Insper

Machine Learning

Clustering

Fábio Ayres <fabioja@insper.edu.br>

Aprendizado supervisionado

- Temos variáveis independentes X e a variável dependente y
- Objetivo: Construir modelo preditivo

$$\hat{y} = h(x, \theta)$$

$\hat{y}$: estimativa

θ: parâmetros do modelo

- Para construir o modelo (aprender os parametros)
  - Define uma função de perda
  - Aplica um algoritmo de otimização

Aprendizado não-supervisionado

- Não temos variavel dependente ← não tem classe verdadeira!
- Objetivo: análise exploratória para obter insight

Exemplo: clustering

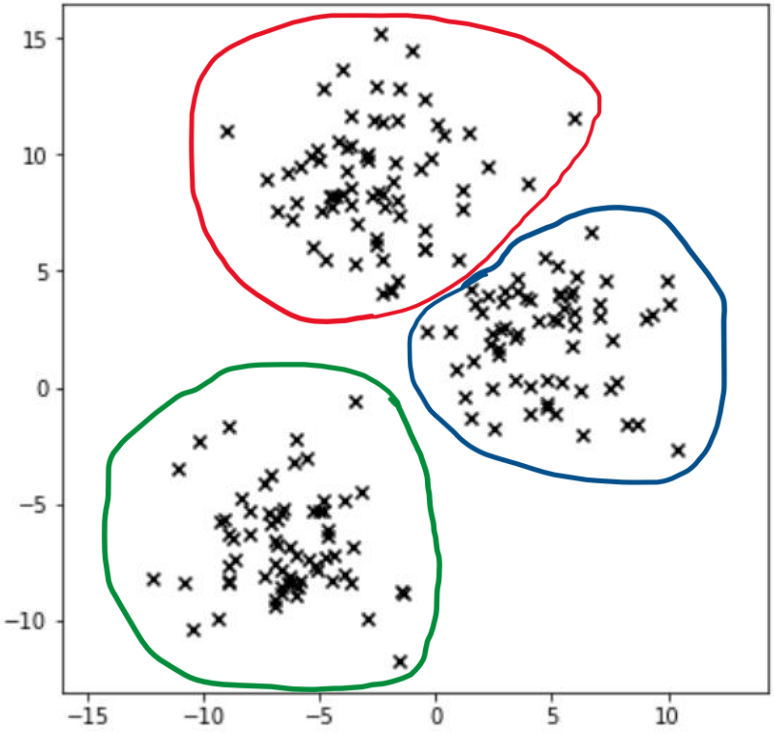
$$\hat{y} = h(x, \theta)$$

$\hat{y}$: cluster

θ: parametros

- Ao invés de otimizar uma "função de perda" (pois não tem y), otimizar algum critério de "qualidade do clustering".

Quais os agrupamentos naturais?



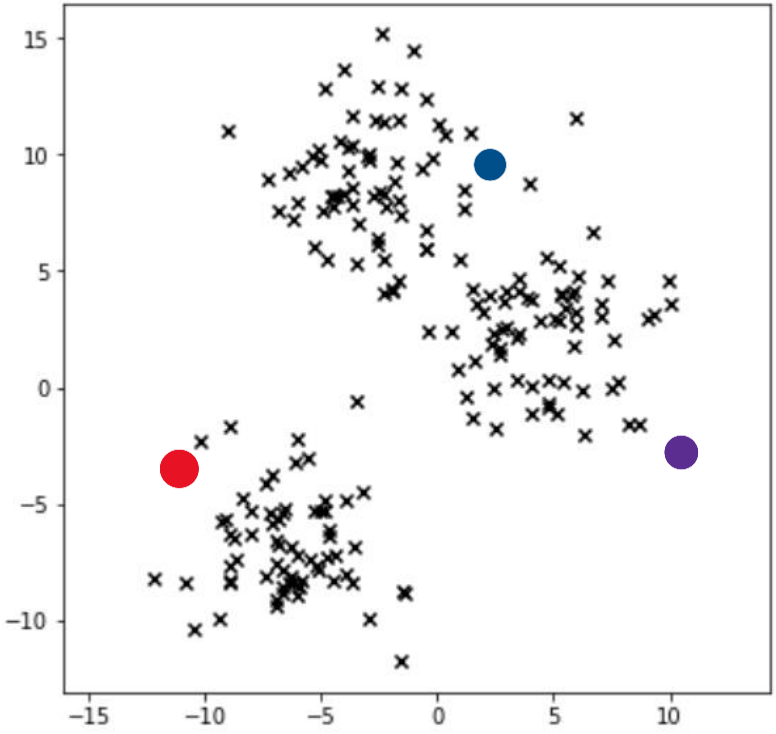
Clustering

- k-means
- Mean Shift
- Clustering Hierarquico

K-means

1. Definir quantos clusters queremos
   k = nº clusters

2. Inicialização:
   Sorteia k pontos como "centroides" de cluster



K-means

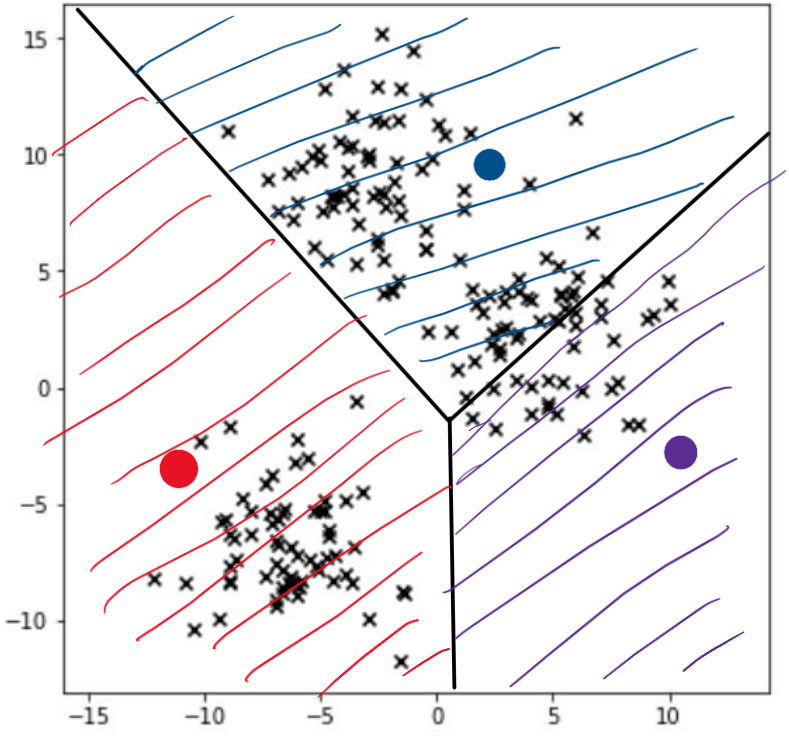
- Associa cada ponto ao centroide mais próximo

mapa de Voronoi

dual

triangulação de Delaunay

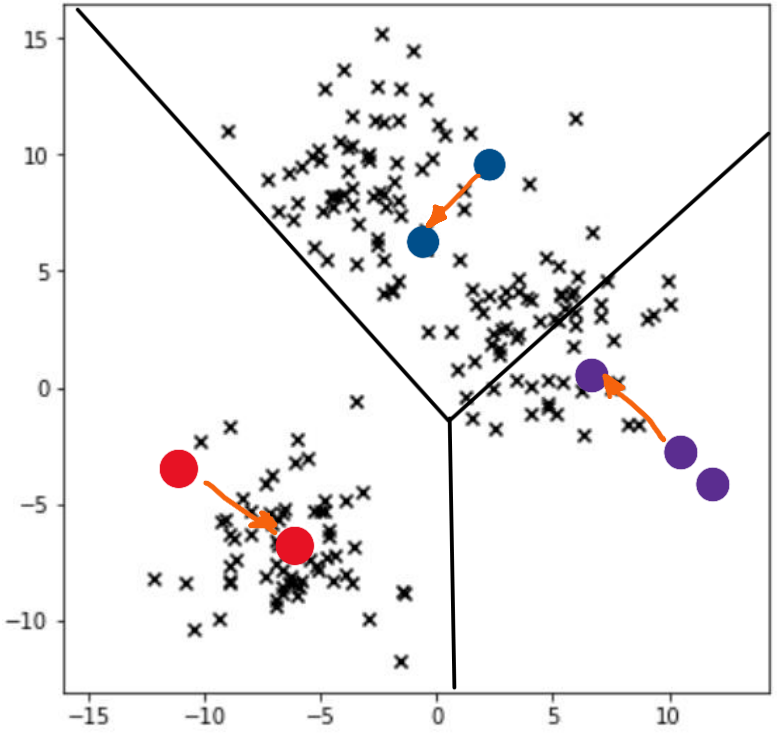
Computational Geometry



K-means

- Recalcula os centroides

- Repetir até convergencia:
  - Associar pontos
  - Recalcular centroide



K-means

Vantagens

- Simples
- Só depende da definição de distância entre pontos
- Escalável

Desvantagens

- Tem que definir a priori o número de clusters.
- Sensível ao chute inicial dos centroides.
- O que mais?

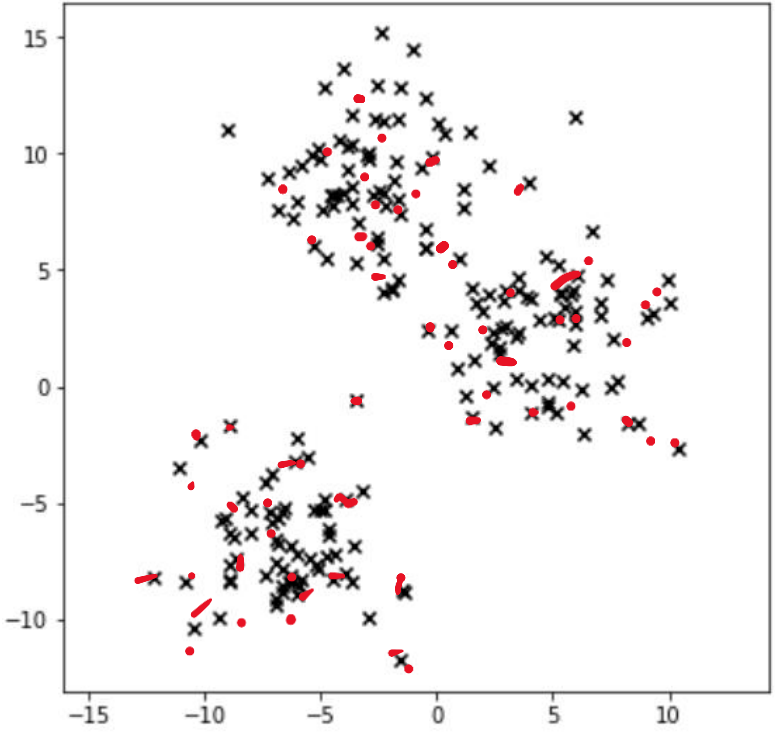
Mean-shift clustering

Inicialização:

Escolhe varios pontos do dataset como "sementes"

Loop:

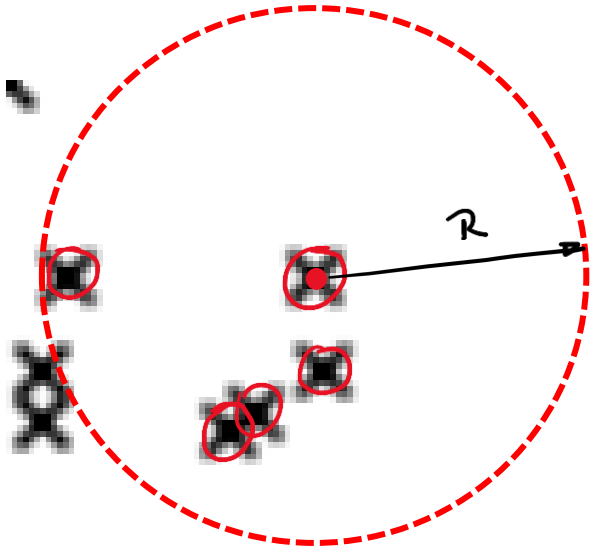
- Cada semente identifica seus novos "amigos" num raio R



Mean-shift clustering

Loop:

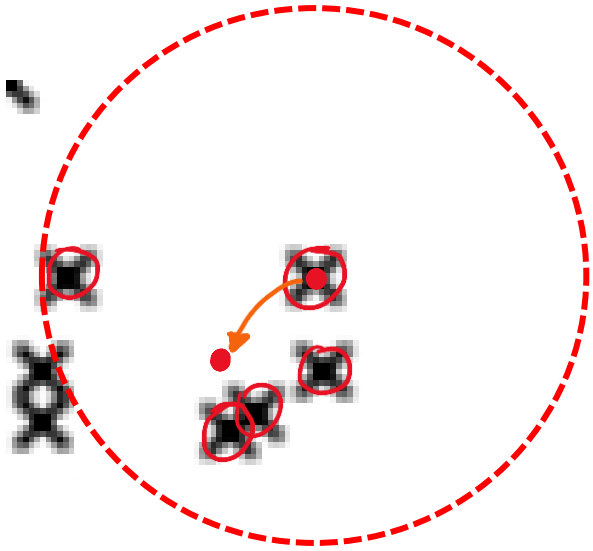
- Cada semente identifica seus novos "amigos" num raio R



Mean-shift clustering

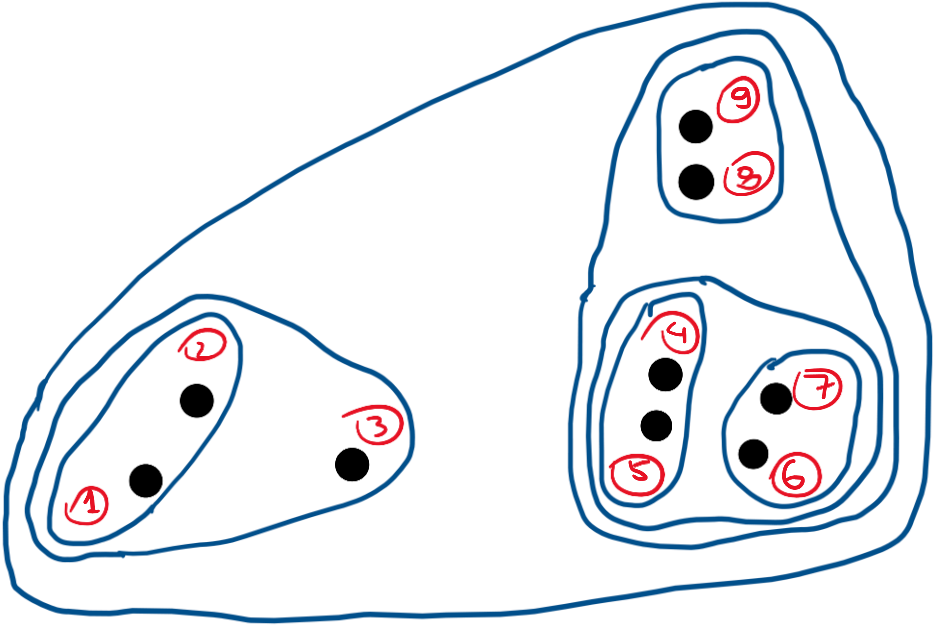
Loop:

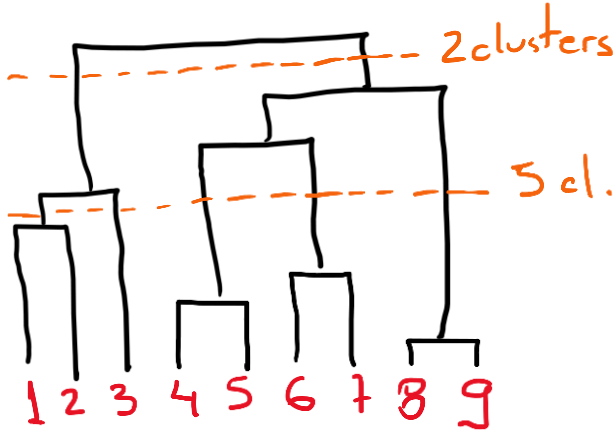
- "Corre abraçar os amigos!" Recalcula posição da semente como a média dos pontos selecionados
- Se duas sementes "se encontram" → fusão!



Repetir até convergir

Agglomerative clustering





Insper