

# ALGORITMO K-NN

Bianca Lara Gomes  
biancalara@gmail.com

## Introdução

O algoritmo de classificação baseado no vizinho mais próximo (Nearest Neighbor - NN) é uma técnica amplamente empregada para reconhecer padrões. O centro de seu funcionamento está em descobrir o vizinho mais próximo de uma dada instância. O algoritmo (K-Nearest Neighbor – K-NN) pertence a um grupo de técnicas denominada de Instance-based Learning. Nesse caso, são encontrados os K vizinhos mais próximos do padrão de consulta, ao invés de apenas o vizinho mais próximo. Dito dessa forma, esse método é bastante simples e de fácil implementação.

## 1. DESCRIÇÃO DO ALGORITMO

Usado para classificar objetos com base em exemplos de treinamento que estão mais próximos no espaço de características. Para utilizar o KNN é necessário um conjunto de exemplos de treinamento, definir uma métrica para calcular a distância entre os exemplos de treinamento, definir o valor de k.

O KNN classifica um dado elemento de acordo com as respectivas classes dos  $k(k \geq 1)$  vizinhos mais próximos – pertencentes a uma base de treinamentos dada. O algoritmo calcula a distância do elemento dado para cada elemento da base de treinamento e então ordenar os elementos da base de treinamento do mais próximo ao de maior distância. Dos elementos ordenados selecionam-se apenas os K primeiros, que servem de parâmetro para a regra de classificação. Exemplo: 1-NN é um K-NN sabendo que  $k = 1$ , ou seja, seleciona-se apenas o elemento do treinamento mais próximo da instância que se pretende classificar. 5-NN vai usar os cinco elementos mais próximos da instância e baseado nas classes dos cinco elementos, infere-se a classe do elemento de teste. Dois pontos importantes no K-NN são: a regra de classificação e função que calcula a distância entre duas instâncias. A regra de classificação diz como o algoritmo vai tratar a importância de cada um dos k elementos selecionados – os k mais próximos. À função de distância cabe a tarefa de mensurar a disparidade entre dois elementos de forma a poder identificar quais são os K-NN. Se tratando de distância entre dois elementos é preciso definir o que são esses elementos: cada elemento tem uma classe – usada na regra de

classificação – e um vetor de atributos, e cada um desses atributos tem um tipo que indica quais operações se pode realizar sobre o atributo. E é por meio dessas operações que se torna possível o cálculo das distâncias.

Pseudocódigo do algoritmo

### Inicialização:

Preparar conjunto de dados de entrada e saída

Informar o valor de k;

### Para cada nova amostra faça

Calcular distância para todas as amostras

Determinar o conjunto das k's distâncias mais próximas

O rótulo com mais representantes no conjunto dos k's vizinhos será o escolhido

### Fim para

**Retornar:** conjunto de rótulos de classificação

## 2. APLICAÇÕES

- Classificação automática de testes;
- Imagens multispectrais para estimativa de parâmetros florestais;
- Seleção de características e classificação de abelhas.

## 3. CONCLUSÃO

As funções de distância são o maior viés de estudo no KNN, pois possibilitam as maiores variações nos resultados obtidos por esse algoritmo, podendo inclusive obter resultados em nível de estado da arte nessa área.

Vantagens: Técnica simples e facilmente implementada. Bastante flexível. Em alguns casos apresenta ótimos resultados.

Desvantagens: Classificar um exemplo desconhecido pode ser um processo computacionalmente complexo. Requer um cálculo de distância para cada exemplo de treinamento. Pode consumir muito tempo quando o conjunto de treinamento é muito grande. A precisão da classificação pode ser severamente degradada pela presença de ruído ou características irrelevantes.

#### 4. REFERÊNCIAS

ALGORITMOS DE APRENDIZAGEM DE MÁQUINA. Disponível em:

<<https://www.wattpad.com/115821129-algoritmos-de-aprendizagem-de-m%C3%A1quina-knn-nearest>>. Acesso em: 05 maio 2018.

INF 171 – INTELIGENCIA ARTIFICIAL.

Disponível em:

<[http://edirlei.3dgb.com.br/aulas/ia\\_2012\\_1/IA\\_Aula\\_16\\_KNN.pdf](http://edirlei.3dgb.com.br/aulas/ia_2012_1/IA_Aula_16_KNN.pdf)>. Acesso em: 05 maio 2018.