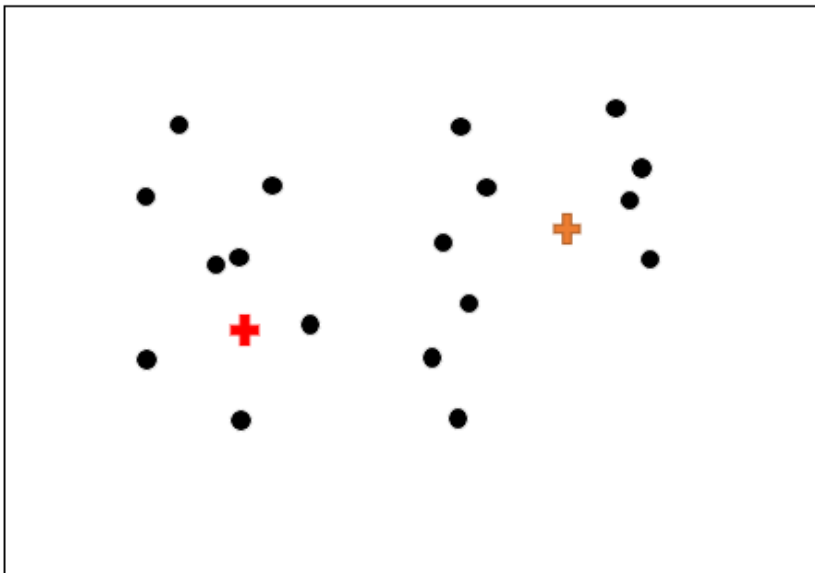


# K-means

---

- Dados sem classe (não supervisionado)
- Agrupamento de forma orgânica (de acordo com algoritmo e seus parâmetros)

# Algoritmo de Agrupamento



- Definir  $K = 2$
- Definir dois centroides
- Cada Instancia busca o centroide mais próximo
- Os centroides são atualizados baseados na media dos pontos do seu grupo
- O processo é repetido até um critério de parada

# Atividade

---

- Agrupar os 4 atributos numéricos usando k-means
- Comparar os grupos com a classe
- Avaliar a performance
- Precisamos alguns transformações:
  - Transformar os 4 atributos em um vetor
  - Transformar a classe em um índice e transformar para inteiro

sepalength	sepalwidth	petallength	petalwidth	class
5.1	3.5	1.4	0.2	Iris-setosa
4.9	3.0	1.4	0.2	Iris-setosa
4.7	3.2	1.3	0.2	Iris-setosa
4.6	3.1	1.5	0.2	Iris-setosa
5.0	3.6	1.4	0.2	Iris-setosa

# Hiper Parâmetros

- distanceMeasure: métrica de distância. Opções Euclidean e cosine. (padrão: euclidean)
- k: Números de cluster a serem criados. (padrão: 2)
- maxIter: número máximo de iteração. (padrão: 20)
- predictionCol: nome da coluna com o cluster. (padrão: prediction)