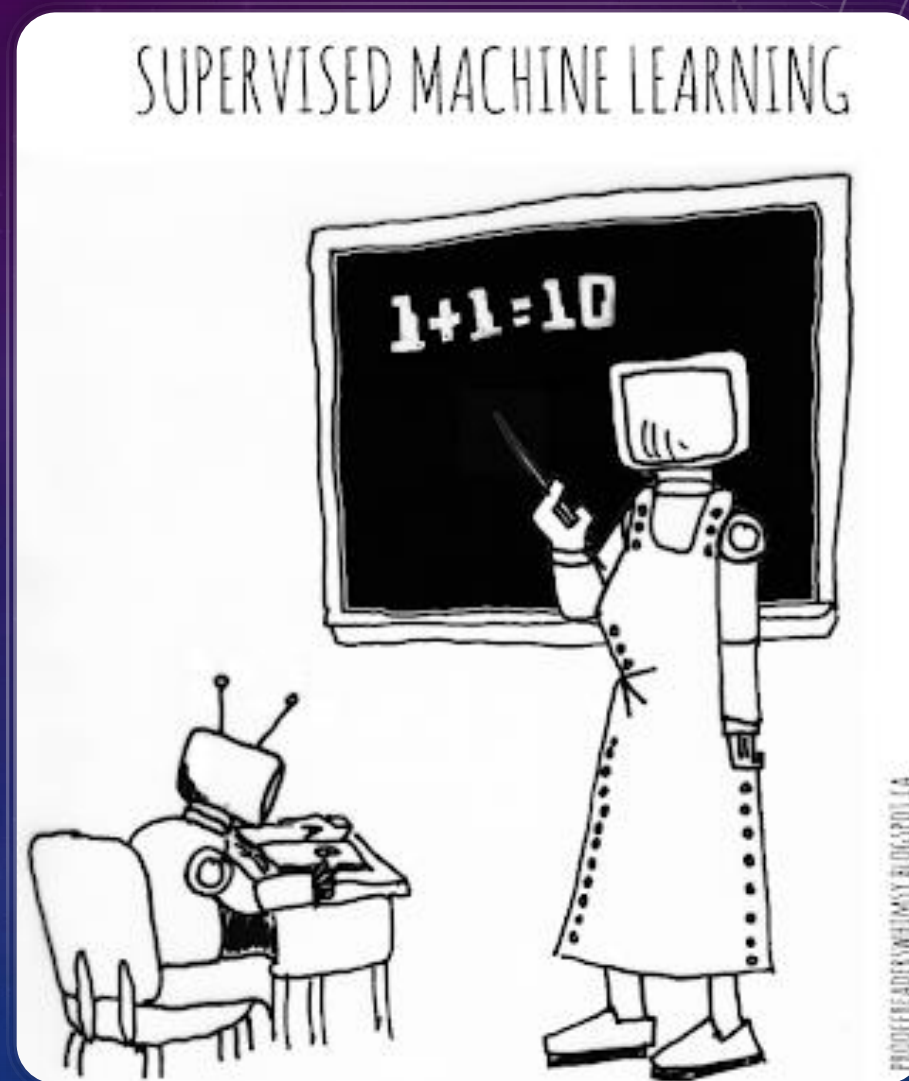




K-MEANS LUCAS MACEDO DA SILVA

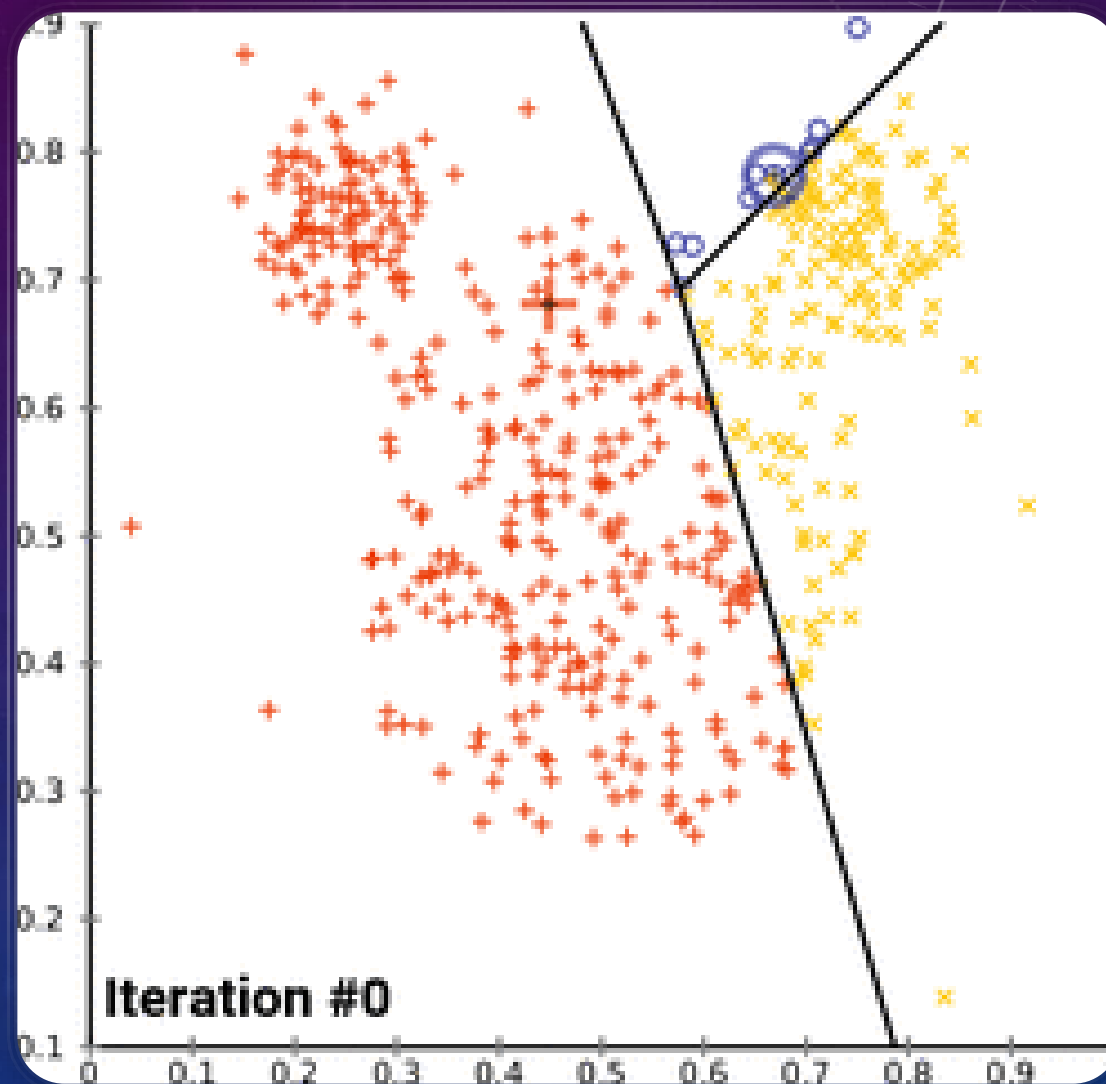
1. INTRODUÇÃO

- Algoritmo de aprendizagem não supervisionada.



2. DEFINIÇÃO

- “É uma abordagem simples e elegante para particionar um *dataset* em k cluster. Sem sobreposição”



3. K-MEANS

- A ideia por trás do k-means é que um bom cluster é aquele que no qual a variação dentro do cluster é a menor.

4. ALGORITMO

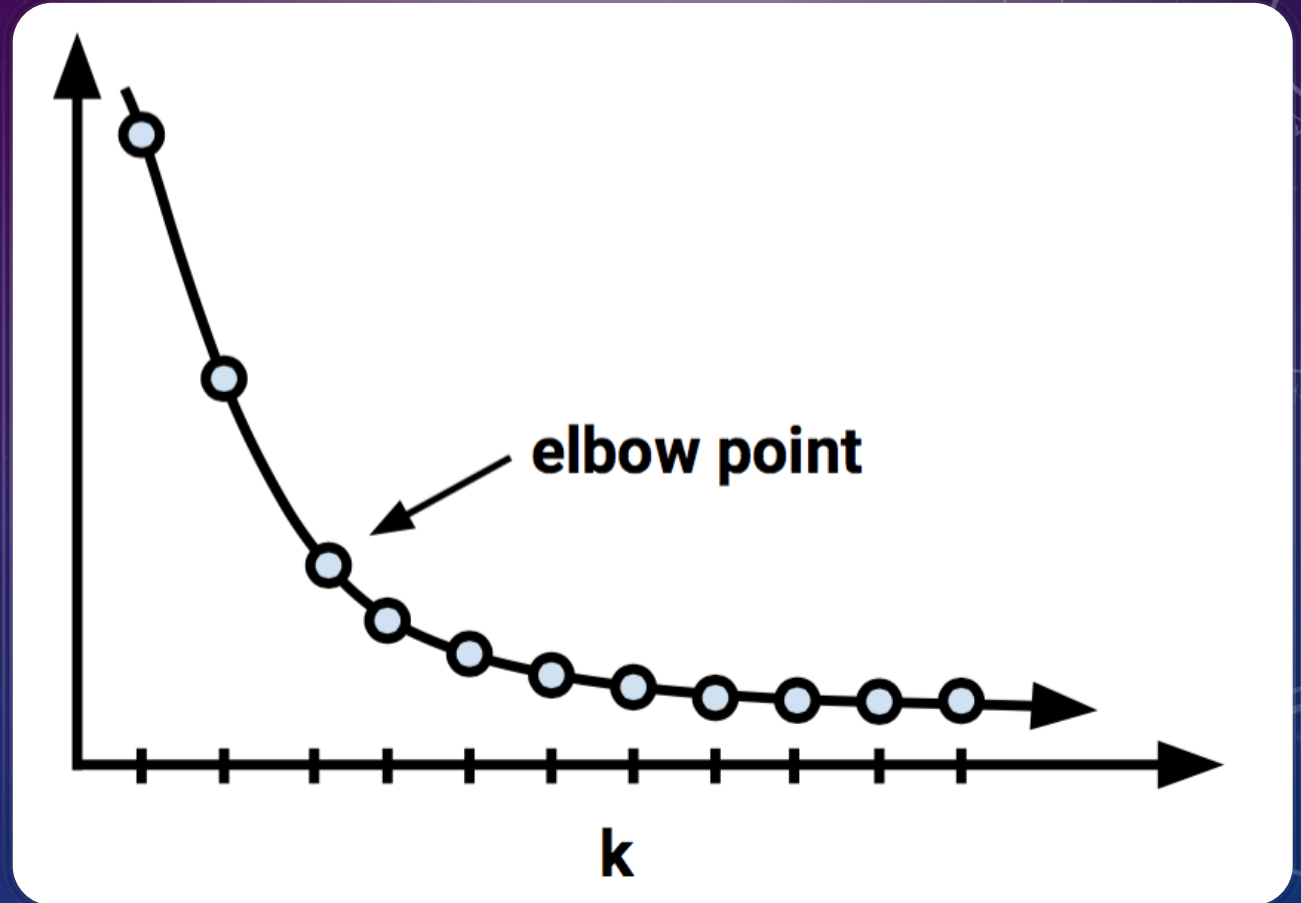
- 1. Atribuir um número (cluster) para cada um dos dados.
- 2. Itere sobre os dados até que as atribuições do cluster pare de mudar:
 - a. Para cada um dos k cluster calcule a centroide. O k -ésimo centroide do cluster é o vetor da característica p para as observações no k -ésimo cluster.
 - b. Atribua cada observação ao cluster cujo centróide está mais próximo.

5. EXEMPLO



6. COMO DETERMINAR O MELHOR NÚMERO DE CLUSTERS?

- Encontrar um valor de k que a distância dos dados e dos centroides não mude.



7. REFERÊNCIAS BIBLIOGRÁFICAS

JAMES, Gareth et al. **An Introduction to Statistical Learning with Applications in R**. New York: Springer, 2013.