



INF 1771 – Inteligência Artificial


Aula 13 – K-Nearest Neighbor (KNN)

Edirlei Soares de Lima
<elima@inf.puc-rio.br>

Formas de Aprendizado

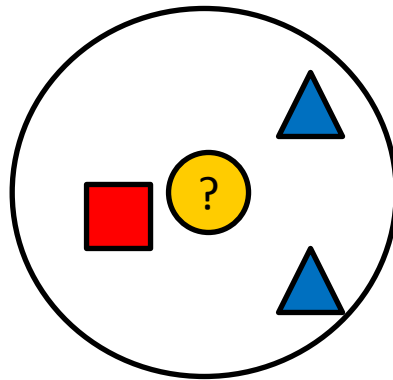
- **Aprendizado Supervisionado**
 - Árvores de Decisão.
 - **K-Nearest Neighbor (KNN).**
 - Support Vector Machines (SVM).
 - Redes Neurais.
- Aprendizado Não Supervisionado
- Aprendizado Por Reforço

Aprendizado Supervisionado

- Observa-se alguns pares de **exemplos de entrada e saída**, de forma a aprender uma **função que mapeia a entrada para a saída**.
 - Damos ao sistema a **resposta correta** durante o processo de treinamento.
 - É eficiente pois o sistema pode trabalhar diretamente com informações corretas.
- 

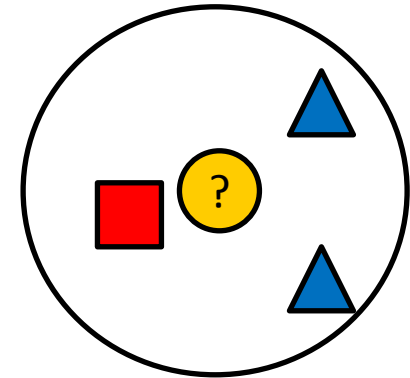
K-Nearest Neighbor

- É um dos algoritmos de classificação clássicos e bem **simples**.
- Usado para classificar objetos com base em **exemplos de treinamento** que estão mais próximos no espaço de características.



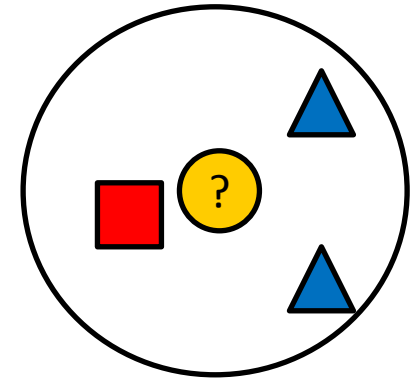
K-Nearest Neighbor

- Para utilizar o KNN é necessário:
 - **(1)** Um conjunto de exemplos de treinamento.
 - **(2)** Definir uma métrica para calcular a distância entre os exemplos de treinamento.
 - **(3)** Definir o valor de K (o número de vizinhos mais próximos que serão considerados pelo algoritmo).

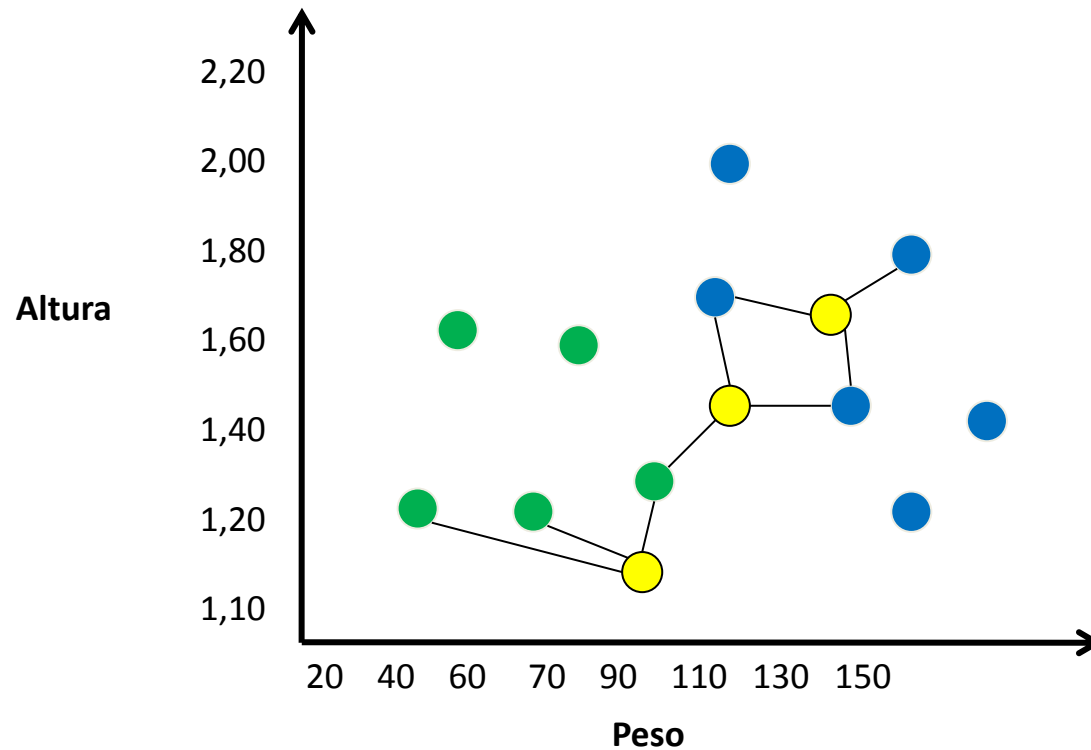


K-Nearest Neighbor

- Classificar um exemplo desconhecido com o algoritmo KNN consiste em:
 - **(1)** Calcular a distância entre o exemplo desconhecido e o outros exemplos do conjunto de treinamento.
 - **(2)** Identificar os K vizinhos mais próximos.
 - **(3)** Utilizar o rótulo da classe dos vizinhos mais próximos para determinar o rótulo de classe do exemplo desconhecido (votação majoritária).



Espaço de Características




K-Nearest Neighbor

- Calculando a distancia entre dois pontos:
 - Existem varias formas diferentes de calcular essa distancia. A mais simples é a distancia euclidiana:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

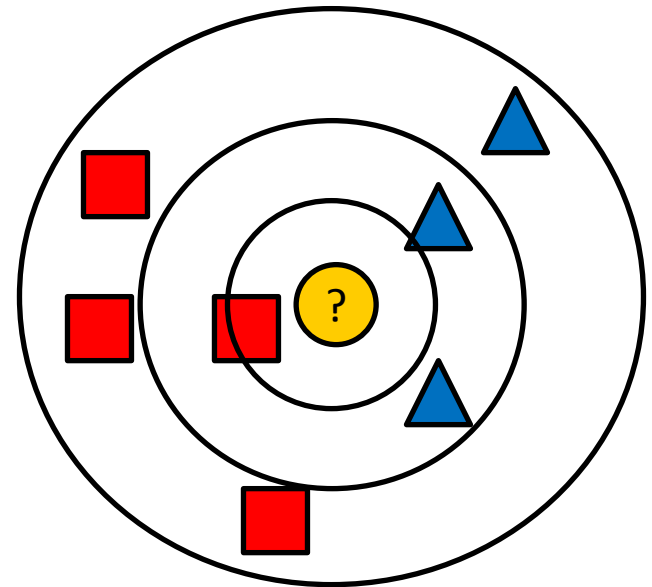
- É importante normalizar os dados.
- Outras formas de medir a distancia:
 - Distância de Mahalanobis.
 - Distância de Minkowsky.
 - Hamming Distance.
 - ...

K-Nearest Neighbor


- Determinando a classe do exemplo desconhecido a partir da lista de vizinhos mais próximos:
 - Considera-se o voto majoritário entre os rótulos de classe dos K vizinhos mais próximos.
 - Como escolher o valor de K?
- 

K-Nearest Neighbor

- **K = 1**
 - Pertence a classe de quadrados.
- **K = 3**
 - Pertence a classe de triângulos.
- **K = 7**
 - Pertence a classe de quadrados.



K-Nearest Neighbor

- Como escolher o valor de K?
 - Se K for muito pequeno, a classificação fica sensível a pontos de ruído.
 - Se k é muito grande, a vizinhança pode incluir elementos de outras classes.
 - Além disso, é necessário sempre escolher um valor ímpar para K, assim se evita empates na votação.
- 

K-Nearest Neighbor

- A **precisão** da classificação utilizando o algoritmo KNN depende fortemente do modelo de dados.
- Na maioria das vezes os atributos precisam ser **normalizados** para evitar que as medidas de distância sejam dominado por um único atributo. Exemplos:
 - Altura de uma pessoa pode variar de 1,20 a 2,10.
 - Peso de uma pessoa pode variar de 40 kg a 150 kg.
 - O salário de uma pessoa podem variar de R\$ 800 a R\$ 20.000.

K-Nearest Neighbor

- **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.

Leitura Complementar

- Mitchell, T. **Machine Learning**, McGraw–Hill Science/Engineering/Math, 1997.
- Duda, R. Hart, P. Stork, D. **Pattern Classification**, John Wiley & Sons, 2000

