

Clustering "formar grupos naturais"

↳ a gosto do fuguês

Aprendizado
supervisionado

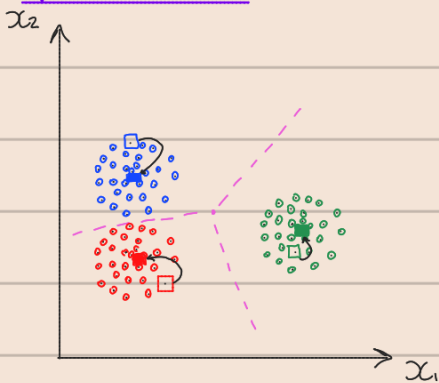
"existe resposta certa"

- tem X, y
- otimiza um erro objetivo
- modelo preditivo

Aprendizado
não supervisionado

- não tem o X
- não tem resposta diretamente correta
- detecção de padrões

K-means



1. Escolhe $K = n^{\circ}$ cluster
2. Cria K centróides colocando K pontos ao acaso
3. Enquanto "não convergiu":
 - 3.1. Associa cada ponto do dataset ao centróide mais próximo
 - 3.2. Atualizar cada centróide p/ posição média dos seus associados

Atenção!

Qualquer algoritmo que empregue a noção de distância ou a magnitude de features requer normalização

Exemplos: PCA, SVM, K-means, Ridge, LASSO, K-NN

Exemplo numérico: