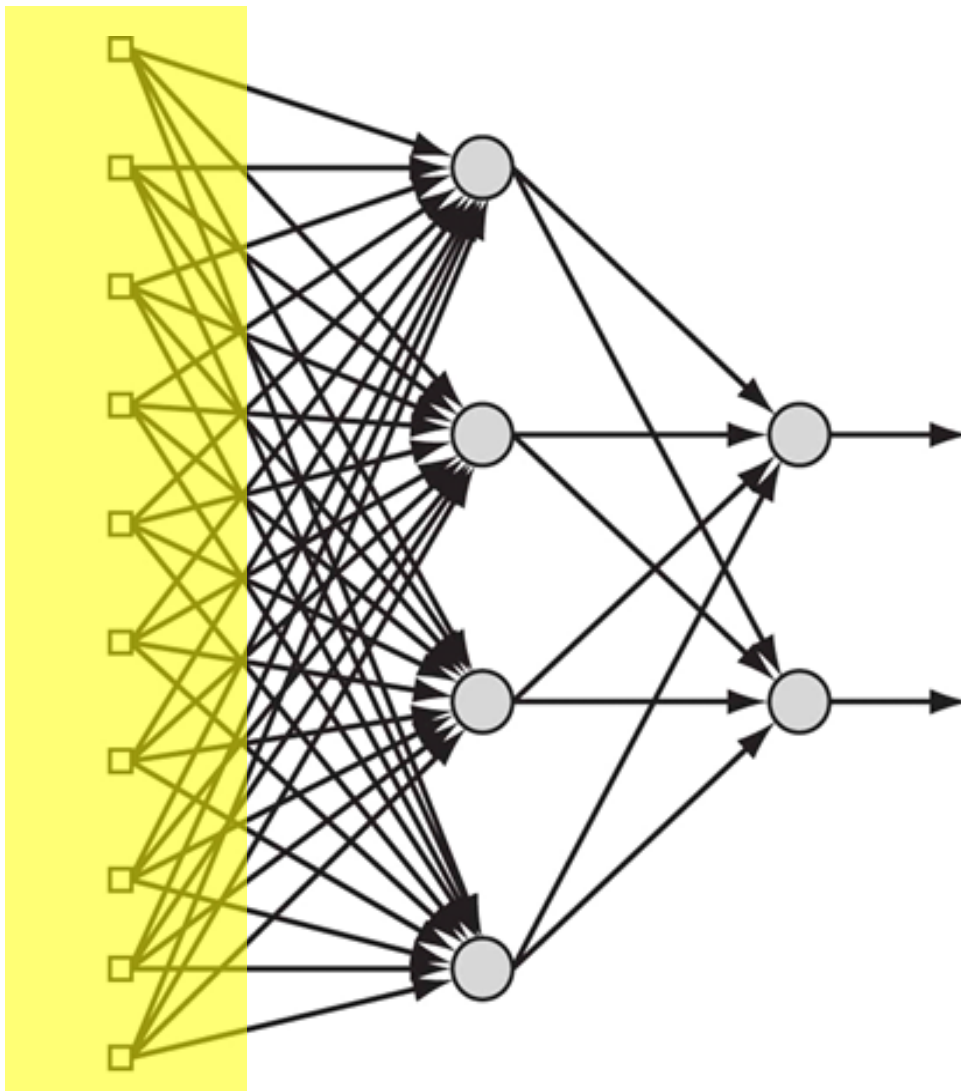


- A arquitetura de uma rede neural define a forma como os seus diversos neurônios estão arranjados, ou dispostos, uns em relação aos outros;
- O primeiro modelo de “rede neural” proposto (neurônio artificial de McCulloch & Pitts), como o próprio nome diz, era composto de um único neurônio.



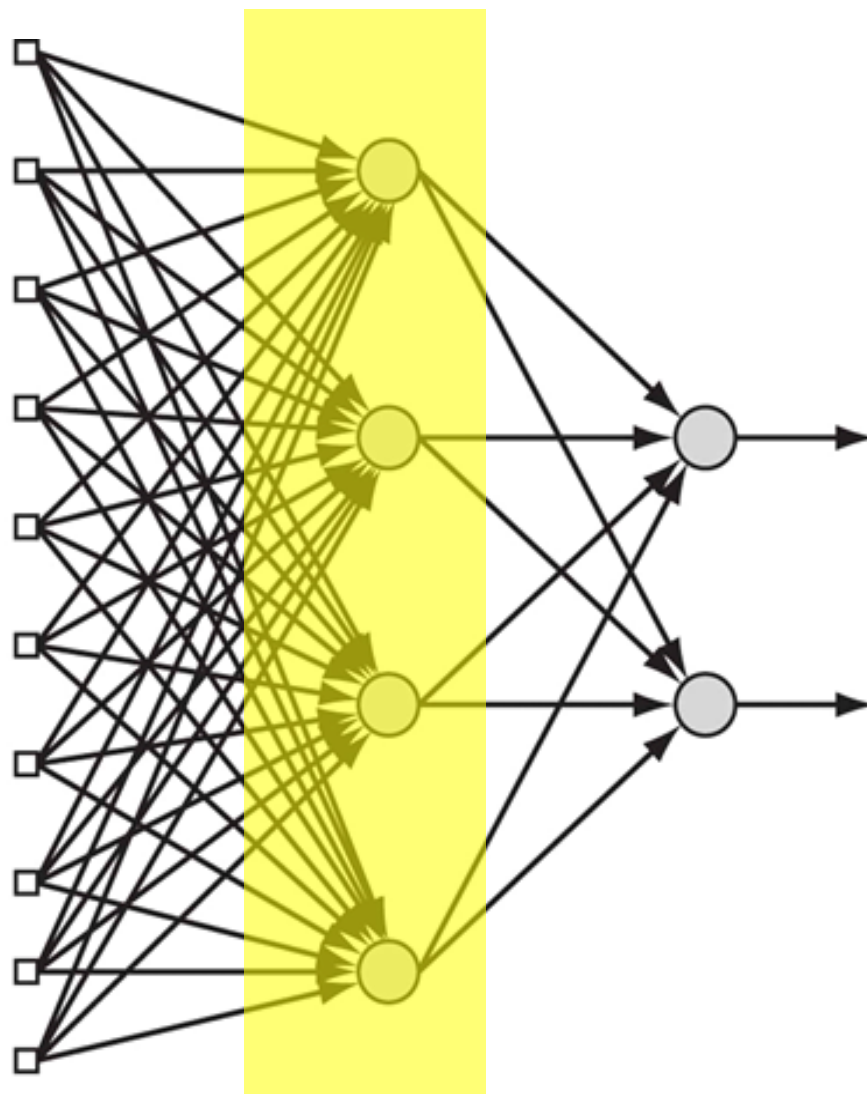
- Uma rede neural de múltiplas camadas pode ser dividida em três partes básicas:
 - a) Camada de entrada: responsável pelo recebimento de informações (dados) do meio externo;
 - b) Camadas escondidas: compostas por neurônios que possuem a responsabilidade de extrair as características associadas ao processo 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

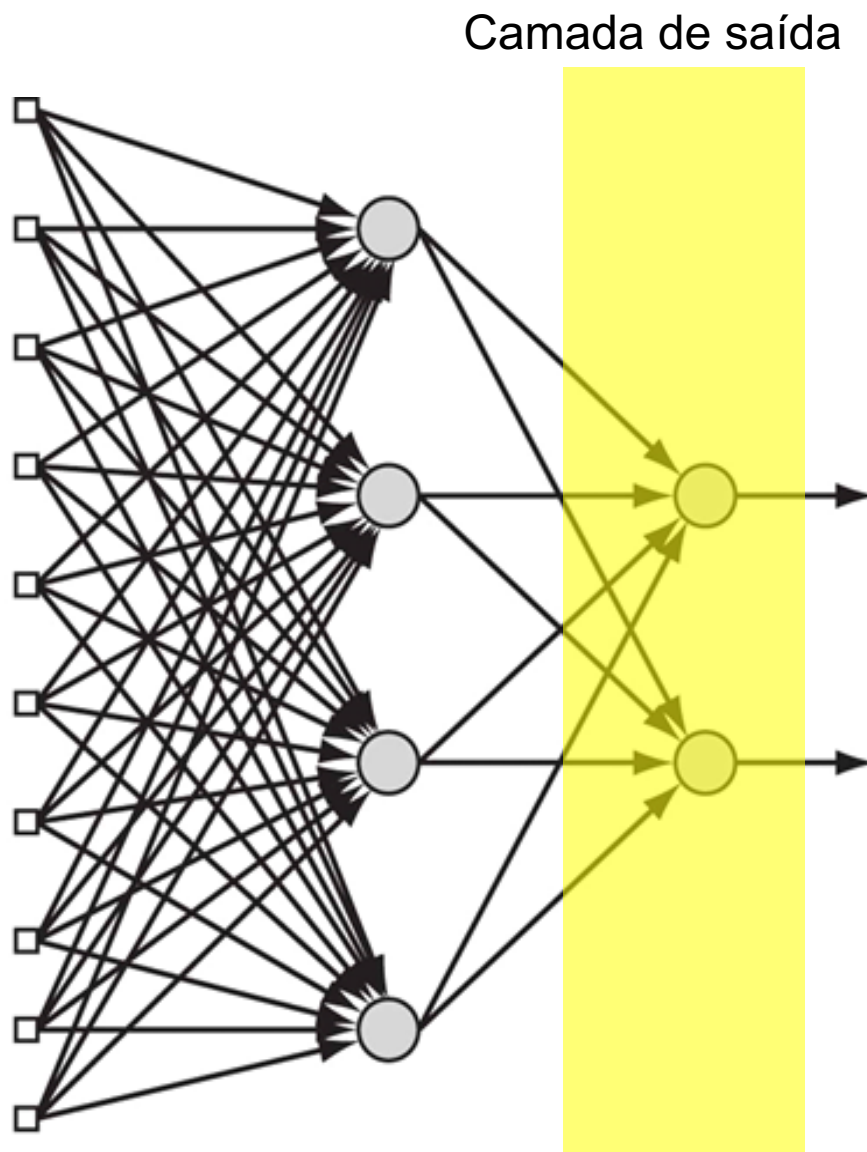


Arquiteturas de redes neurais

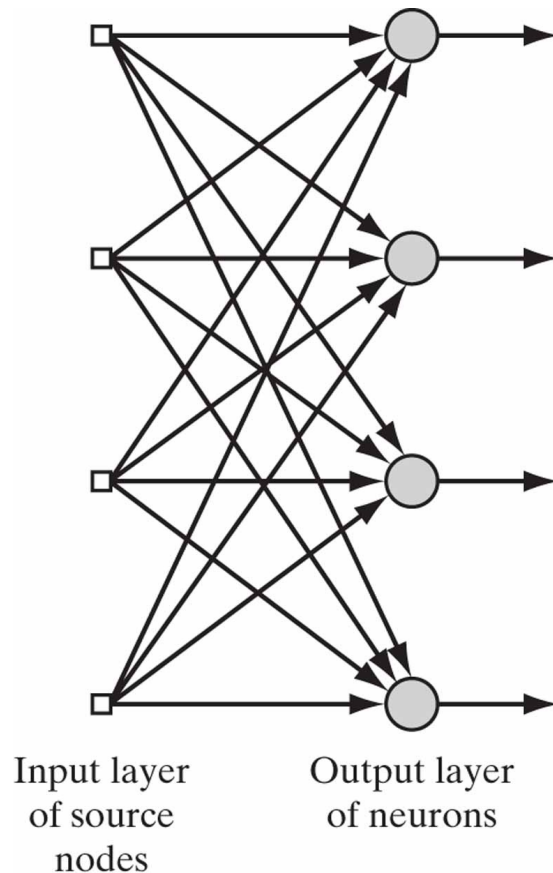
Camada escondida ou intermediária



Arquiteturas de redes neurais

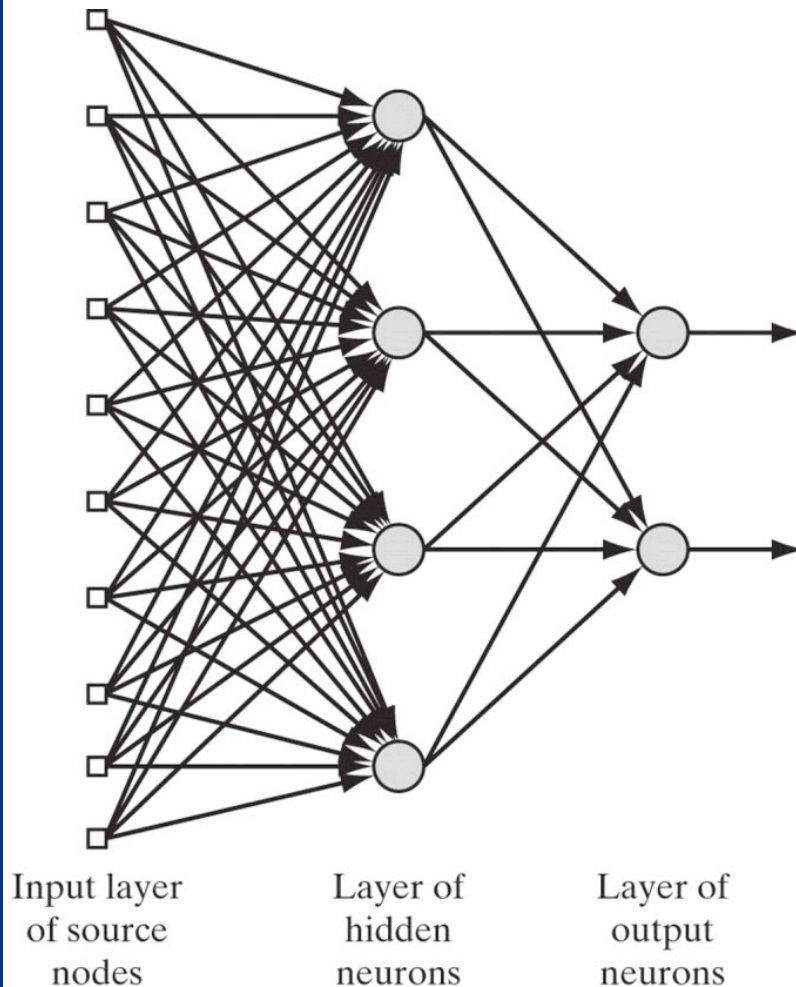


- Redes alimentadas adiante (*feedforward*) com camada única



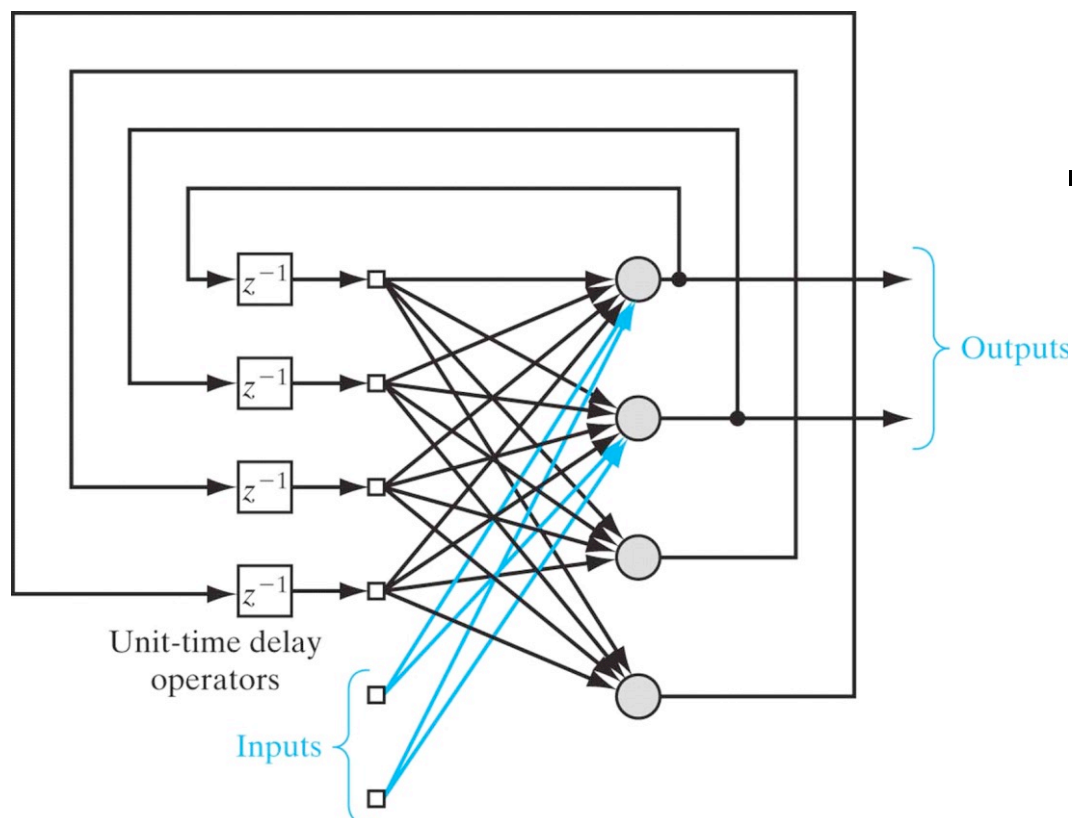
- Uma camada de entrada de nós-fonte que se projeta sobre uma camada de saída de neurônios.

➤ Redes alimentadas adiante com múltiplas camadas



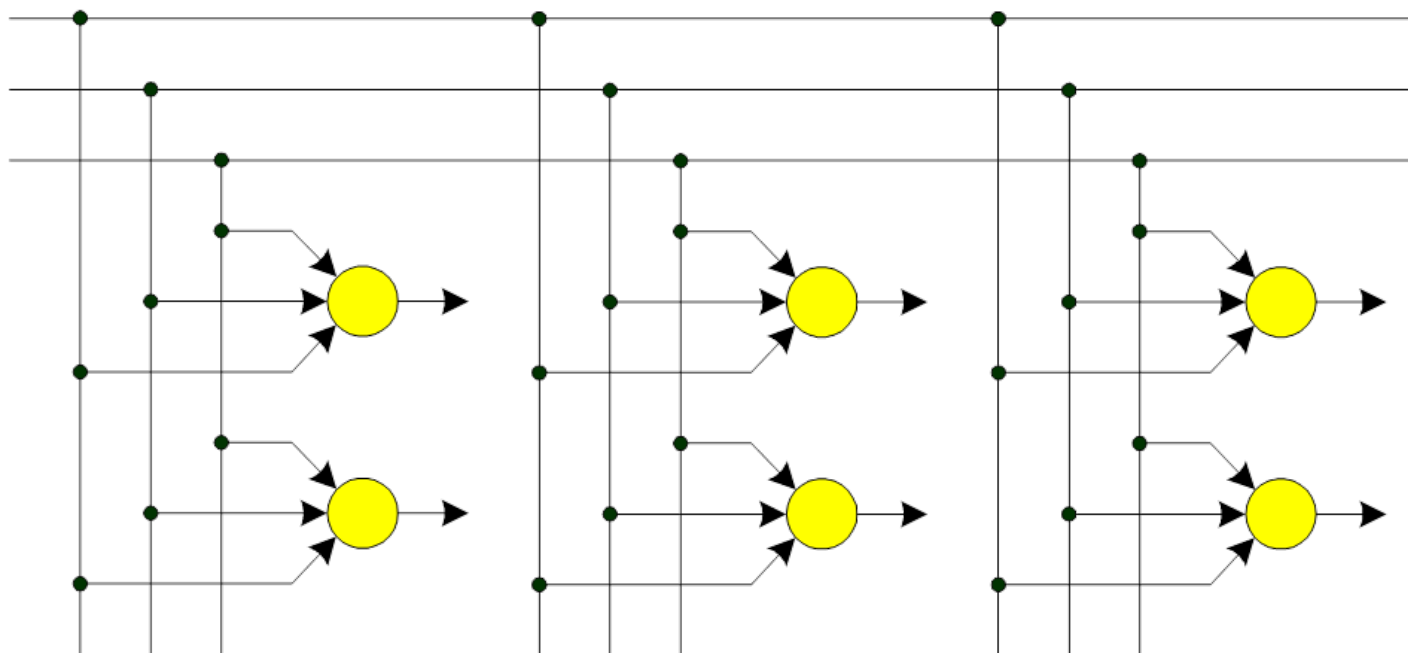
- Os nós-fonte da camada de entrada da rede fornecem os respectivos elementos do padrão de ativação, que constituem os sinais de entrada aplicados aos neurônios na segunda camada.

➤ Redes recorrentes ou realimentadas



- Difere da rede neural alimentada adiante por possuir 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 (mapas de Kohonen)

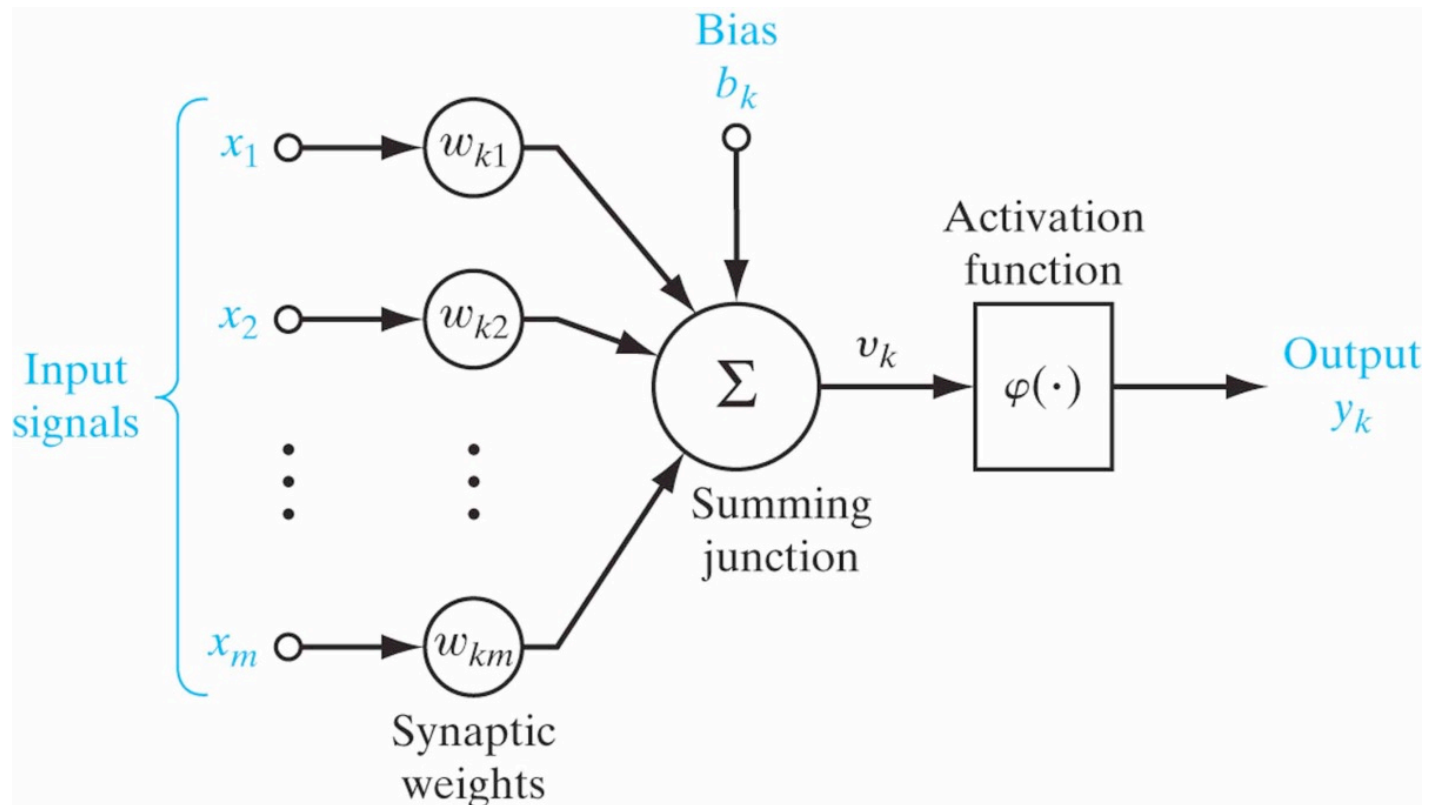
Redes Neurais Artificiais

Neurônio artificial



Neurônio artificial

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

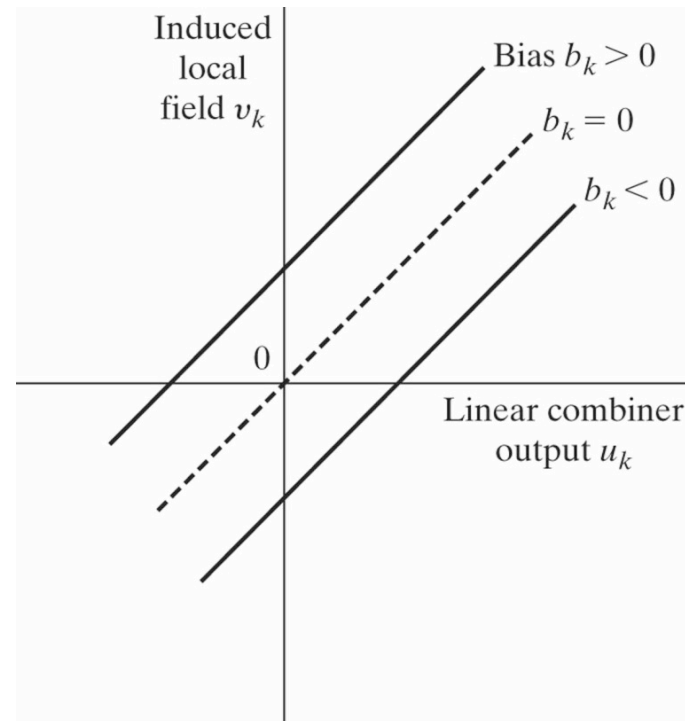


- **Sinais de entrada** $\{x_1, x_2, \dots, x_m\}$: sinais ou medidas do meio externo que representam os valores das variáveis;
- **Pesos sinápticos** $\{w_{k1}, w_{k2}, \dots, w_{km}\}$: valores que servirão para ponderar cada uma das variáveis de entrada (indicam a relevância de cada entrada na rede);
- **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;

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

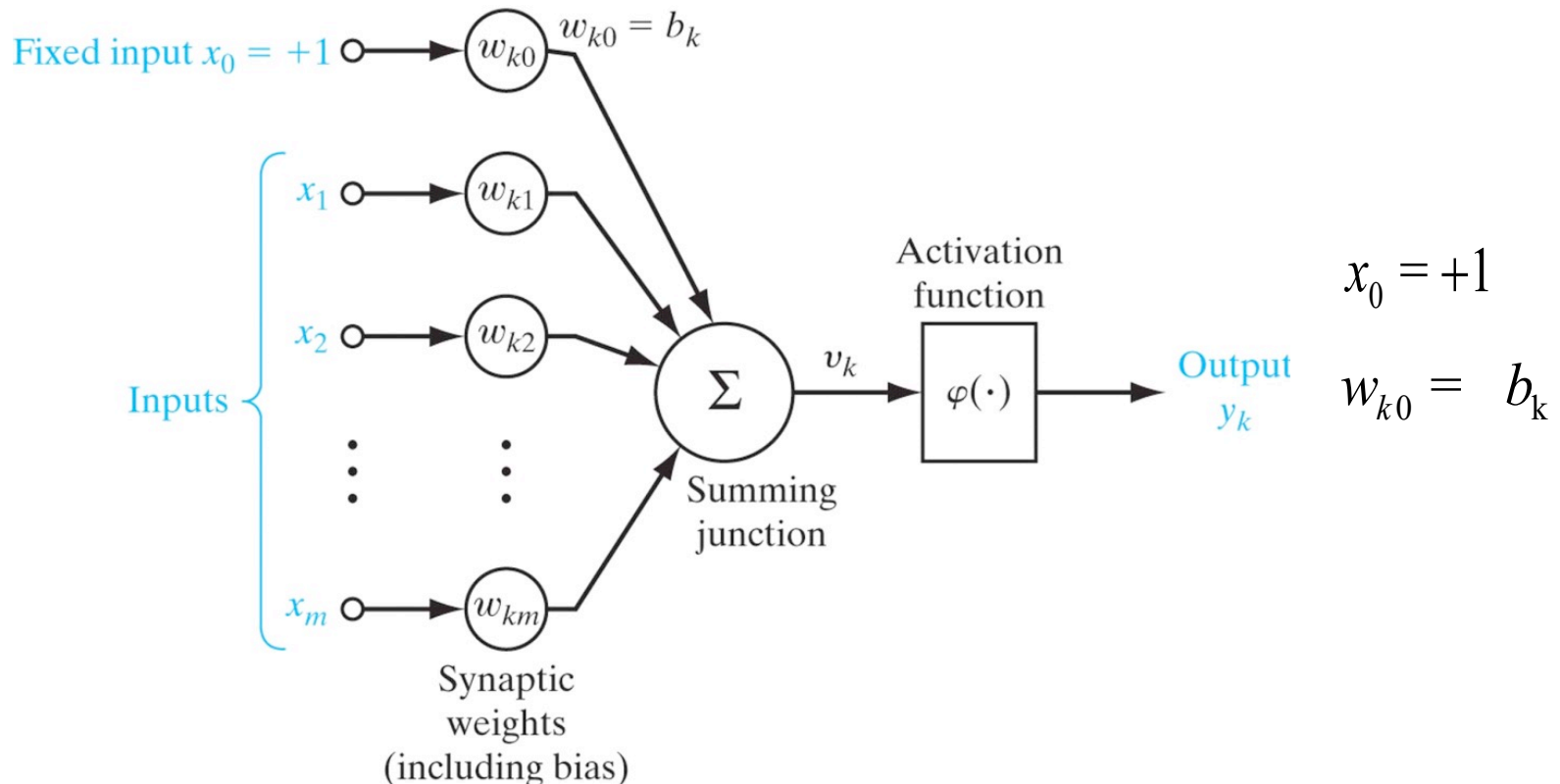
$$v_k = u_k + b_k$$

- **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.;

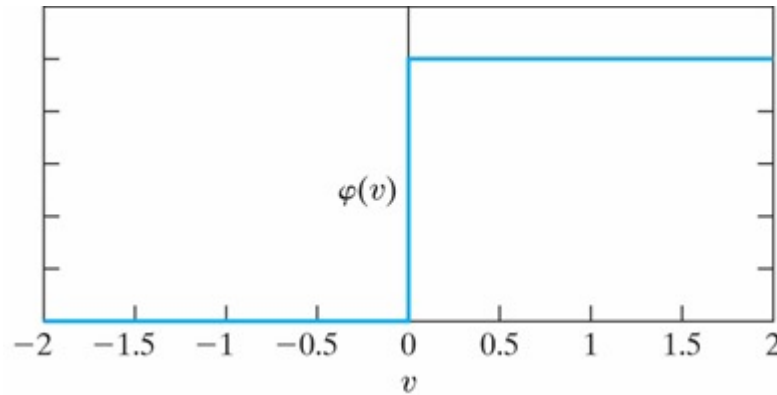
$$y_k = \varphi(u_k + b_k) = \varphi(v_k)$$

Outro modelo de neurônio

- Outro modelo matemático não-linear de um neurônio

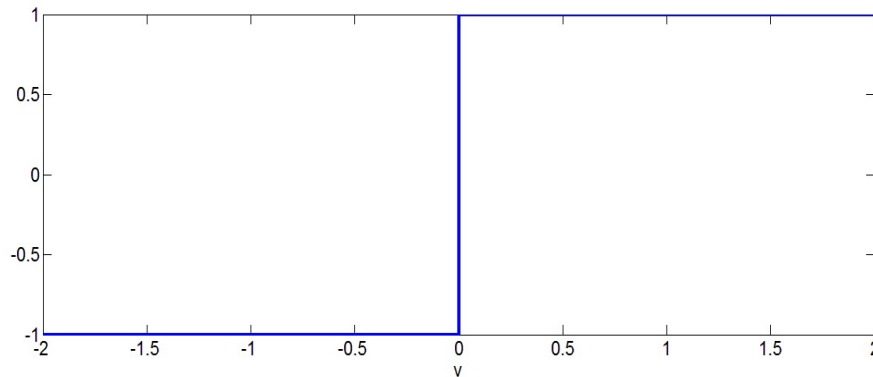


(a) Função de limiar ou função de *heaviside*.



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

(b) Função de *heaviside simétrica*.

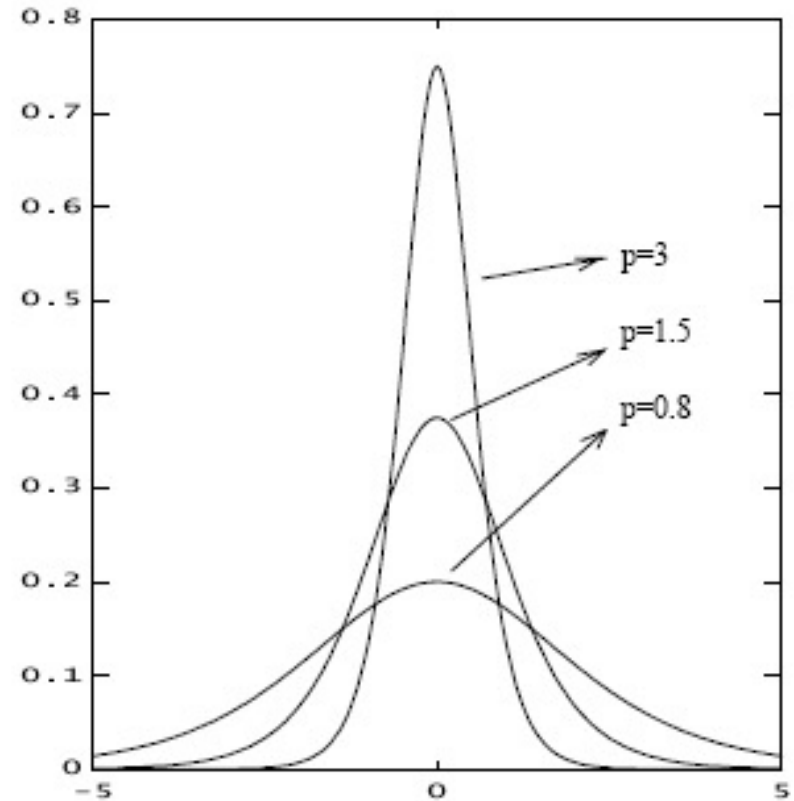
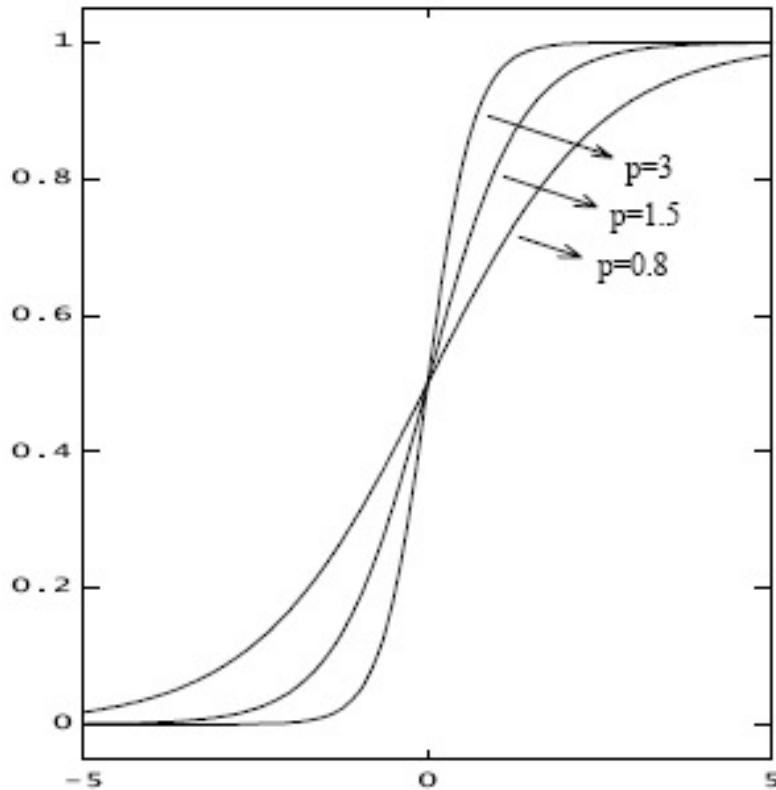


$$\varphi(v) = \begin{cases} +1 & \text{se } v > 0 \\ 0 & \text{se } v = 0 \\ -1 & \text{se } v < 0 \end{cases}$$

(c) Função de ativação logística (sigmóide).

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

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

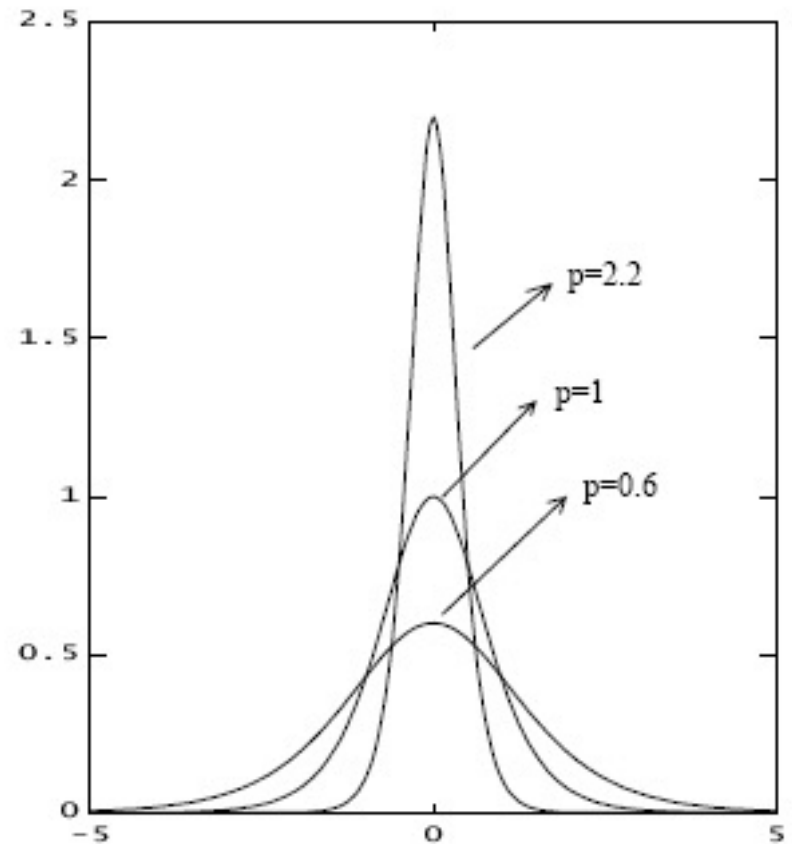
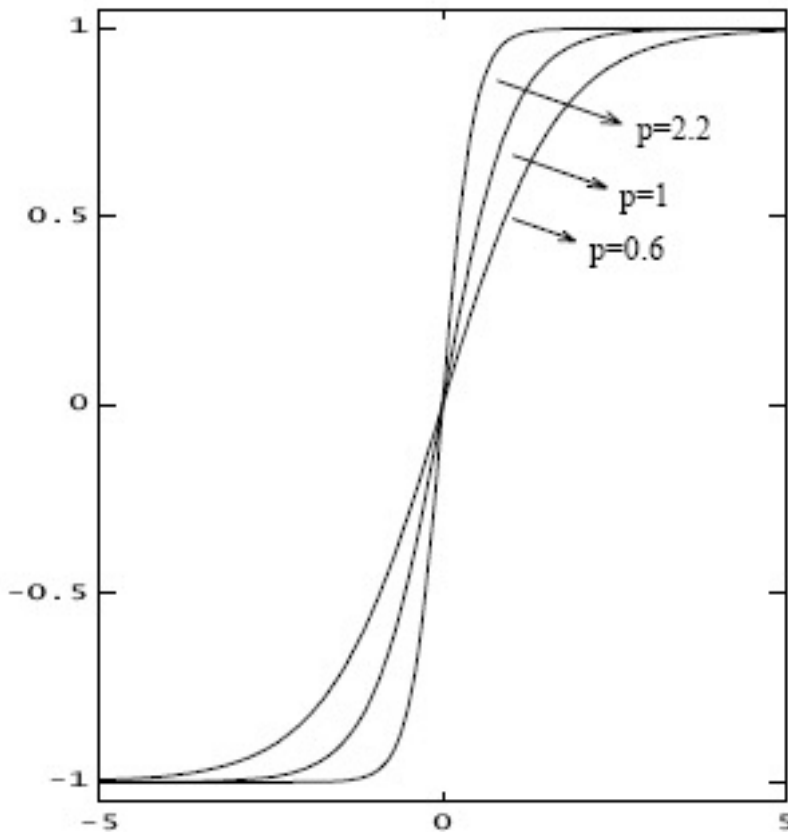


Tipos de função de ativação

(d) Função de ativação tangente hiperbólica.

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

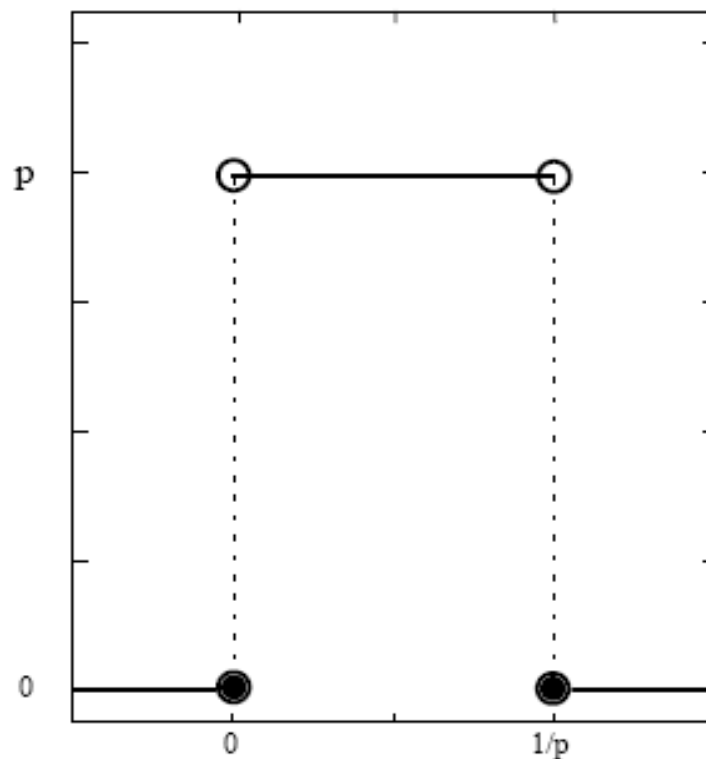
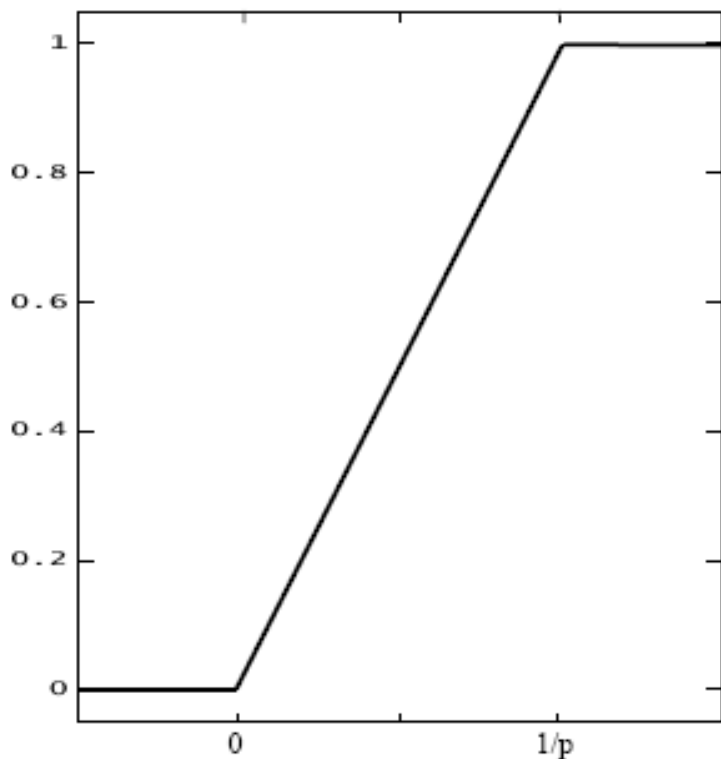
$$\frac{\partial \varphi(v)}{\partial v} = p(1 - v^2) > 0$$



(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

Processo de treinamento e aspectos de aprendizado

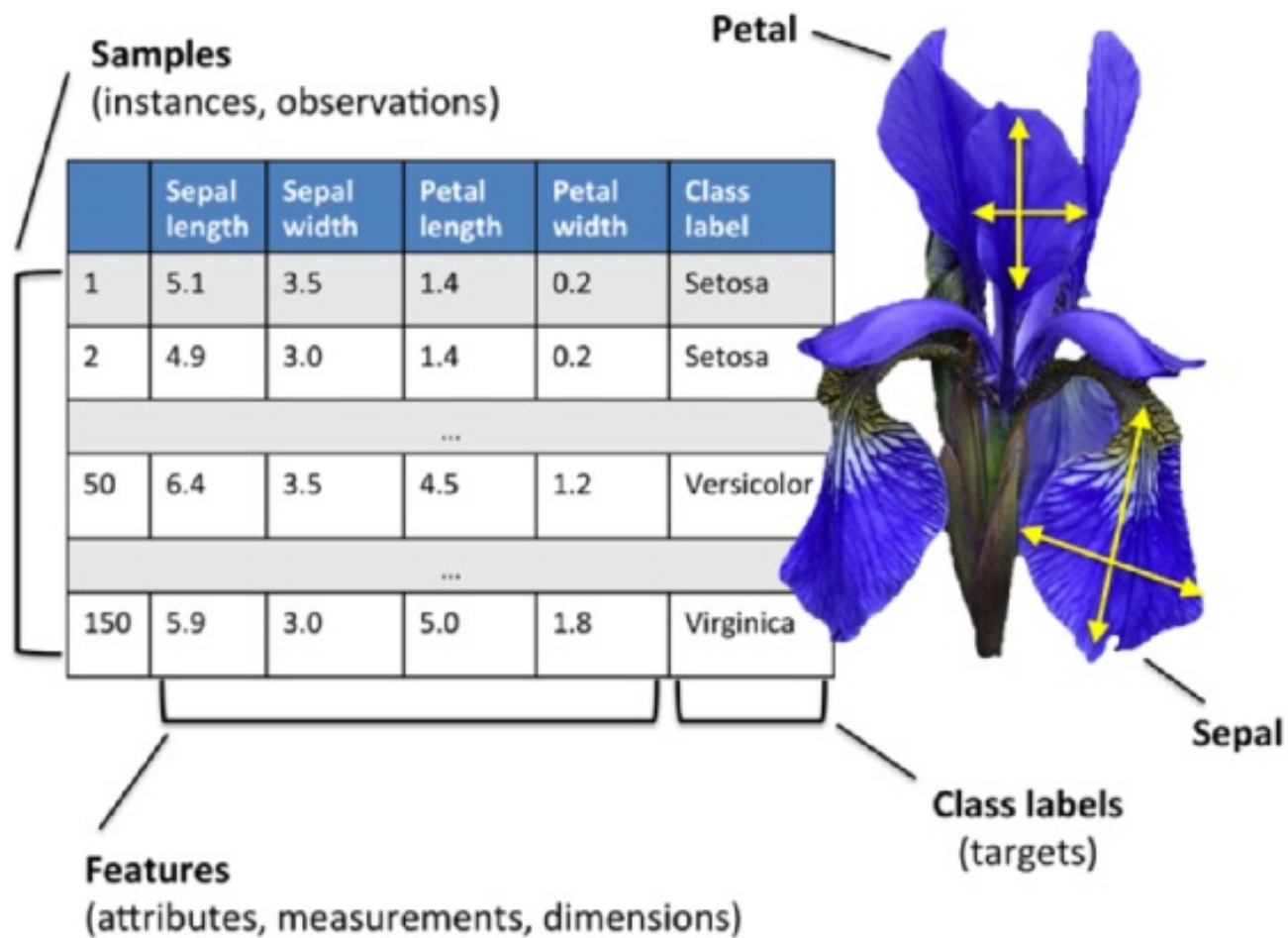


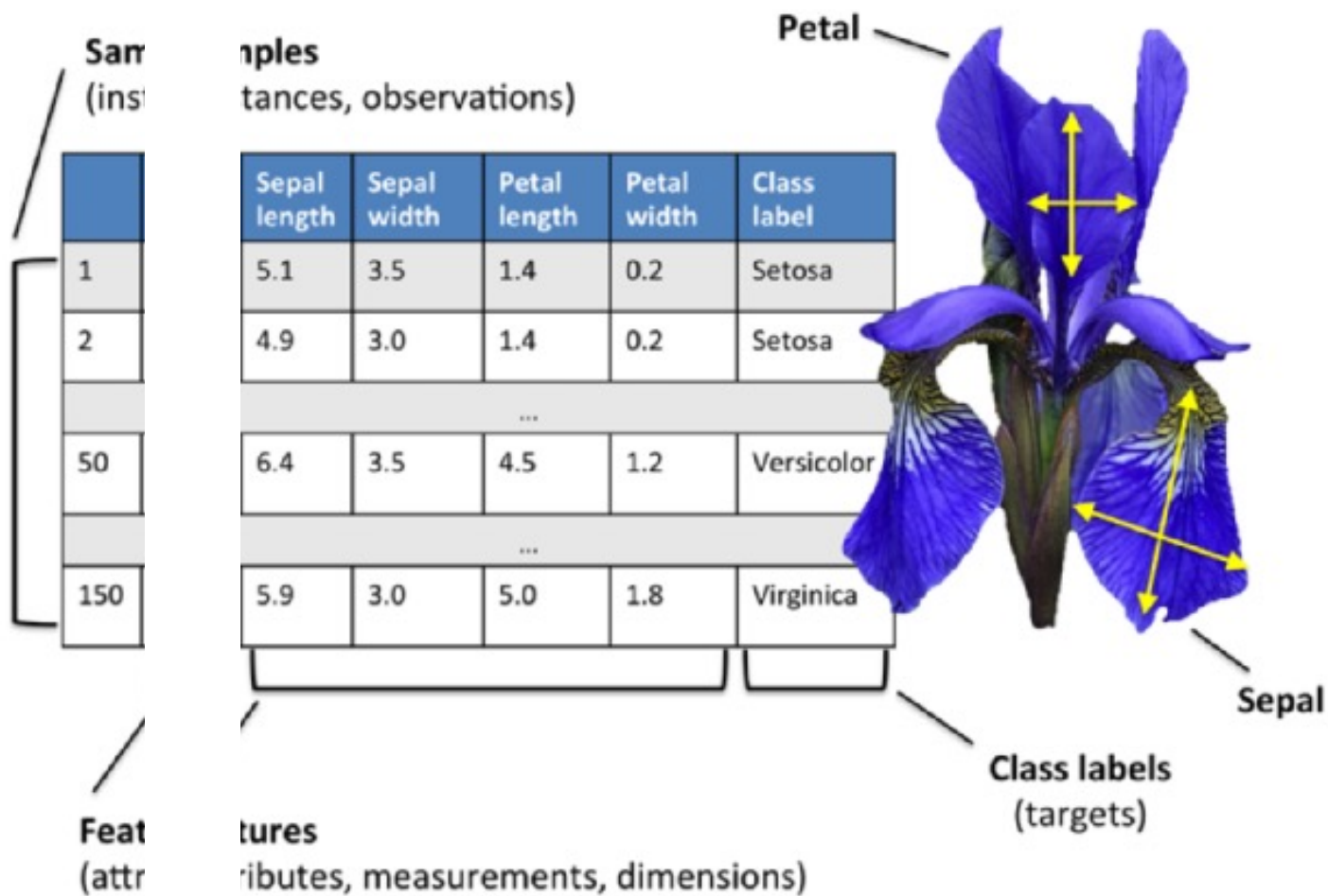
Processo de aprendizado

- 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**.

- 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.





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

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

➤ Distância euclidiana

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

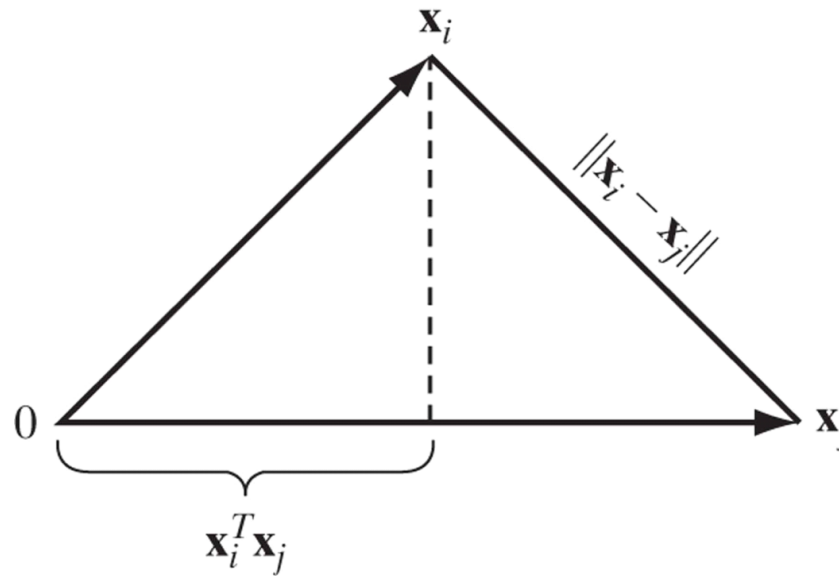
➤ Produto escalar ou produto interno

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

**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 diferentes forem X_i e X_j , maior será a distância euclidiana.
 - Quanto mais similares forem X_i e X_j , maior será o produto interno.



✓ Algoritmos

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

✓ Paradigmas

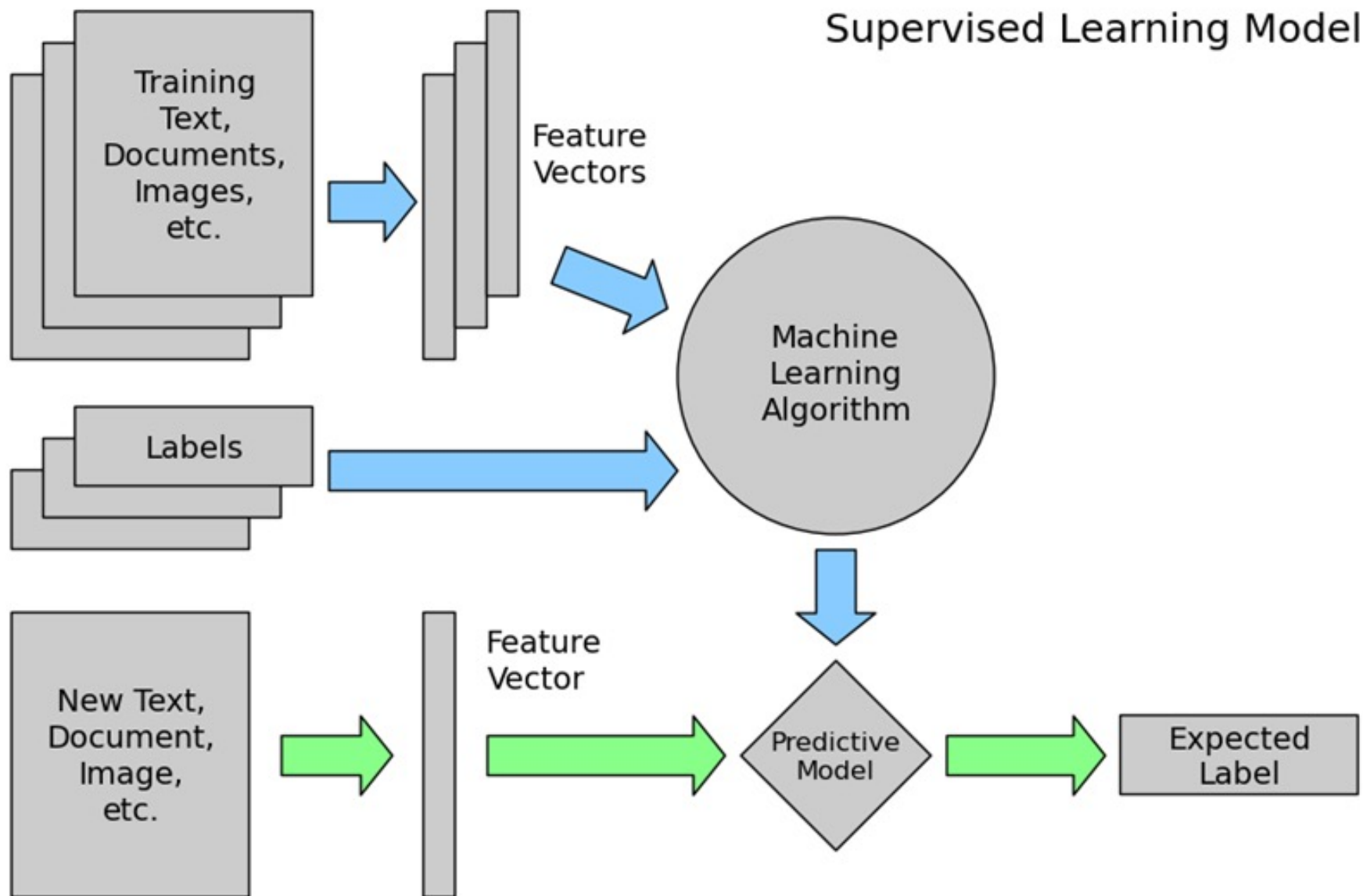
- Supervisionado
- Não supervisionado
- Reforço

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

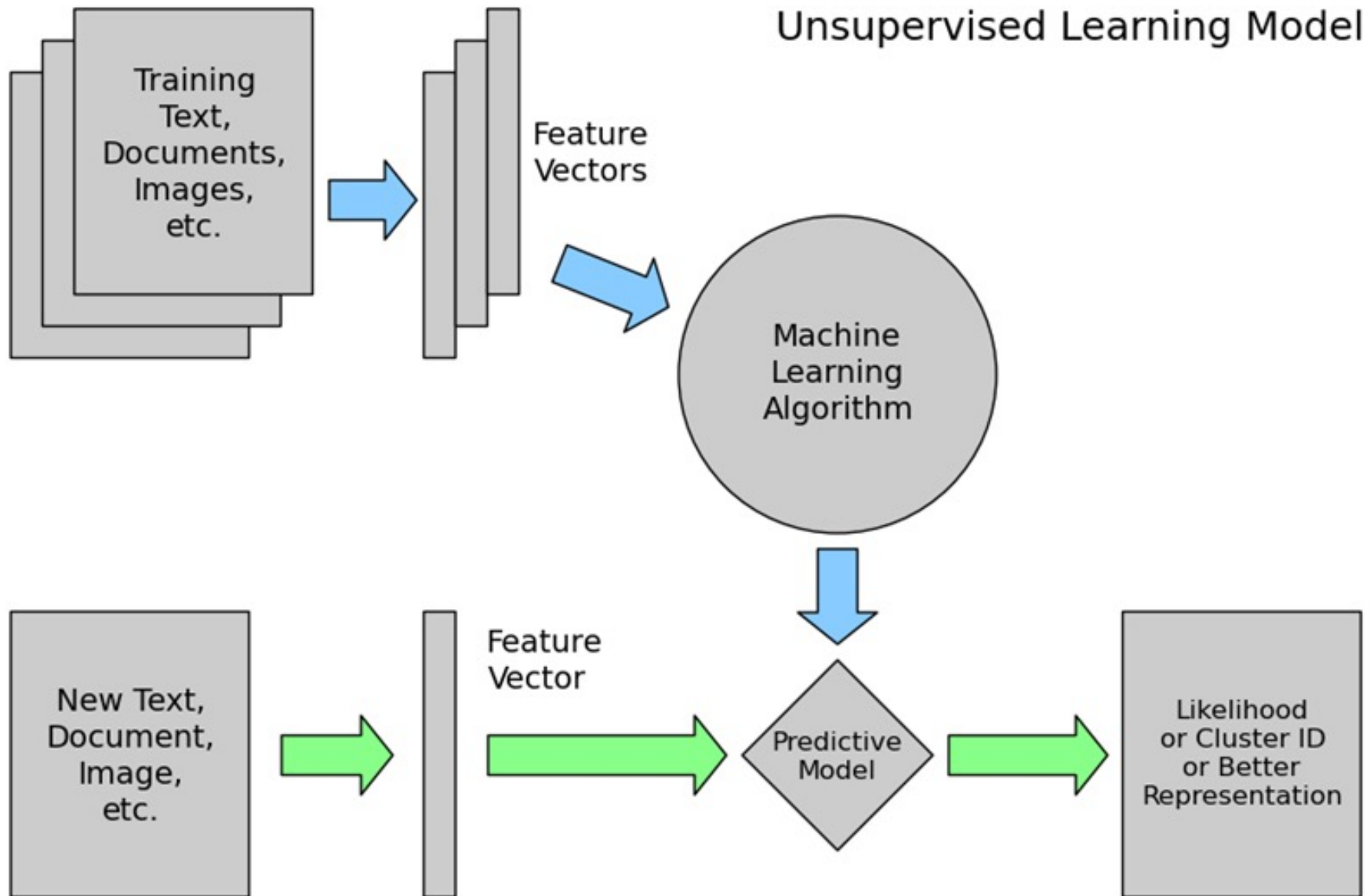
Aprendizado em redes neurais artificiais, cont.

- **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.

Aprendizagem supervisionada



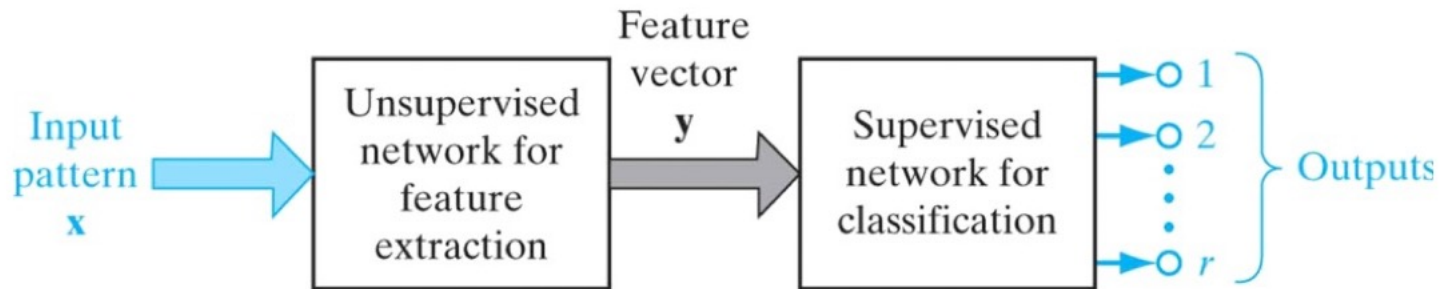
- Ambiente desconhecido pela rede neural.
- O conhecimento é apresentado por um conjunto de exemplos entrada-saída.
- A partir de um conhecimento prévio, o professor é capaz de fornecer uma resposta desejada para o vetor de treinamento;
- Os parâmetros da rede são ajustados sob a influência combinada 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.



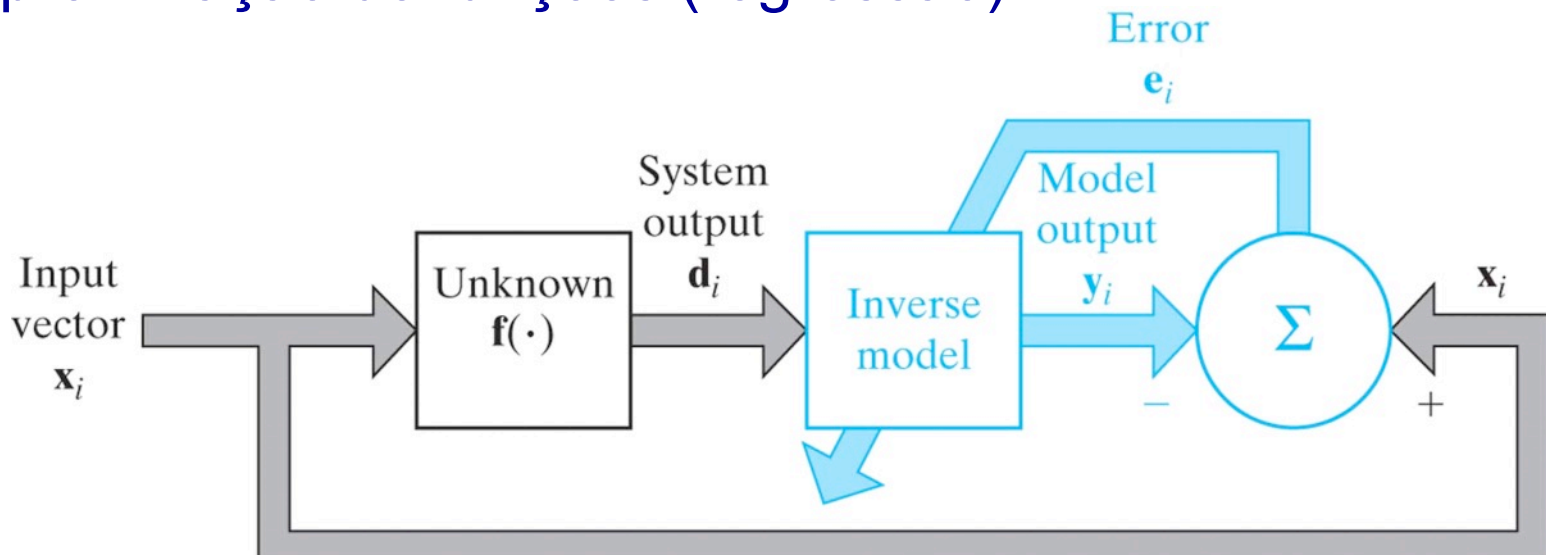
- Não há professor para supervisionar o processo de aprendizagem.
- Não há exemplos rotulados da função a ser aprendida pela rede.
- São dadas condições para realizar um **medida independente da tarefa** da qualidade da representação que a rede deve aprender.
- Pode-se utilizar a regra de aprendizagem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

Tarefas de aprendizagem

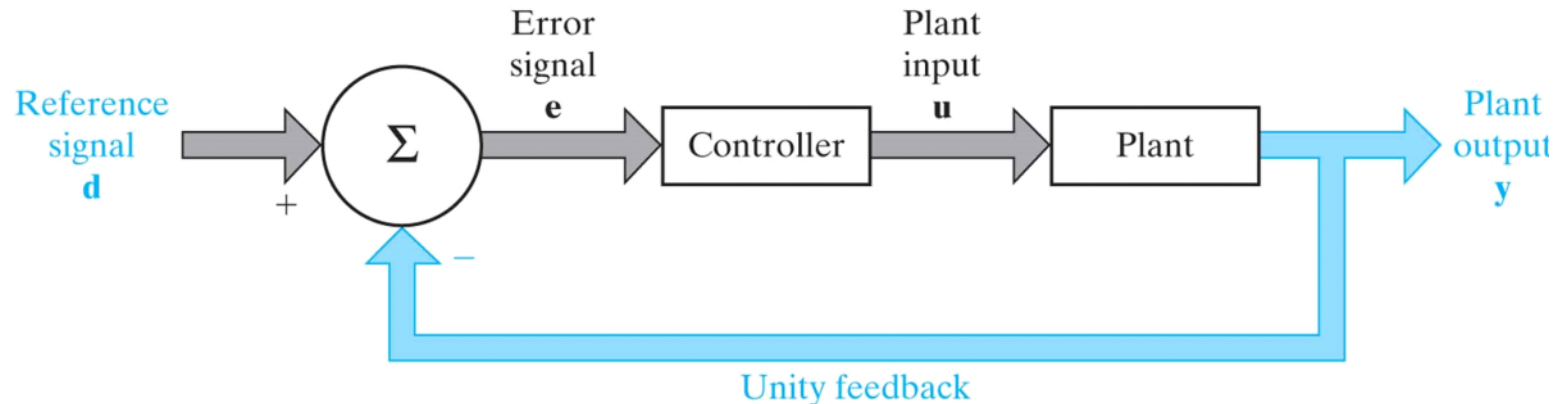
- ✓ Reconhecimento de padrões e associação



- ✓ Aproximação de funções (regressão)



✓ Controle (com redes recorrentes)



✓ Filtragem (suavização ou previsão)

- ✓ 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.



Marcelo Vinícius C. Aragão
marcelovca90@inatel.br