

Aprendizado de Máquina

Classificação - Parte II

Métodos Baseados em Otimização

- Algumas técnicas de AM buscam hipótese recorrendo à otimização de uma função:
 - Ex. erro médio quadrático
- Em problemas supervisionados, os rótulos dos objetos são considerado na formulação
- Estudaremos uma técnica:
 - Redes Neurais Artificiais (RNAs)

Redes Neurais Artificiais

- Cérebro humano é responsável pelo processamento e controle de diversas informações
- Realizamos ações que requerem atenção a diversos eventos ao mesmo tempo e processamentos variados
 - Ex. pegar objeto, caminhar, envolvem ação de diversos componentes, como memória, coordenação, aprendizado
- ⇒ Motivação na construção de máquinas inteligentes

Redes Neurais Artificiais

- Sistemas distribuídos inspirados na estrutura e funcionamento do sistema nervoso
- Objetivo: simular capacidade de aprendizado do cérebro na aquisição de conhecimento

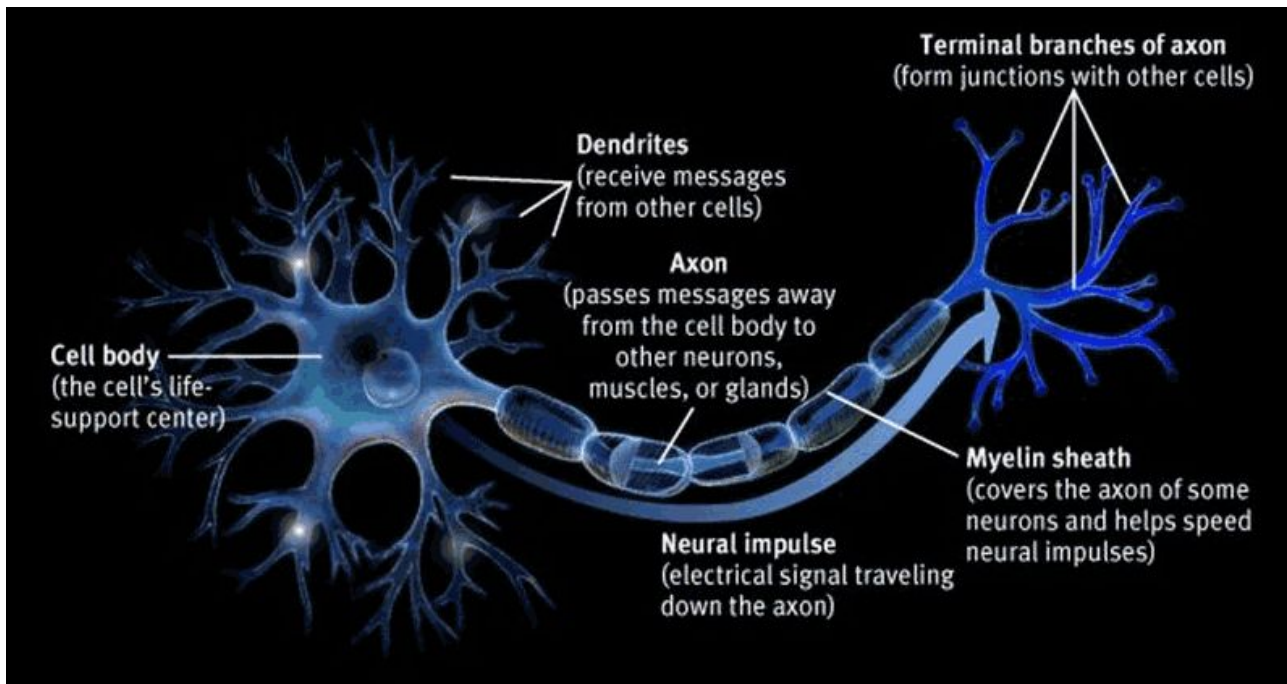
Compostas por várias unidades de processamento (“neurônios”)

Interligadas por um grande número de conexões (“sinapses”)

Redes Biológicas

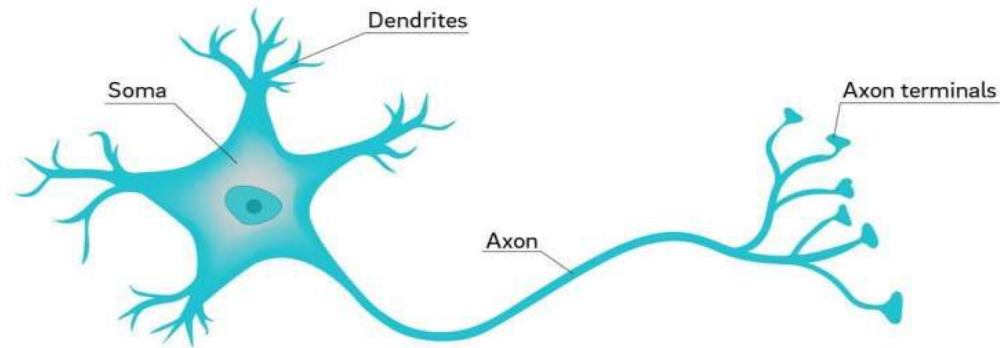
- Cérebro humano: 10^{11} neurônios
- Cada neurônio processa e se comunica com milhares de outros continuamente e em paralelo
- Cérebro: responsável por funções cognitivas e execução de funções sensoriomotoras e autônomas
- Tem capacidade de reconhecer padrões e relacioná-los, usar e armazenar conhecimento por experiência e interpretar observações

Neurônio Natural



Neurônio Natural

Neuron



Redes Biológicas



Redes Biológicas

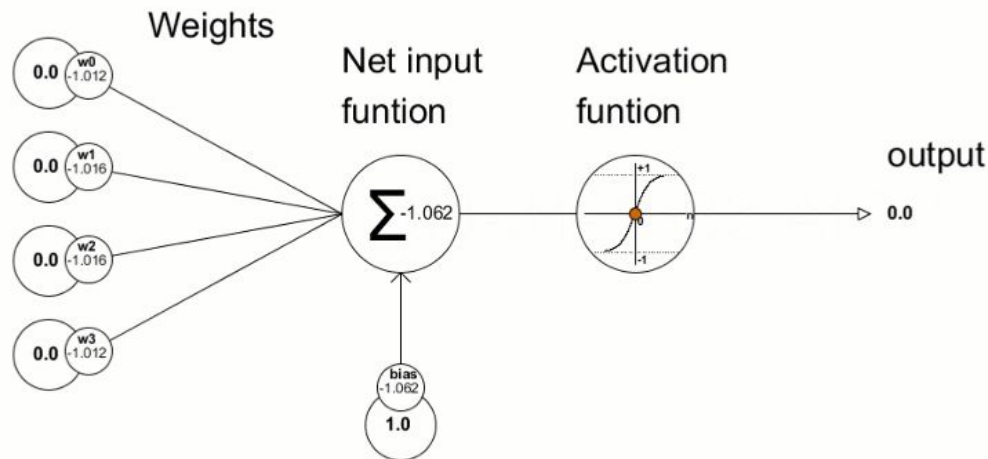
- Neurônios são bem mais lentos que os circuitos elétricos, mas o cérebro é capaz de realizar muitas tarefas mais rápido que qualquer computador
 - Redes neurais biológicas trabalham de forma massivamente paralela
 - Neurônios estão organizados em cerca de 1000 núdulos principais, cada um com 500 redes neurais
 - E cada neurônio pode estar ligado a centenas ou até milhares de outros neurônios

Rede Neural Artificial

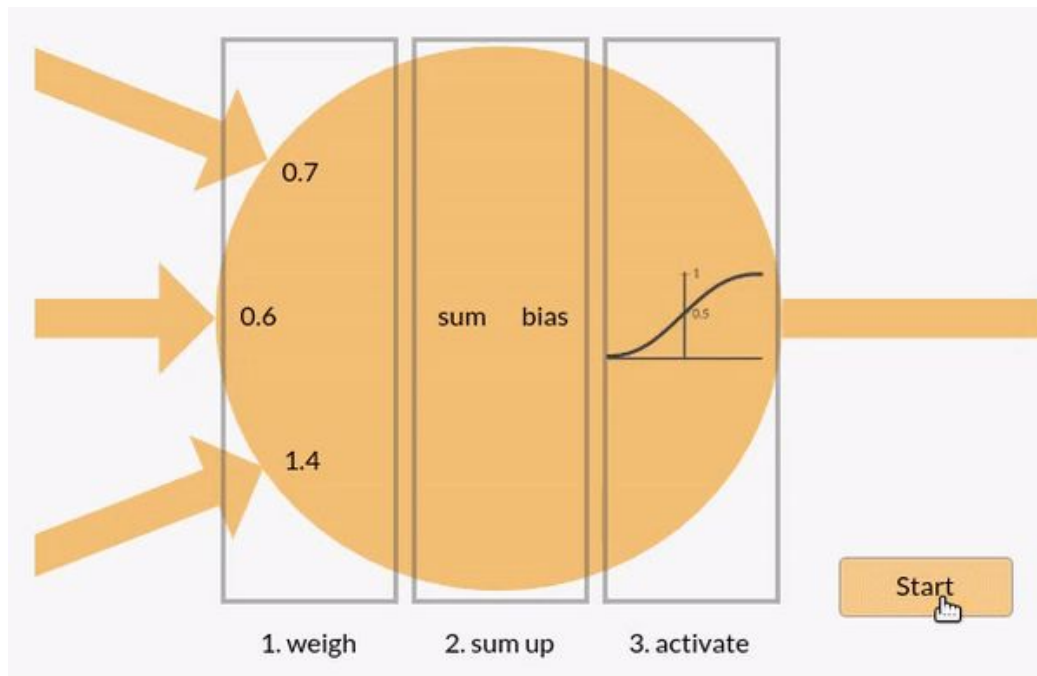
- Uma Rede Neural Artificial (RNA) é um sistema computacional que apresenta um modelo inspirado na estrutura neural do cérebro humano
- Componentes básicos:
 - Neurônio: unidade computacional básica da rede
 - Arquitetura: estrutura topológica de como os neurônios são conectados
 - Aprendizagem: processo que adapta a rede de modo a computar uma função desejada, ou realizar uma tarefa

Neurônio Artificial

- Objeto x com d atributos fornece entrada
- Pesos para as entradas são dados pelo vetor w
- É realizada uma soma ponderada da entrada, à qual é aplicada uma função de ativação, que fornece a saída final (previsão)



Funções de Ativação



Redes Multicamadas

- Modelo de rede mais popular
 - Resolvem problemas mais complexos do que o Perceptron simples
- Possuem uma ou mais camadas intermediárias
 - Funções de ativação não-lineares em pelo menos uma das camadas intermediárias
 - Sempre vai existir uma rede com uma camada equivalente a uma multicamadas com funções de ativação lineares
 - Transformações não-lineares sucessivas

Redes Multicamadas

- Grande Funcionalidade
 - Uma camada intermediária aproxima:
 - Qualquer função contínua ou Booleana
 - Duas camadas intermediárias aproximam:
 - Qualquer função
 - Qualidade da aproximação depende da complexidade da rede
 - Usar muitas camadas e/ou neurônios
 - Risco de overfitting!

Redes Multicamadas

- MLP – Multilayer Perceptron
 - Uma ou mais camadas intermediárias de neurônios
 - Função de ativação Sigmoide ou Tangente Hiperbólica
 - Arquitetura mais comum: completamente conectada
 - Cada neurônio realiza uma função específica
 - Função implementada é uma combinação das funções realizadas pelos neurônios da camada anterior conectados a ele

*Termo θ usado para desvio da função em relação à origem
⇒ Todos neurônios vão ter um $w_0 = \theta$ (ou $-\theta$) com entrada fixa -1 (ou $+1$)*

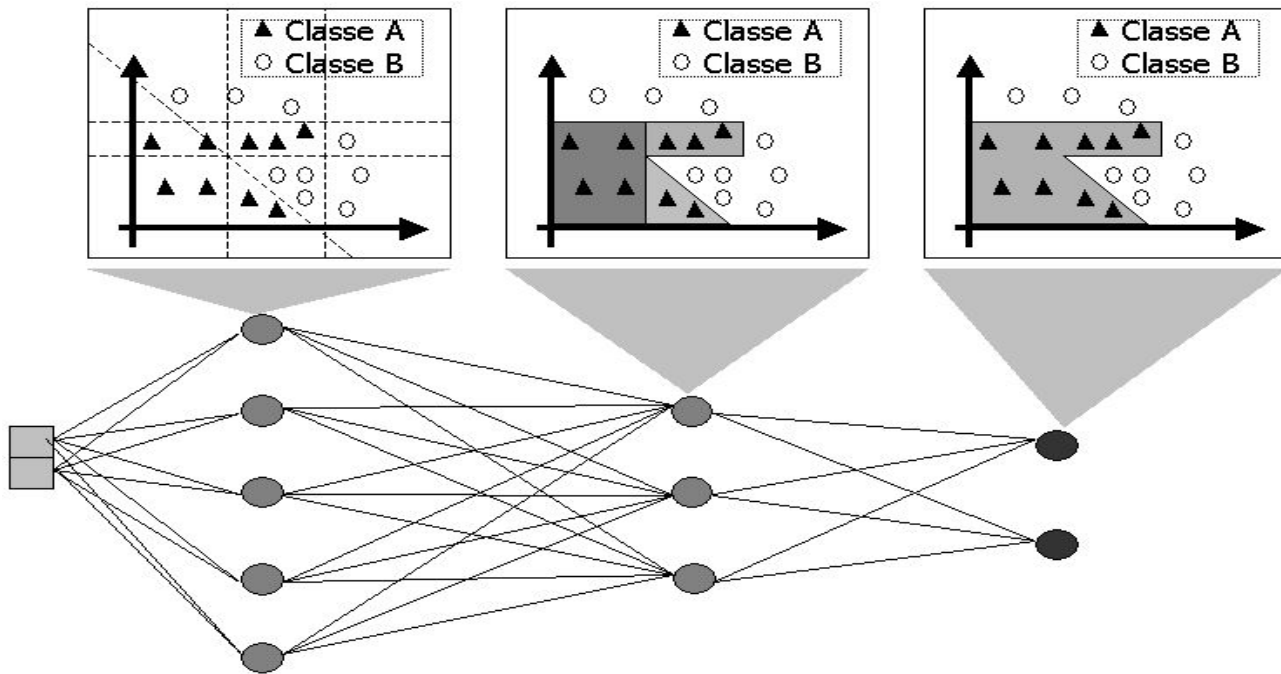
MLP

- Função implementada por cada neurônio
 - Formada pela combinação das funções implementadas por neurônios da camada anterior
 - Camada 1: hiperplanos no espaço de entradas
 - Camada 2: regiões convexas
 - Número de lados = número de unidades na camada anterior
 - Camada 3: Combinações de figuras convexas, produzindo formatos abstratos
 - Número de figuras convexas = número de unidades da camada anterior



*Funções cada vez
mais complexas*

Intuição



Combinação das
funções
desempenhadas
por cada neurônio
define a função
associada à RNA

MLP

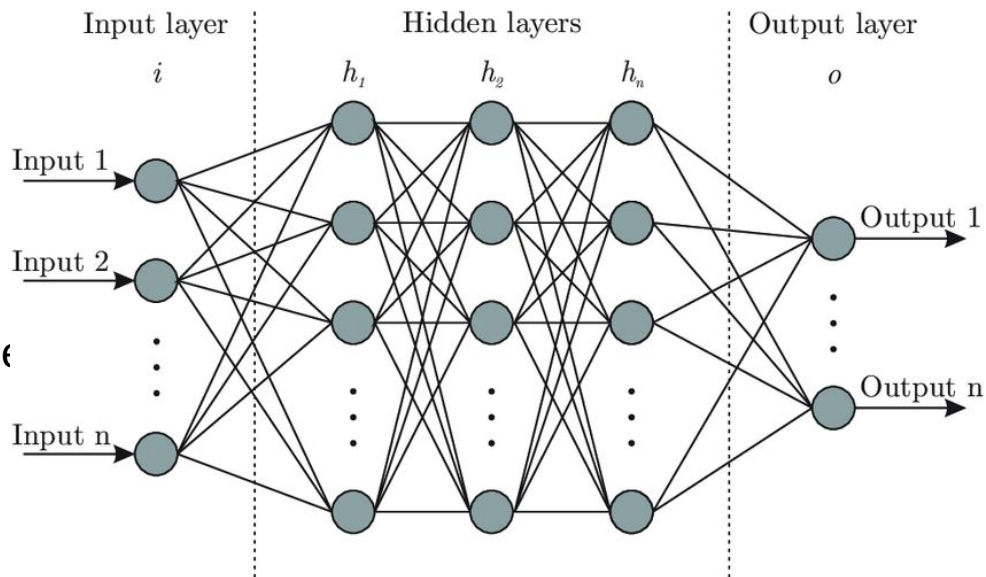
- Camada de saída: um neurônio para cada um dos rótulos presentes
 - Classificação: função de ativação Sigmoidal ou tangente hiperbólica
 - Regressão: função de ativação linear
- Saída para um objeto \mathbf{x} : $\mathbf{y} = [y_1, y_2, \dots, y_k]^t$, k = número de rótulos
- Classificação: vetor \mathbf{y} para cada objeto de entrada tem valor 1 na posição associada à classe do objeto e 0 nas demais posições

Topologia

- Definida por:
 - Número de camadas da RNA
 - Número de neurônios em cada camada
 - Grau de conectividade dos neurônios
 - Presença ou não de conexões de retropropagação

Topologia

- Neurônios podem estar dispostos em camadas
 - Neurônio pode receber como entrada a saída de neurônios da camada anterior
 - E enviar sua saída para entrada de neurônios em camada seguinte



Função custo

- Toda rede neural gera uma saída.
- Essa saída deve ser analisada e comparada com o dado que se tem sobre o objeto que foi classificado. Para realizar essa tarefa, é utilizada uma função custo da rede para um conjunto de dados é:

$$L = \frac{1}{N} \sum_i L_i(f(x_i, W), y_i)$$

- Essa função mede o quão diferente a saída da rede está do rótulo real dos objetos.

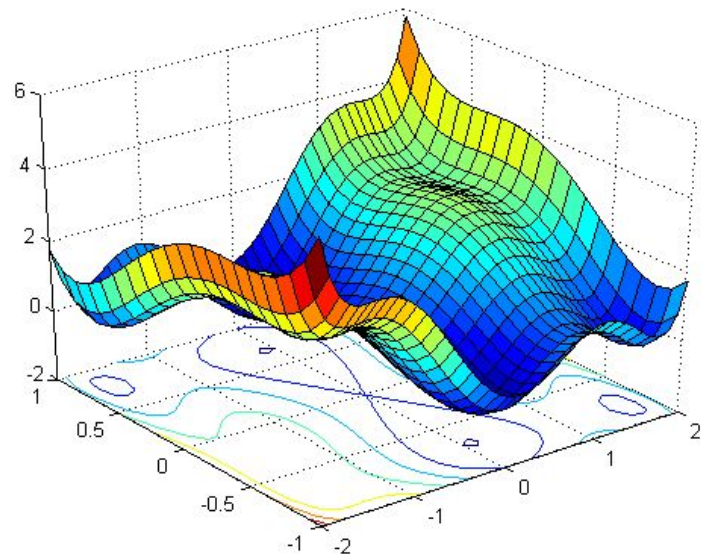
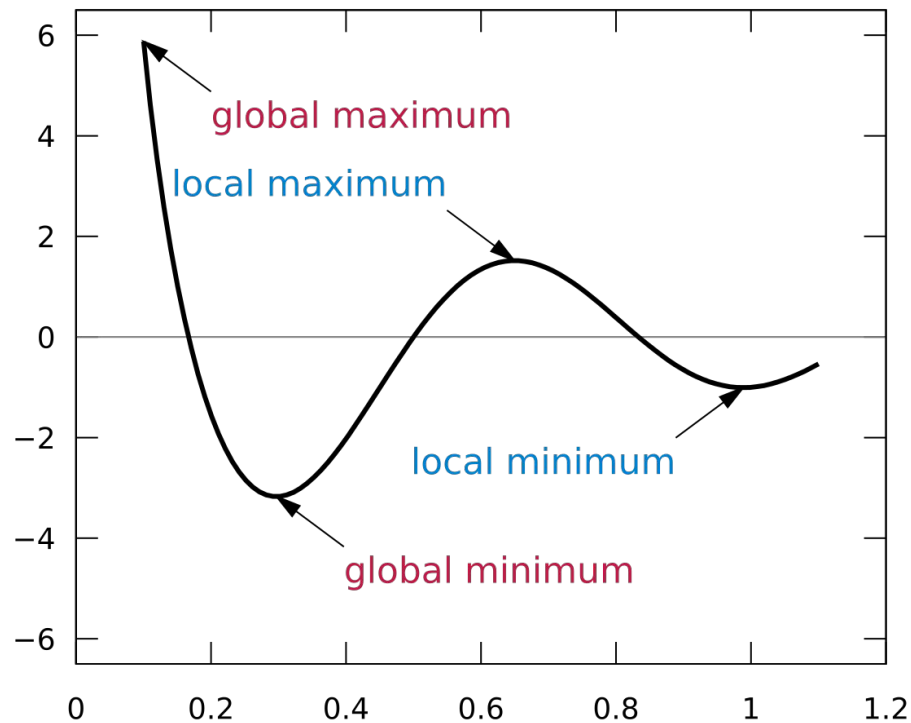
Aprendizado

- Três grupos de algoritmos:
- Correção de erros: procuram ajustar pesos para reduzir erros cometidos pela rede
 - (supervisionado)
- Hebbiano: baseados na regra de Hebb, se dois neurônios estão simultaneamente ativos, a conexão entre eles deve ser reforçada
 - (não supervisionado)
- Competitivo: promovem competição entre neurônios para definir quais terão pesos ajustados, geralmente os que respondem mais fortemente à entrada
 - (não supervisionado)

Aprendizado por Correção de Erro

- Superfície de erro
 - Superfície multi-dimensional representando gráfico da função de custos X peso
 - Objetivo do aprendizado:
 - A partir de um ponto qualquer da superfície, mover em direção a um mínimo global

Superfície de erro



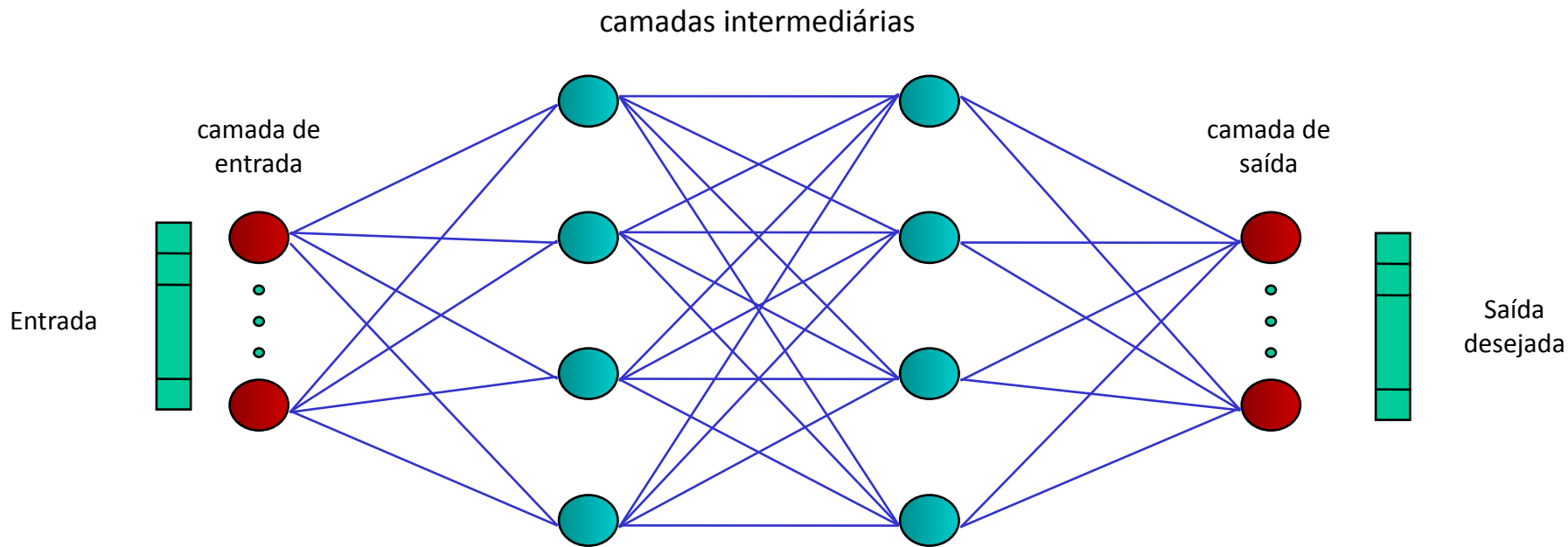
Backpropagation

- Treinamento por correção de erros
 - Camada de saída: comparação entre vetor de saída dos neurônios e vetor de valores desejados
 - Classificação: rede classifica objeto corretamente quando a saída mais elevada é a do neurônio correspondente à classe correta do exemplo
 - Se valores são baixos ou mais de um neurônio dá valor de saída alto, a rede não tem condições de prever
 - Camadas intermediárias: Qual a saída desejada de uma camada intermediária?
 - Algoritmo Backpropagation (retropropagação de erros)

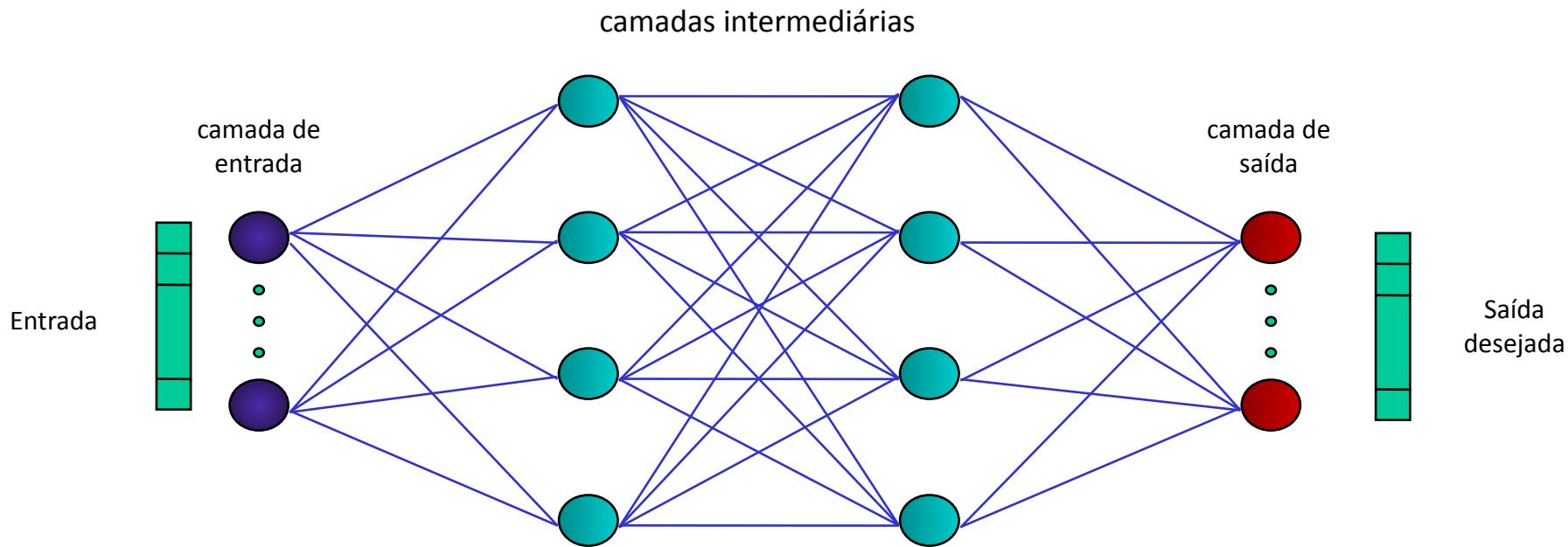
Backpropagation

- Treinamento: iteração de duas fases
 - Cada fase percorre a rede em dois sentidos
 - Sinal (forward)
 - Erro (backward)

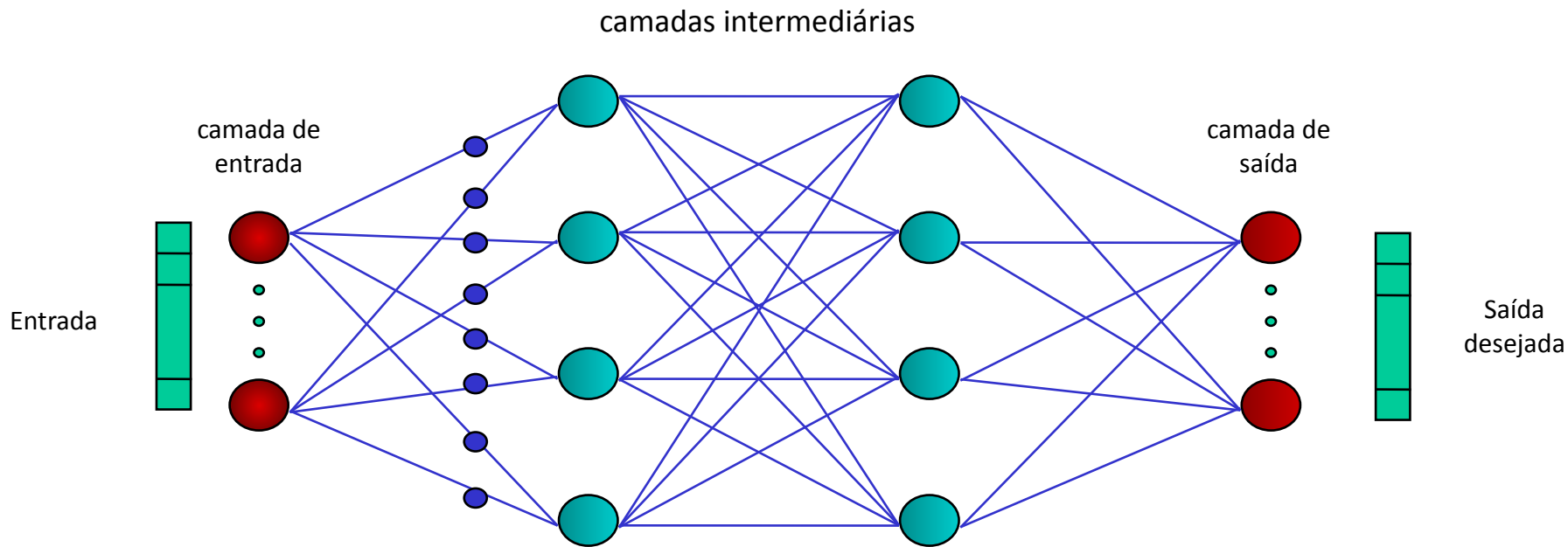
Aprendizado



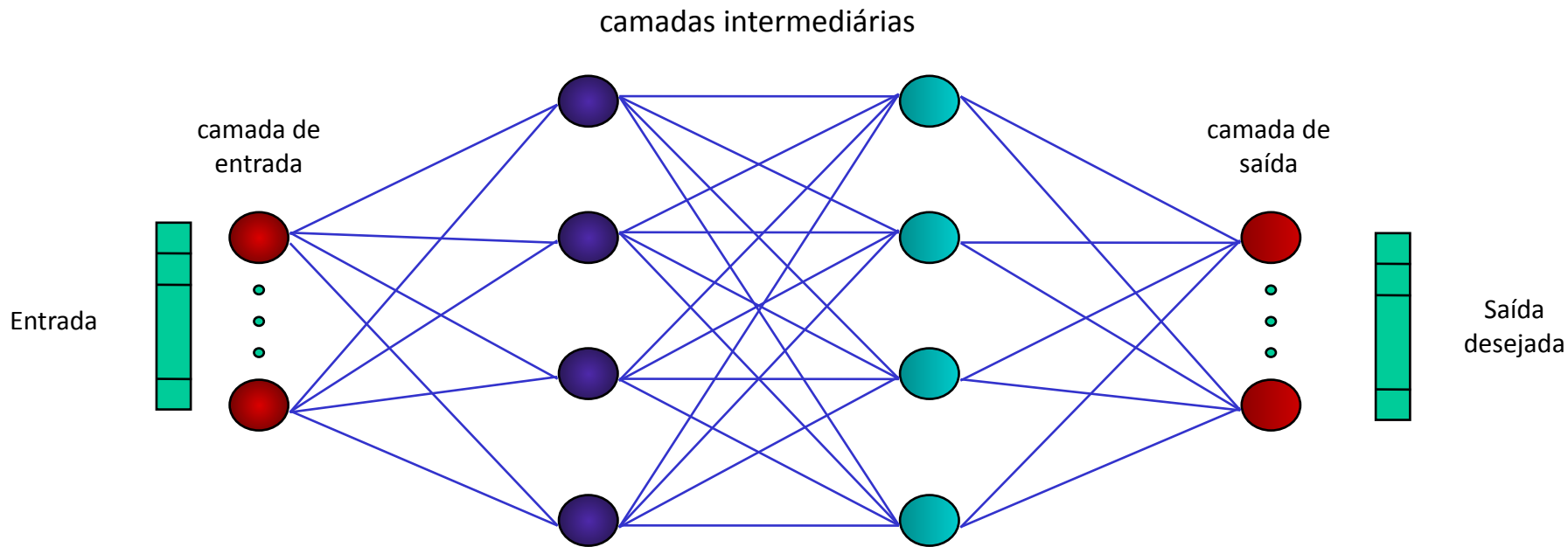
Aprendizado



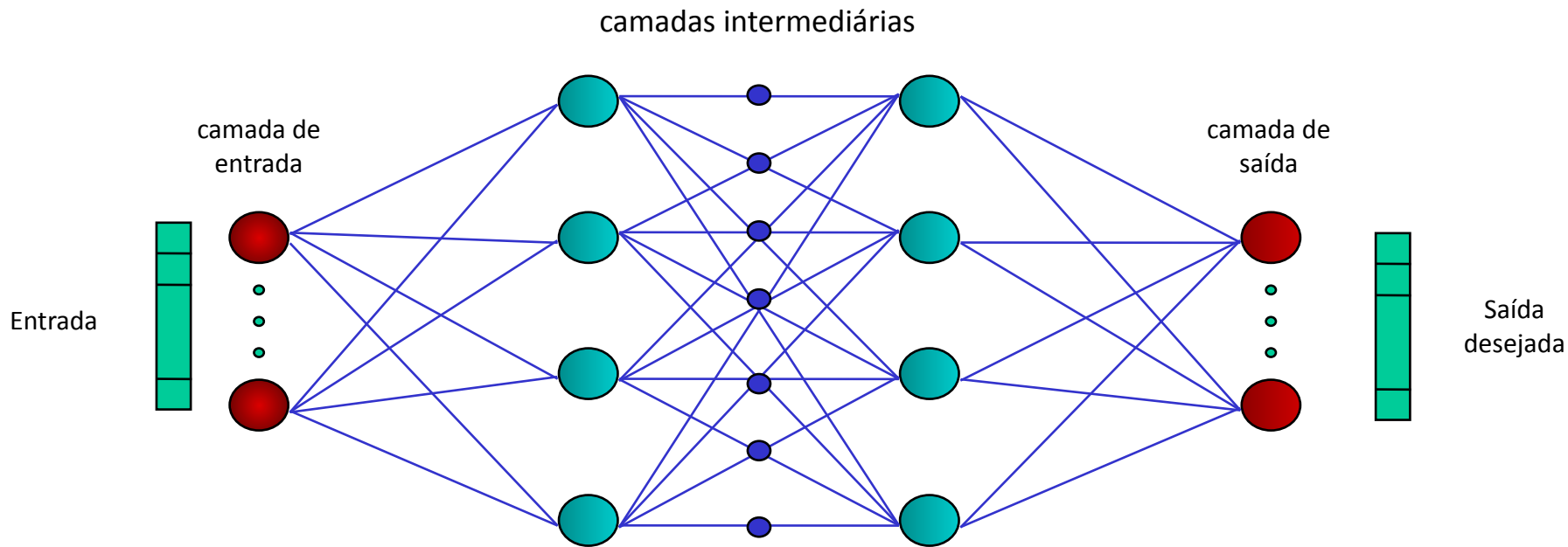
Aprendizado



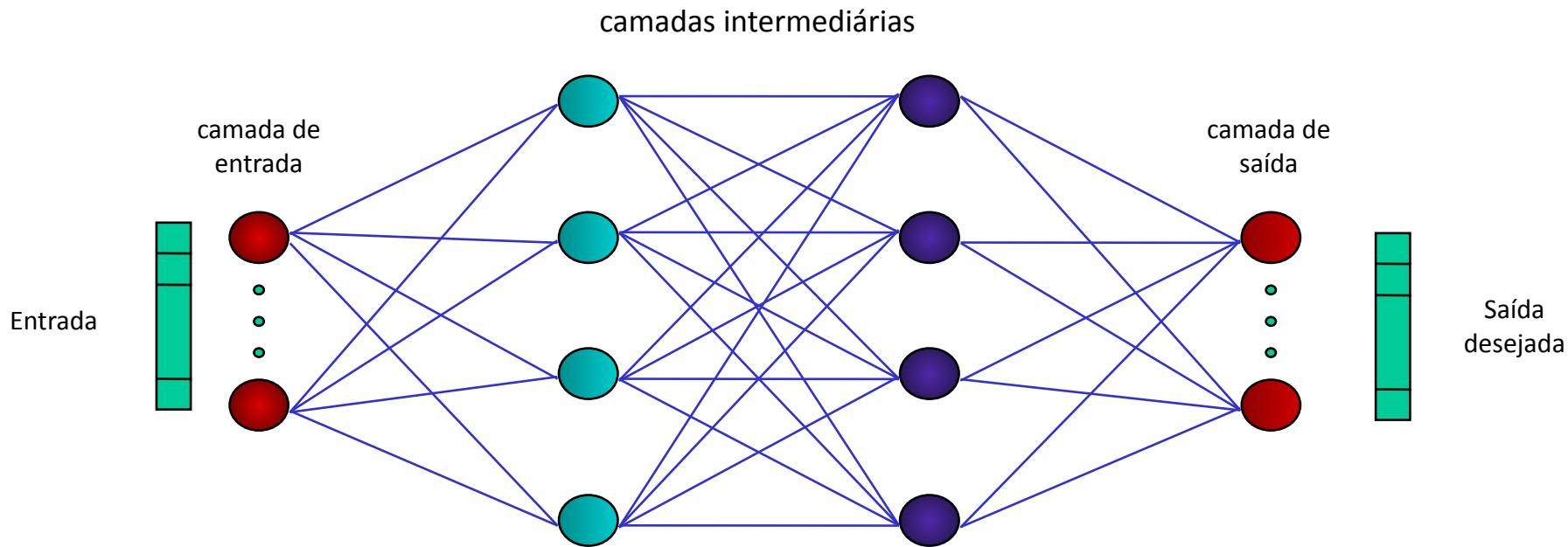
Aprendizado



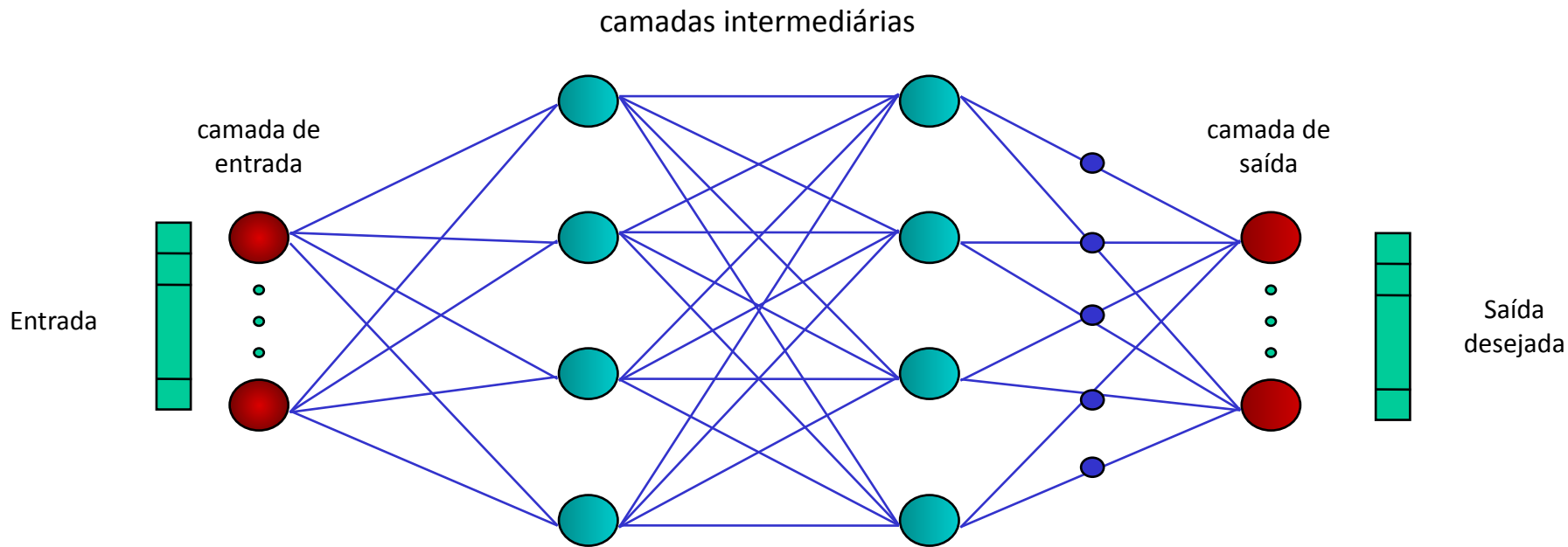
Aprendizado



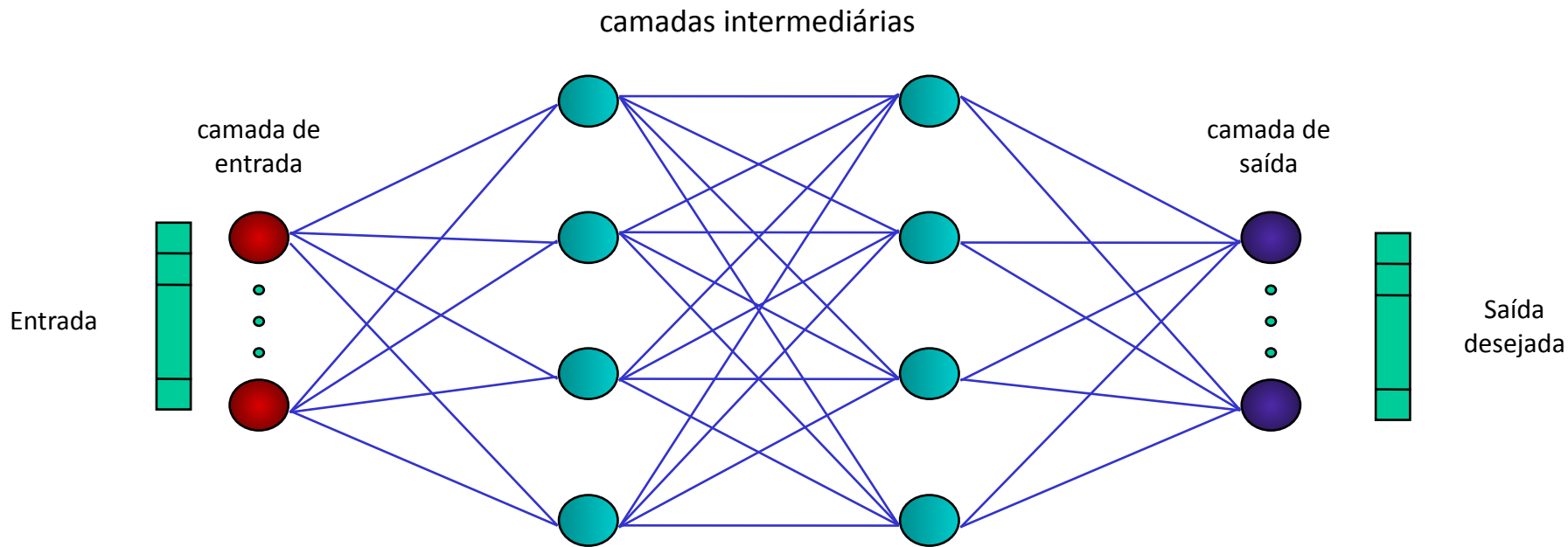
Aprendizado



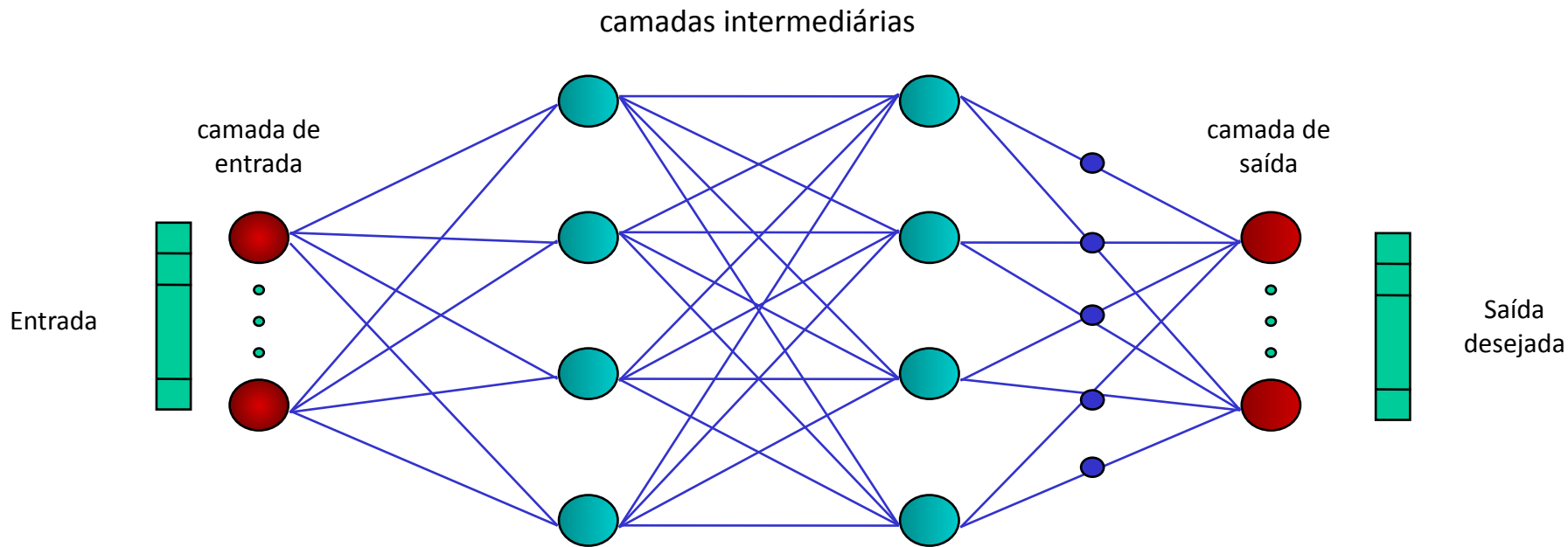
Aprendizado



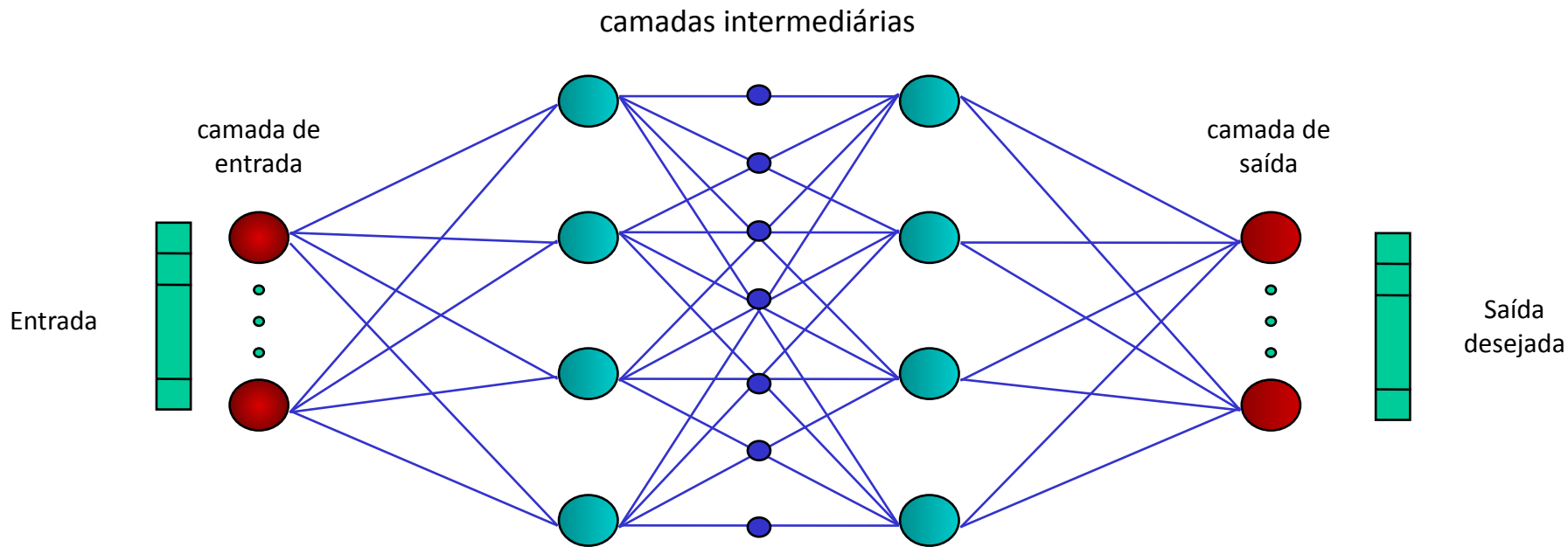
Aprendizado



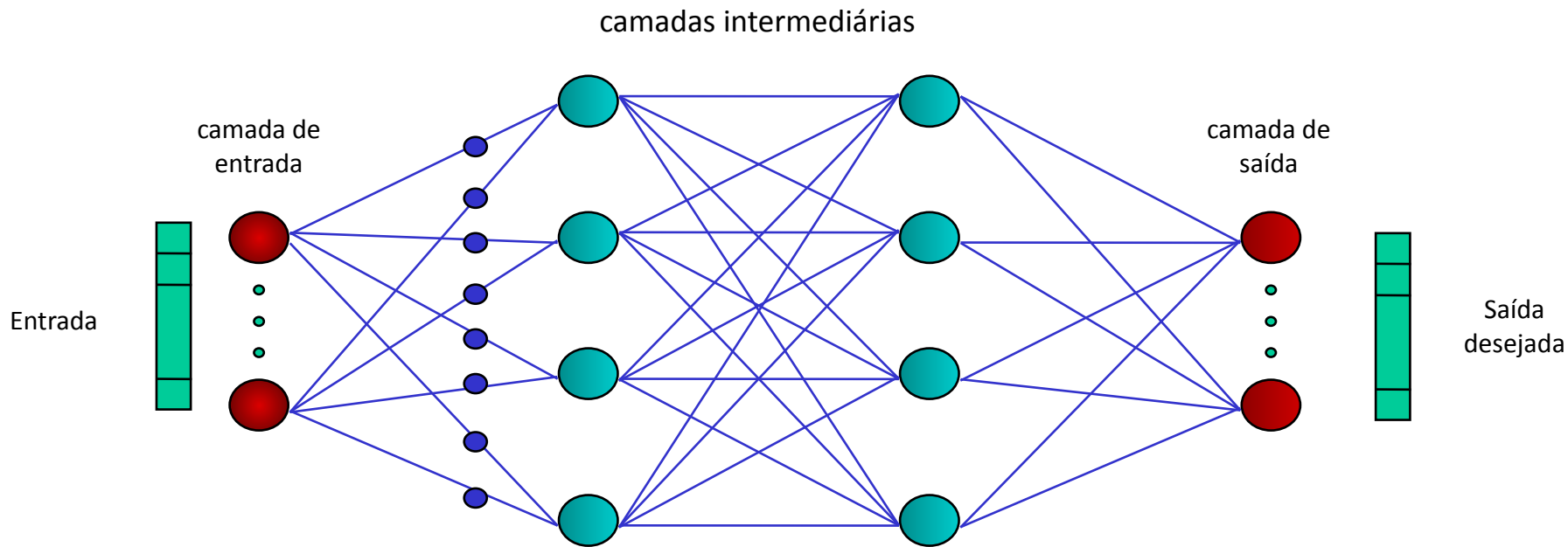
Aprendizado



Aprendizado



Aprendizado



Aprendizado

- Versão padrão: ajuste de pesos para cada objeto individualmente
- Variação batch: pesos são ajustados uma única vez para cada conjunto de exemplos.

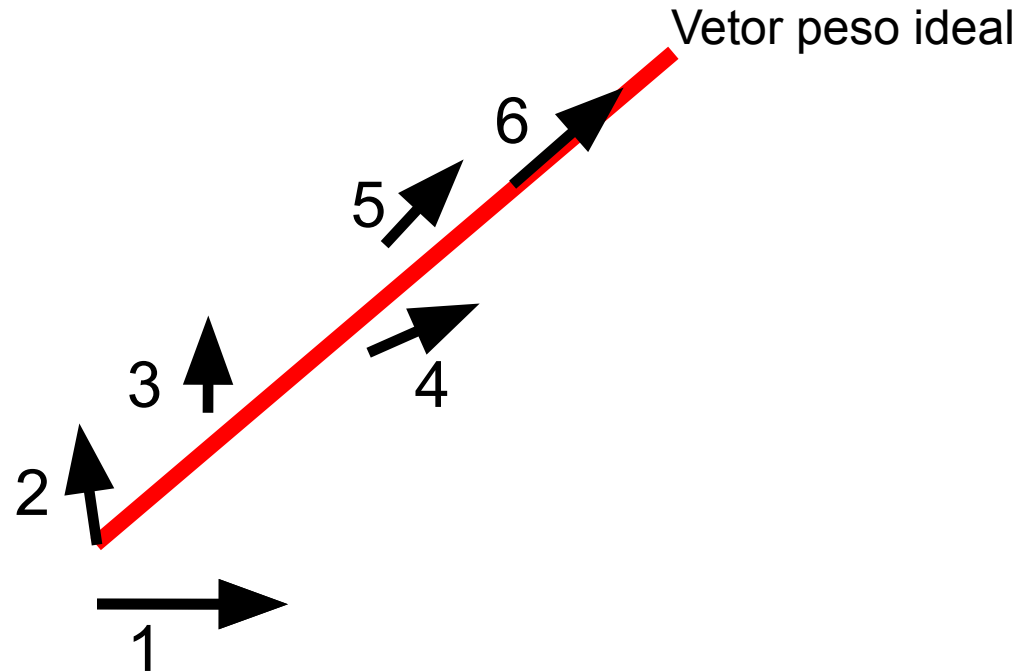
Aprendizado

- Ajuste dos pesos das conexões

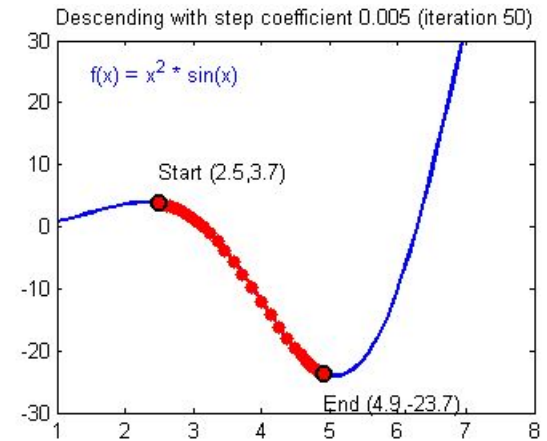
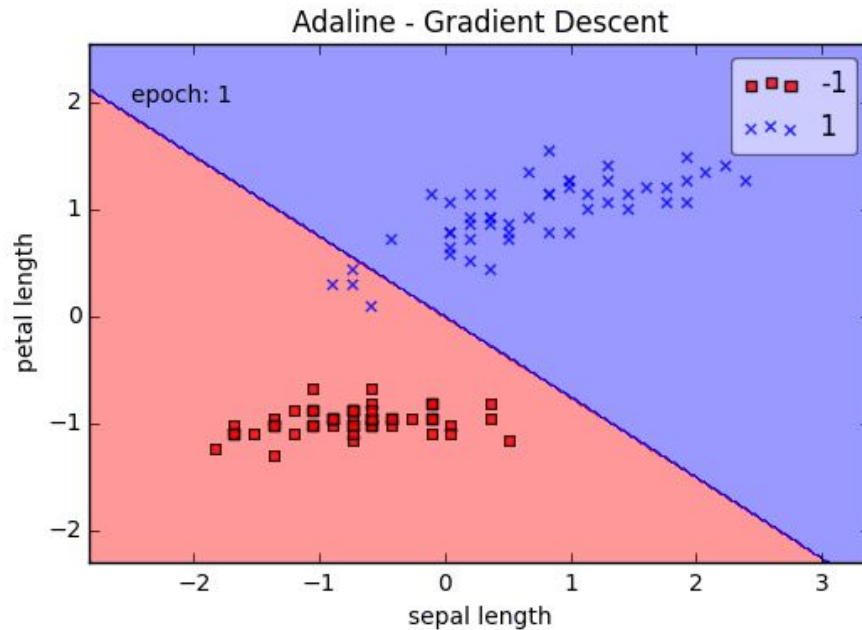
$$w(t + 1) = w(t) + \nabla w(t)$$

- Algoritmos de aprendizado
 - Conjunto de regras bem definidas para ensinar a rede a resolver um dado problema
 - Divergem na maneira como os pesos são ajustados
 - Em como ∇w é calculado

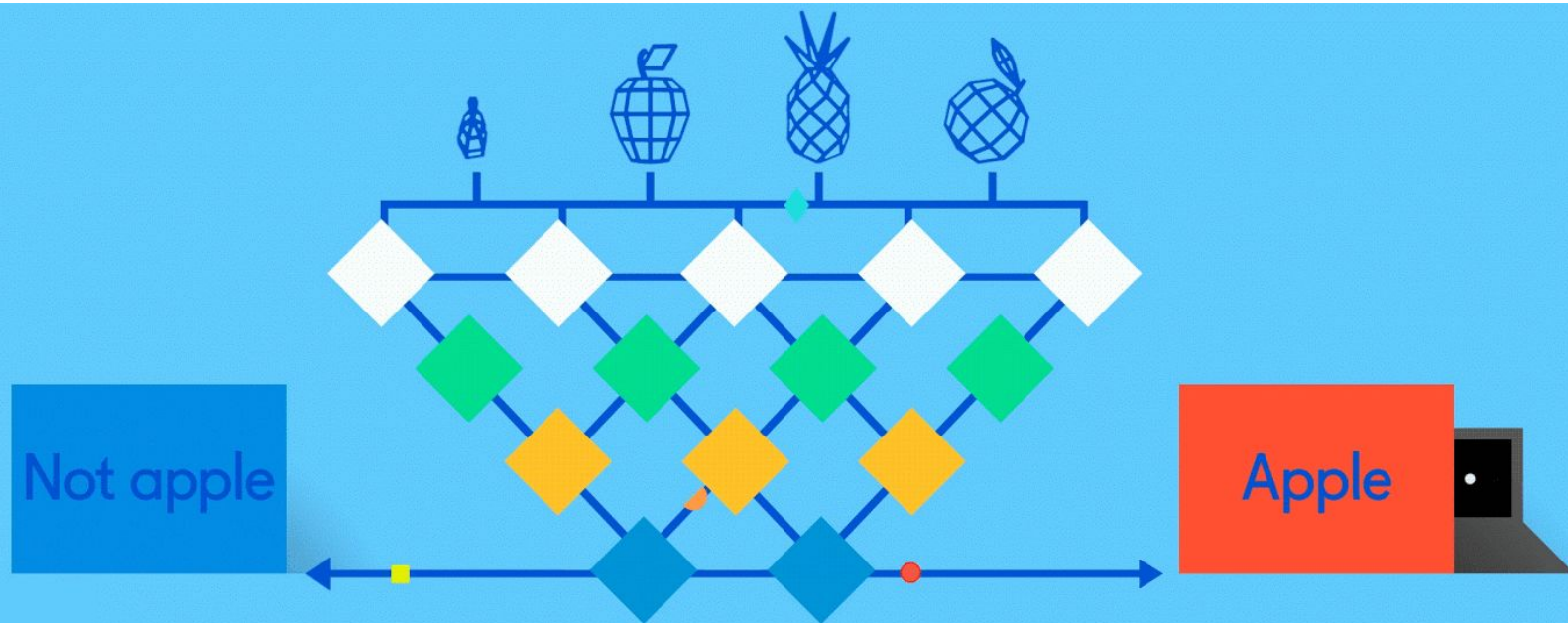
Treinamento



Treinamento



MLP

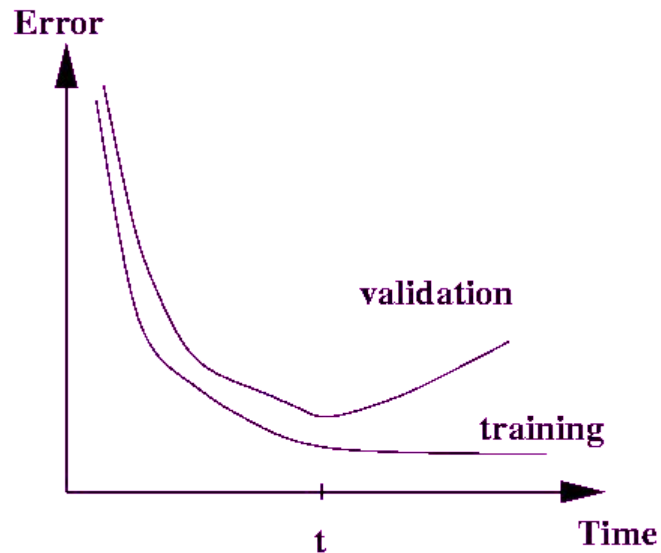


Critérios de Parada

- Diferentes critérios podem ser usados:
 - Número máximo de ciclos
 - Taxa máxima de erro
- Early stop: estratégia para evitar overfitting
 - Separa parte dos dados de treinamento para validação
 - Dados de validação são apresentados à rede a cada l ciclos
 - Treinamento é finalizado quando erro em validação começa a aumentar

Critérios de Parada

Early stop



Projeto de Arquitetura da Rede

- Escolhas de função de ativação e topologia da rede
 - Número de camadas e neurônios, padrão das conexões
 - Geralmente empíricas (tentativa e erro)
 - Problema de busca

Observações

- Atributos devem ser numéricos
 - Categóricos devem ser pré-processados
- É necessário normalizar os dados
 - Similar a k-NN
 - Relacionado também a ter crescimento muito grande dos valores de peso

Vantagens

- Várias soluções de sucesso em problemas práticos (principalmente percepção e controle)
- Tolerância a falhas e ruídos

Desvantagens

- Dificuldade de interpretação do modelo gerado (caixas-pretas)
- Escolha de melhor conjunto de parâmetros

Considerações

- Mínimos locais: solução estável que não é a melhor solução
 - Incidência pode ser reduzida
 - Empregando taxa de aprendizado decrescente
 - Adicionando nós intermediários
 - Utilizando termo de momentum
- Backpropagation é muito lento em superfícies complexas
 - Utilizar métodos de segunda ordem
 - Outros algoritmos
 - Ex.: RPROP, Newton, etc.

Slides construídos com base no material fornecido pela autora do livro 'inteligência artificial: uma abordagem de aprendizado de máquina' (Faceli et al, 2021).