

# Introdução a reconhecimento de padrões

## Definição

O reconhecimento de padrões por computador é uma ferramenta no campo da inteligência de máquina. Este tipo de tecnologia está presente em Visão Computacional, reconhecimento de faces, reconhecimento de caracteres impressos e manuscritos, classificação de imagens e dentre outros.

## Classificadores

### 1 Amostras

São informações que podem ser utilizadas para a etapa de treino ou para a avaliação dos classificadores. As amostras são dados que o algoritmo precisa interpretar para gerar a classificação.

### 2 Características

São os atributos utilizados para representar as amostras. As características podem ser representadas de diversas formas, por exemplo: quantitativas, qualitativas ou em forma de palavras. No caso de um carro, as características podem ser o modelo, tração, marca e dentre outros.

### 3 Classes

São as categorias das amostras. No caso de um texto, a categoria pode ser política, esporte, lazer e moda.

### 4 Conjunto de Treino

Representa uma parte da amostra separada para treinar o algoritmo para estimar uma solução para o problema.

### 5 Conjunto de Teste

Representa uma parte da amostra separada para validar o algoritmo. Vale destacar que o conjunto de treino e teste devem ser distintos. É comum encontrar na literatura abordagens que indicam 70% a 80% dos dados para o treino e o restante dos dados para o teste.

### 6 Tipos de aprendizagem

#### 6.1 Aprendizado Supervisionado

O algoritmo aprende a desenvolver determinadas tarefas por meio de resultados conhecidos. Para isso, a categoria dos dados devem ser definidas e passadas para o algoritmo. Por exemplo, para classificar um veículo por imagem é necessário que o algoritmo desenvolva um modelo com o auxílio da categoria, neste caso se é carro, moto, ônibus, dentre outros.

#### 6.2 Aprendizado não Supervisionado

Busca identificar informações semelhantes ou ausentes nos dados desconhecidos para gerar um modelo de classificação. Por exemplo, para classificar um veículo o algoritmo busca encontrar padrões para agrupar os veículos semelhantes, porém desconhece se o veículo é carro, moto, ônibus, dentre outros.

### 6.3 Aprendizado Reforçado

Neste caso, os dados para o treino são conhecidos e a classificação é realizada em uma amostra dos dados. Com base no resultado da classificação, um custo é gerado e deve ser baixo ou zero. O treino é refeito caso o valor do custo não seja satisfatório.

## 7. Tipos de classificadores

### 7.1 K-Nearest Neighbors (KNN)

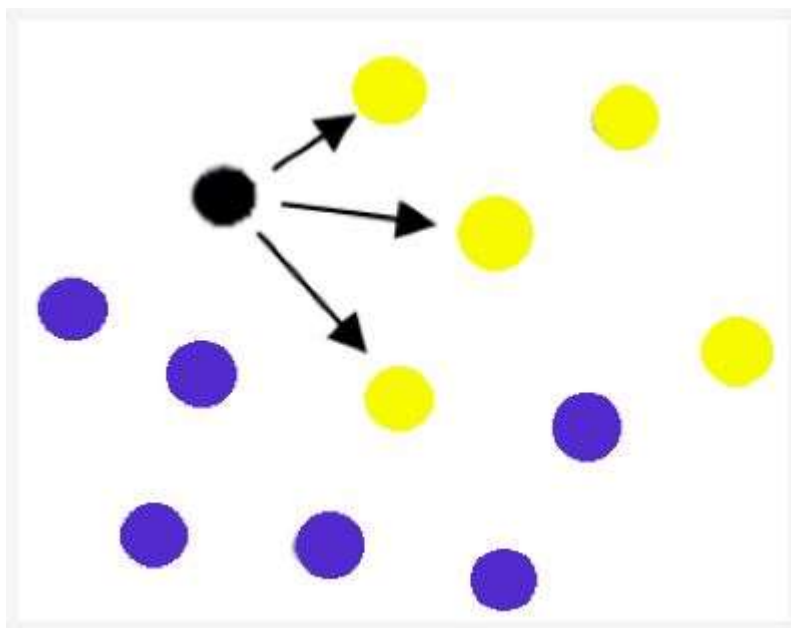
O KNN é o classificador mais simples dentre as técnicas de aprendizagem. O KNN adota que todas as amostras são representadas por pontos em um espaço  $n$ -dimensional, onde  $n$  é o número de descritores usados para representar as amostras. Para classificar, o KNN determina a vizinhança da nova amostra. Neste caso, os pontos definidos mais próximos. Para calcular a proximidade das amostras utiliza-se medidas de distância. A distância euclidiana é a mais utilizada. Considerando uma instância  $x$  descrita por um vetor de características

$$(a_1(x), a_2(x), \dots, a_n(x))$$

onde  $a_r(x)$  define o valor do  $r$ -ésimo atributo da instância  $x$ . Então a distância euclidiana entre duas instâncias  $x_i$  e  $x_j$  é definida por  $d(x_i, x_j)$ .

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2}$$

Concluído o cálculo da distância entre a nova amostra e as outras já definidas, o algoritmo classifica a nova amostra como sendo pertencente à classe a qual ela tenha  $k$  vizinhos mais próximos. Então se definirmos um  $k = 3$  e a nova amostra possuir um vizinho próximo da classe 0, mas dentre os quatro mais próximos os outros três forem da classe 1, a amostra será classificada como pertencente a classe 1.



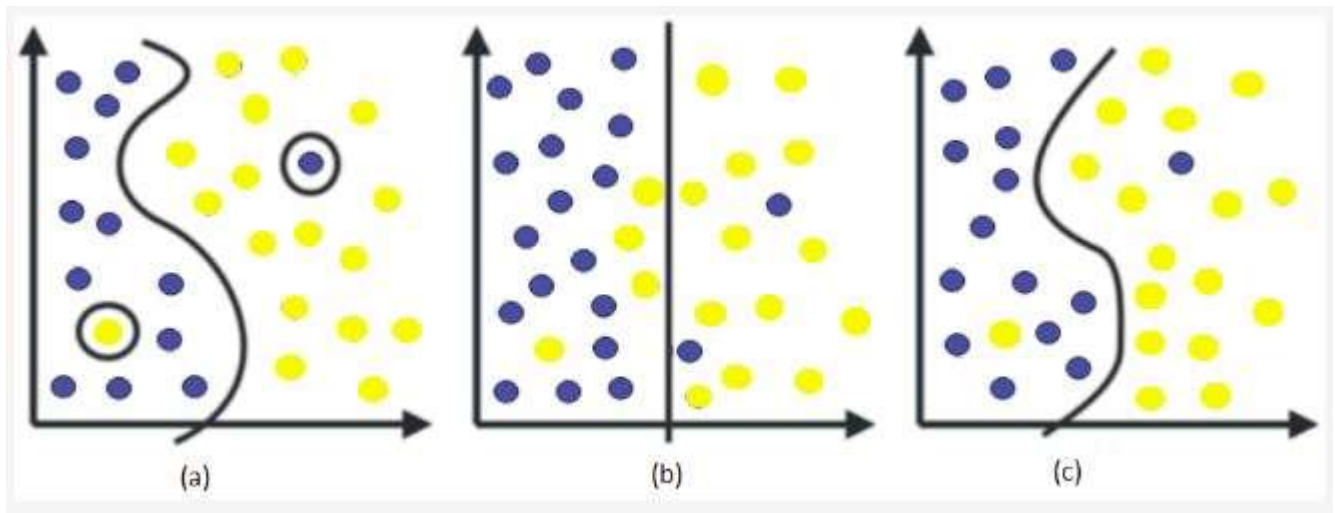
### 7.2 Multi-Layer Perceptron (MLP)

MLP é inspirado no sistema nervoso humano, neste caso as informações são processadas por meio de neurônios interconectados. MLP é um tipo de rede feed-forward, significa que a informação se propaga da entrada para a saída, passando por múltiplas camadas intermediárias. A Figura abaixo ilustra exemplos de diferentes possibilidades de classificação, em que:

(a) ilustra um caso de overfitting que classifica todas as amostras do conjunto, mas pode gerar erros em testes futuros;

(b) ilustra uma classificação linear que não satisfaz o problema e no caso;

(c) ilustra método balanceado que consegue generalizar a classificação;



Excluindo a camada de entrada, cada nó do MLP é um neurônio com uma função de ativação não-linear, que é treinada com a abordagem de retropropagação realizando uma otimização repetido dos pesos que ligam os neurônios minimizando a taxa média de erro quadrático da classificação. Após a rede ser submetida a um novo padrão, é gerada uma resposta que é comparada com a resposta esperada e se estiver errada o erro é estimado. Os valores do erro são transmitidos da saída para a entrada ajustando os pesos até que se obtenha a resposta desejada.

### 7.3 Stochastic Gradient Descent (SGD)

É uma abordagem muito eficiente para ajustar classificadores lineares, como máquinas de vetor de suporte (linear) e regressão logística. Embora o SGD esteja presente na comunidade de aprendizado de máquina há muito tempo, o SGD recebeu uma atenção considerável recentemente no contexto do aprendizado em larga escala.

O SGD foi aplicado com sucesso em problemas de aprendizado de máquina em larga escala, os mais comuns são encontrados na classificação de texto e no processamento de linguagem natural.

#### VANTAGENS

Eficiência;

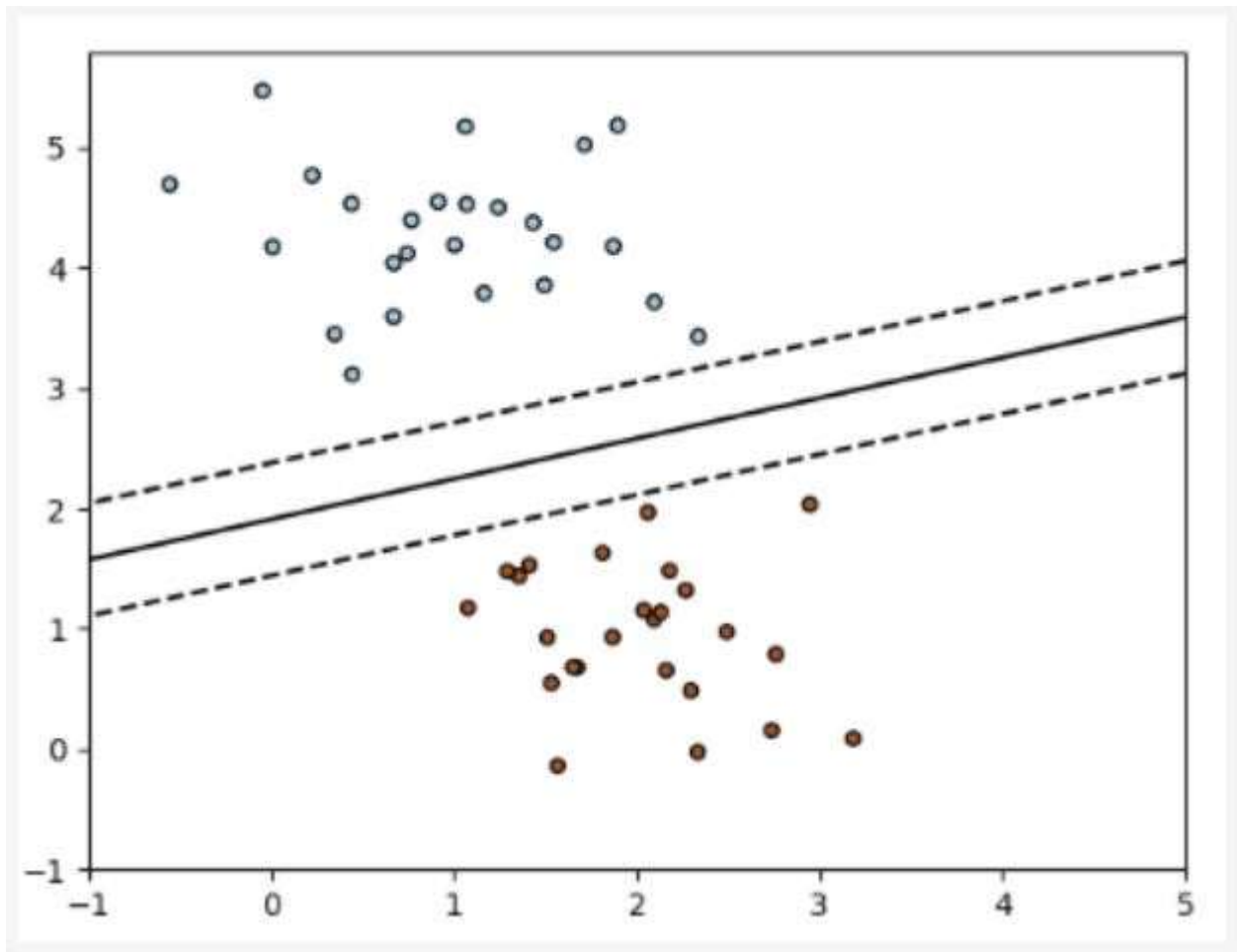
Facilidade de implementação (muitas oportunidades para ajuste de código);

#### DESVANTAGENS

Requer vários hiperparâmetros, como o parâmetro de regularização e o número de iterações;

Sensível ao dimensionamento de recursos;

Como outros classificadores, o SGD deve ser ajustado com dois arrays: um array  $X$  de forma  $(n\_samples, n\_features)$  contendo as amostras de treinamento e um array  $y$  de forma  $(n\_samples)$ , contendo os valores alvo (rótulos de classe) para as amostras de treinamento. A Figura abaixo ilustra a distribuição dos dados e as retas desenvolvidas pelo algoritmo para separar as regiões.



#### 7.4 Support Vector Machine (SVM)

No SVM um hiperplano linear é desenvolvido para separar as amostras positivas das negativas. Dois outros hiperplanos são criados de cada lado do primeiro, com o objetivo de maximizar a distância entre os dois novos hiperplanos. Portanto, os hiperplanos são utilizados para definir as regiões onde ocorre cada uma das classes. Sendo assim, quando o SVM recebe uma nova amostra a mesma é classificada de acordo com o seu posicionamento no plano em relação ao hiperplano divisor.

##### VANTAGENS

Eficaz em casos que o número de dimensões é maior que o número de amostras.

Diferentes funções do Kernel podem ser especificadas para a função de decisão;

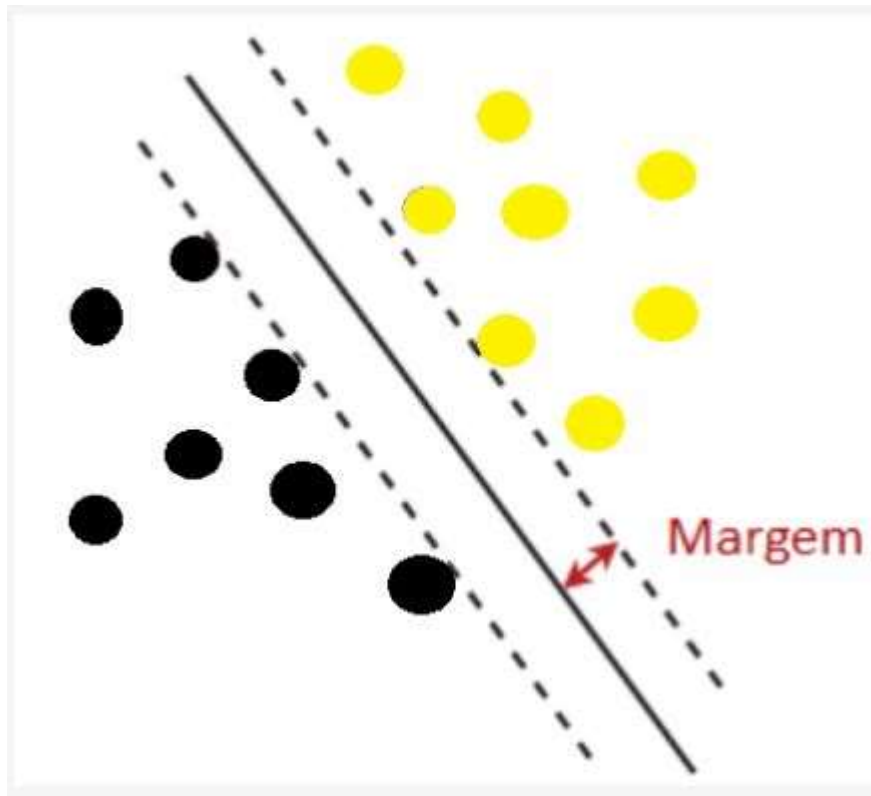
Existe a possibilidade de especificar kernels personalizados conforme a necessidade do problema;

##### DESVANTAGENS

É necessário definir um bom Kernel para a aplicação.

O tempo de treinamento pode ser bem longo dependendo do número de exemplos e dimensionalidade dos dados;

A Figura abaixo ilustra um exemplo dos hiperplanos lineares em um conjunto de dados bidimensionais. As linhas tracejadas representam os hiperplanos.

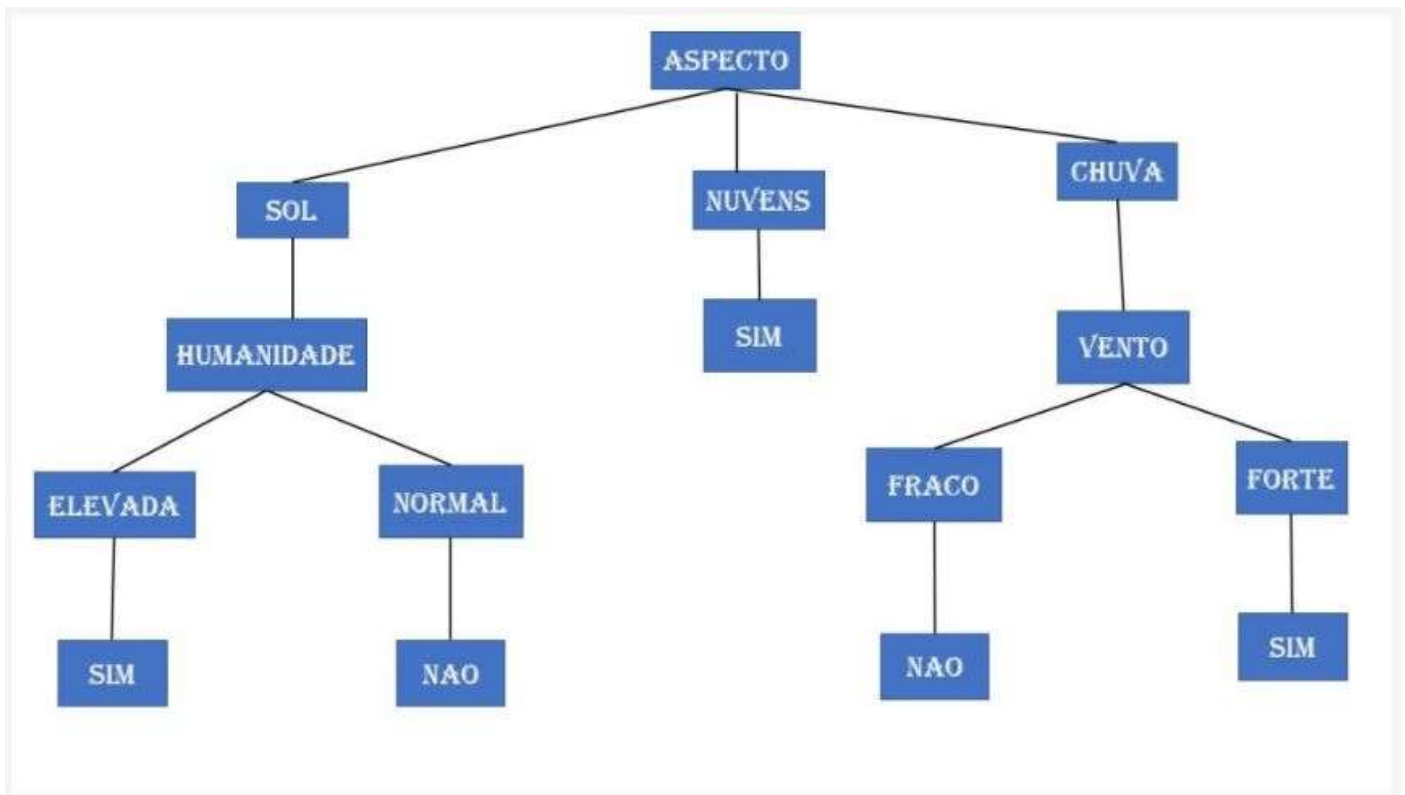


### 7.5 Decision tree

É uma ferramenta popular para classificação e previsão. Uma árvore de decisão é uma estrutura de árvore semelhante a um fluxograma, onde cada nó interno denota um teste em um atributo, cada ramo representa um resultado do teste e cada nó folha (nó terminal) contém um rótulo de classe.

Assim como um fluxograma, a árvore de decisão estabelece nós (decision nodes) que se relacionam entre si por uma hierarquia. Existe o nó-raiz (root node), que é o mais importante, e os nós-folha (leaf nodes), que são os resultados finais. No contexto de machine learning, o raiz é um dos atributos da base de dados e o nó-folha é a classe ou o valor que será gerado como resposta.

Na ligação entre nós, temos regras de “se-então”. Ao chegar a um nó A, o algoritmo se pergunta acerca de uma regra, uma condição, como “se a característica X do registro analisado é menor do que 15?”. Se for menor, então o algoritmo vai para um lado da árvore; se for maior, então ele vai para outro. No próximo nó, segue a mesma lógica. A figura abaixo ilustra um exemplo de um fluxo de uma árvore de decisão.



## 7.6 Naive Bayes

O Naïve Bayes é um classificador probabilístico que define a classificação pela probabilidade posteriori de uma nova amostra pertencer a uma das classes conhecidas. Considere um vetor de característica  $v$ , a probabilidade  $P(c_i|v)$  da classe  $c_i$ , é determinada utilizando o teorema de Bayes, definido como:

$$P(c_i|\vec{v}) = \frac{P(\vec{v}|c_i)P(c_i)}{P(\vec{v})}$$

Considerando que a probabilidade de ambas as classes são iguais,  $P(c_i) = 0,5$  tanto para  $c_0$  quanto para  $c_1$ . Para calcular  $P(v|c_i)$ , considere que as amostras possuem uma função de densidade gaussiana, definida da seguinte forma:

$$P(x) = \frac{1}{\sigma\sqrt{2}} e^{-(x-u)^2/2\sigma^2}$$

A eficácia do naive bayes depende muito da distribuição das informações extraídas pelos descritores, pois o classificador elabora sua função de decisão a partir da suposição de que os dados possuem distribuição normal, portanto quanto mais próxima a distribuição for da normal, melhor o desempenho do classificador.

## 7.7 Random forrest

É um método de aprendizado conjunto. Random forrest busca treinar várias árvores de decisões (descorrelacionadas) obtidas a partir de amostras do dataset, e fazer previsões utilizando os resultados que mais aparecem em caso de um problema de classificação, ou a média dos valores obtidos em caso de regressão.

Random forrest estabelece o resultado da classificação com base nas previsões das árvores de decisão. Neste caso, a previsão é tomada como a média ou a média da saída de várias árvores. Aumentar o número de árvores aumenta a precisão do resultado.

Características:

Mais preciso que o algoritmo da árvore de decisão;

Fornecer uma maneira eficaz de lidar com dados ausentes;

Pode produzir uma previsão razoável sem ajuste de hiperparâmetro;

Resolve o problema de overfitting em árvores de decisão;

Em cada árvore de floresta aleatória, um subconjunto de feições é selecionado aleatoriamente no ponto de divisão do nó;

## **Aplicações**

Reconhecimento de faces e de voz;

Classificação de imagens;

Identificação de impressão digital;