

Aprendizado não supervisionado

Professor: Gabriel
Oliveira Assunção

Tipos de Tarefas

Aprendizado supervisionado

Classificação

- Detecção de fraude
- Diagnóstico de uma doença
- Classificação de SPAM

Regressão

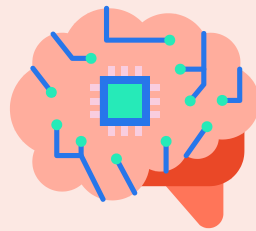
- Predição de score
- Preço de aluguel

Aprendizado não supervisionado

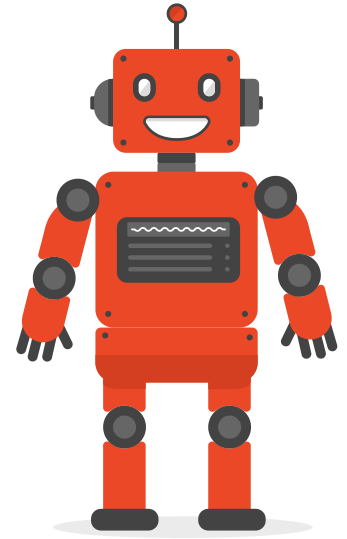
- Segregação de clientes
- Detecção de vídeo
- Detecção de áudio
- Análise de texto

Aprendizado por reforço

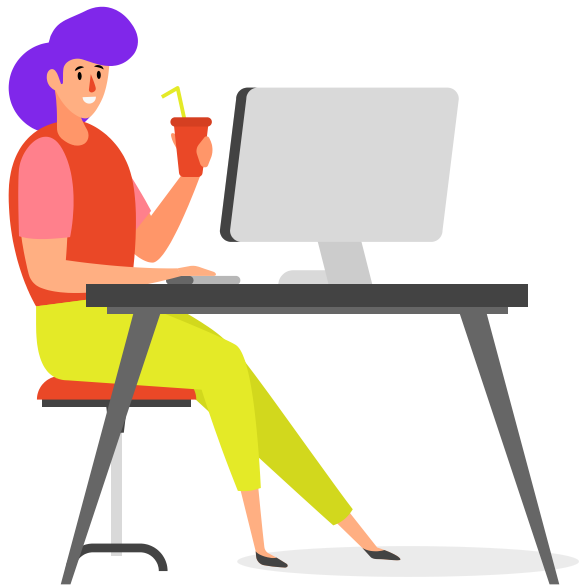
- Carro autonomo
- Reposição de estoque



Redução de dimensão



Componentes principais (PCA)

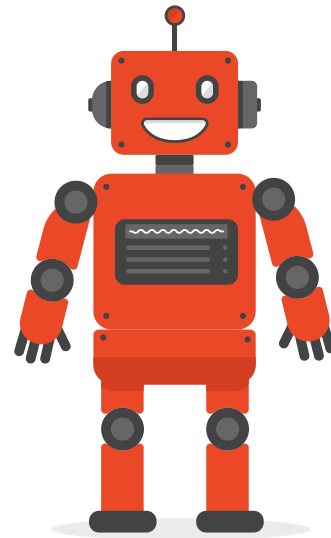


- Redução de dimensão para geração de features.
 - Visualização dos dados.
- Transformar os dados de forma que mantenha uma explicação da variabilidade.



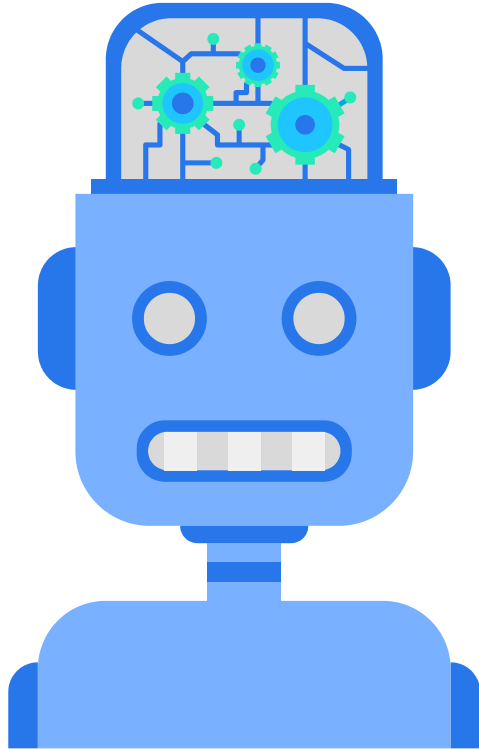
Hora da prática!!!!

Clustering



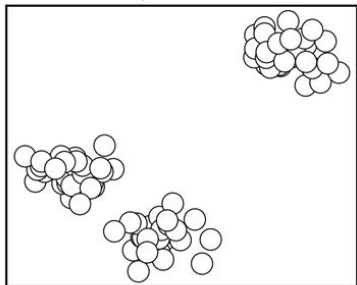


K-Means

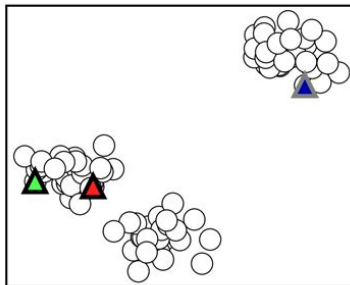


- Número de grupos (cluster) definido.
- Algoritmo define os pontos iniciais dos grupos.
- A cada iteração são definidos os grupos utilizando a proximidade
- Um ponto de centro é definido e é refeito os agrupamentos usando esse centro

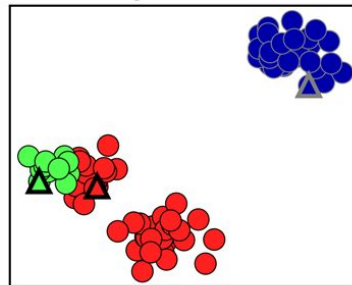
Input data



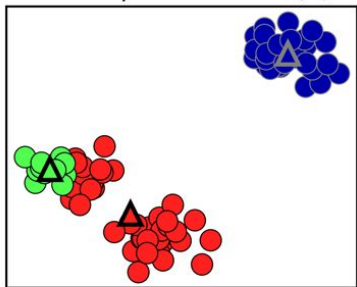
Initialization



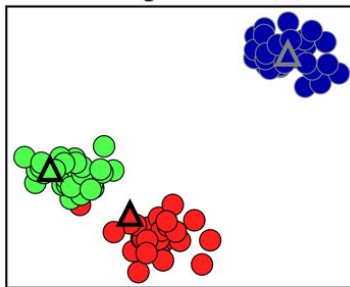
Assign Points (1)



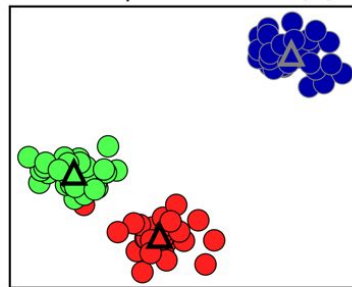
Recompute Centers (1)



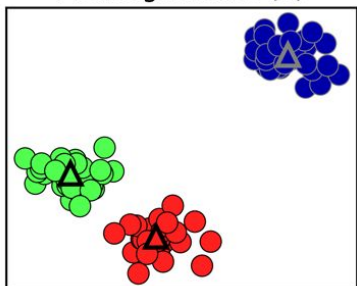
Reassign Points (2)



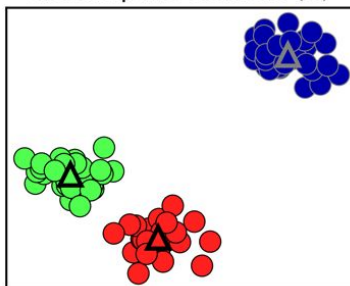
Recompute Centers (2)



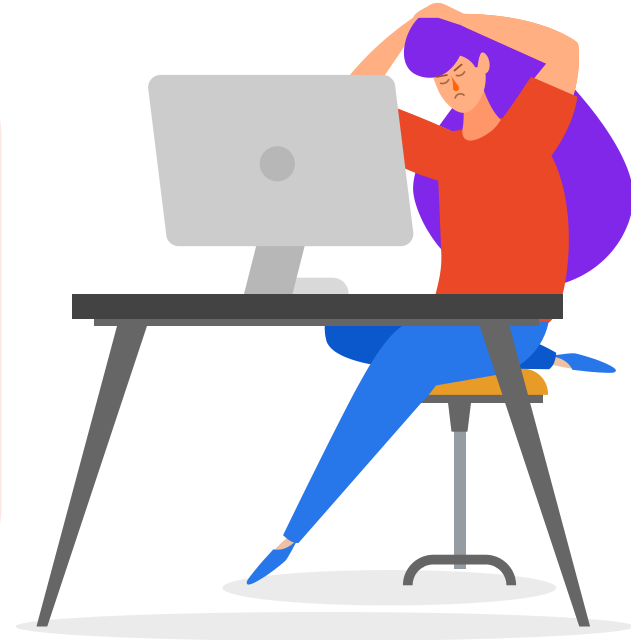
Reassign Points (3)



Recompute Centers (3)

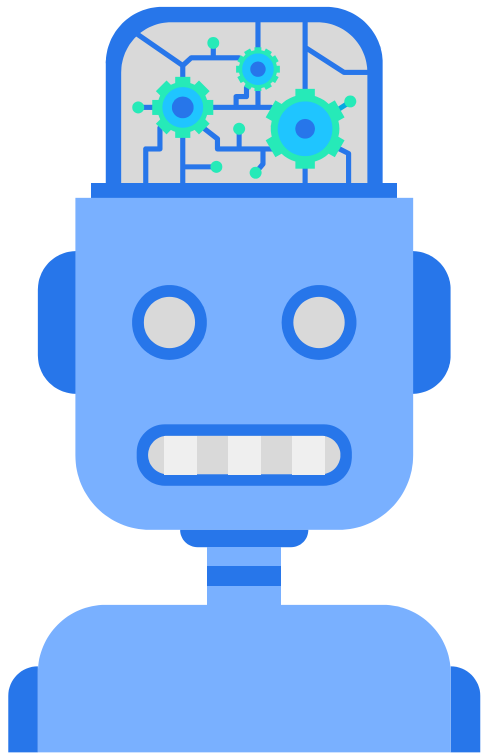


- Depende da definição do número de grupos.
- Inicialização aleatoria pode dar resultados diferentes.
- Grupos de forma complexa



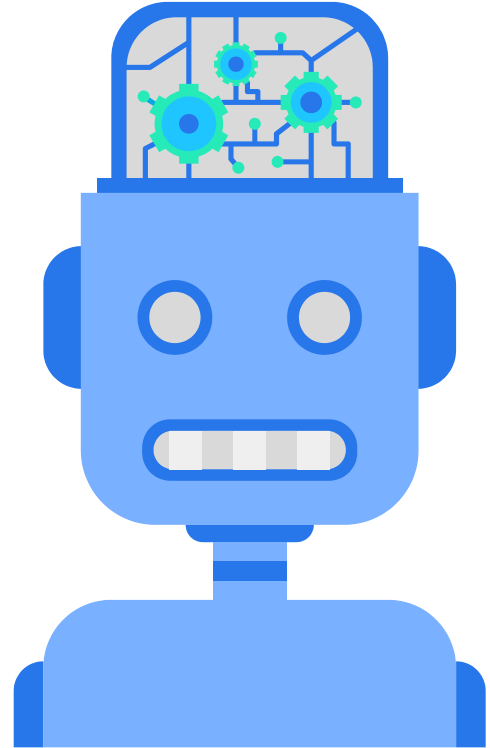


Cluster aglomerativo

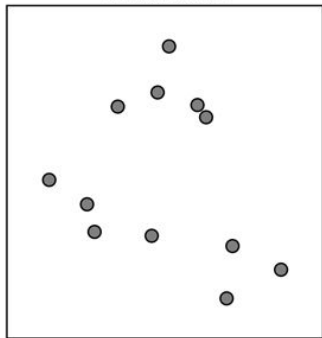


- Cada ponto é um grupo
- Os grupos são unidos considerando a distância
- Os grupos são unidos até um critério de parada

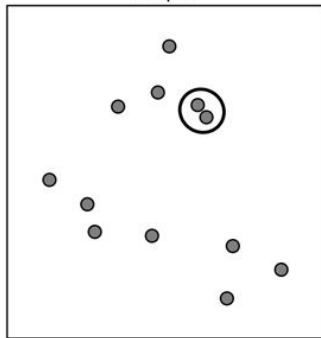
- Métodos de unificar os clusters:
 - Ward: dois grupos são unificados tal que a variabilidade dentro dos clusters aumente menos. (Cluster do mesmo tamanho)
 - Average: junta dois grupos com a menor distância média entre todos os pontos
 - Complete: junta dois grupos com a menor distância máxima entre dois pontos.



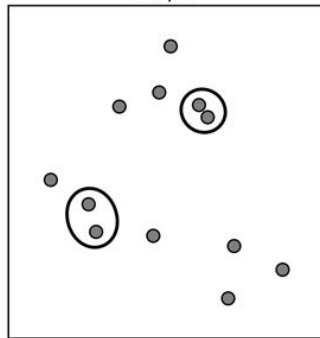
Initialization



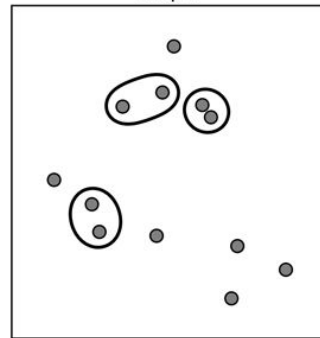
Step 1



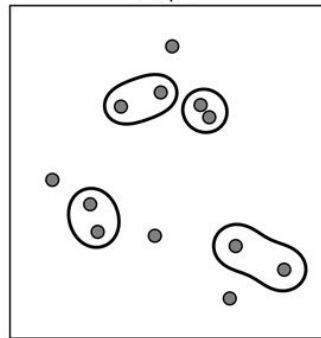
Step 2



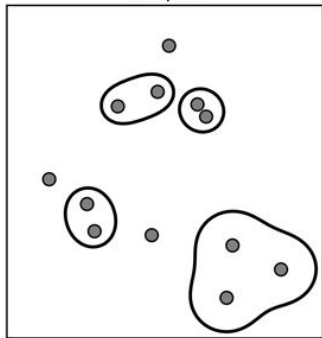
Step 3



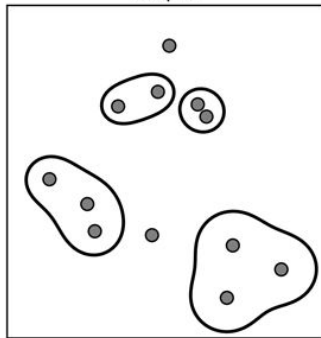
Step 4



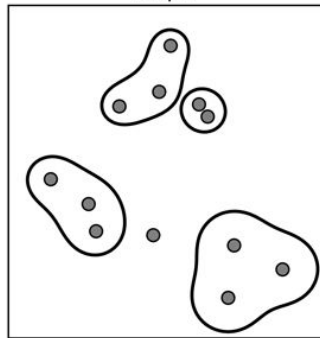
Step 5



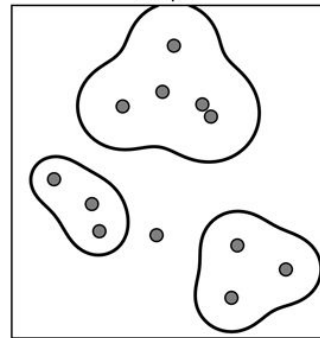
Step 6



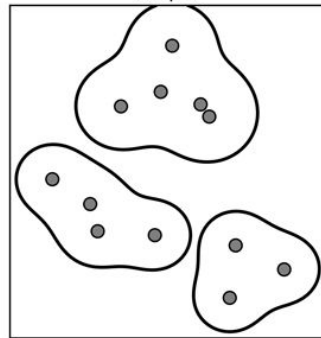
Step 7

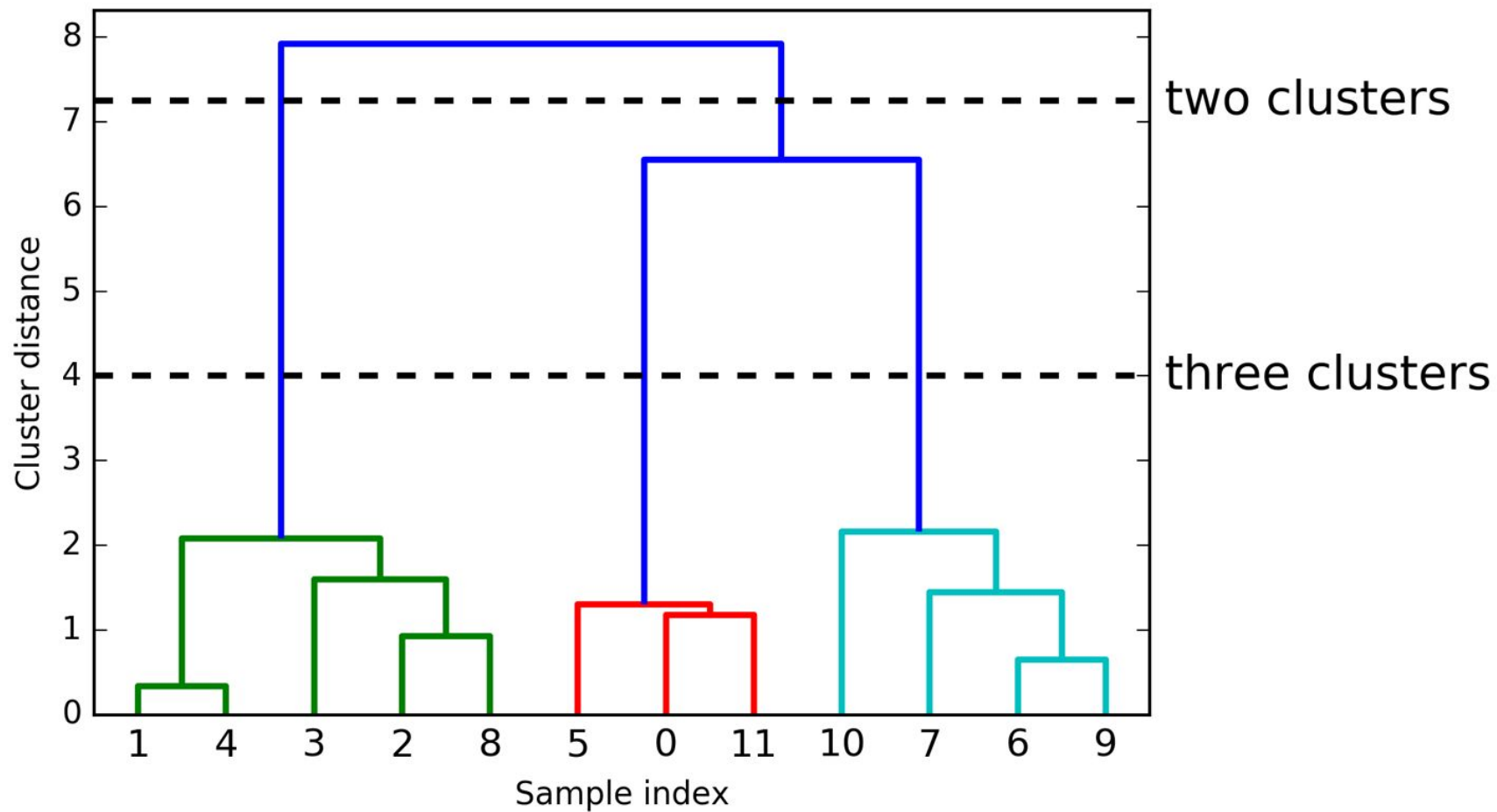


Step 8

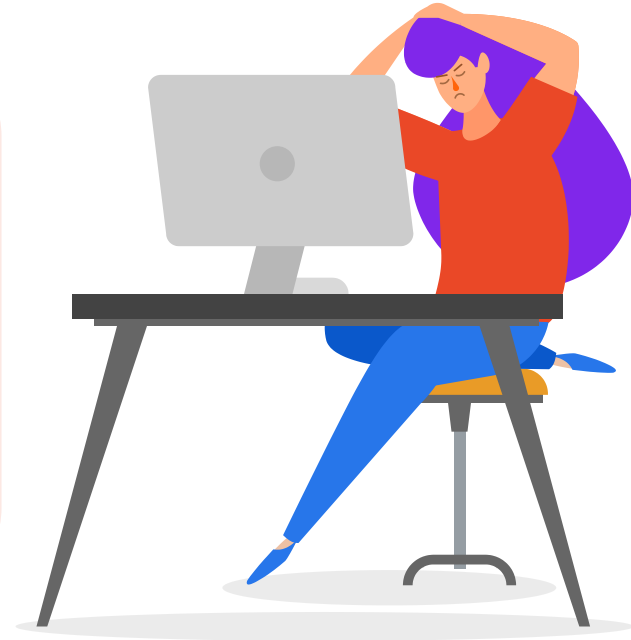


Step 9



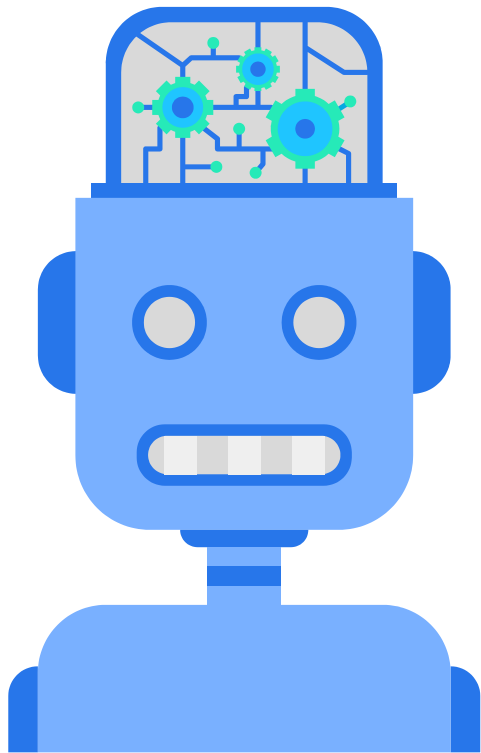


- ~~Depende da definição do número de grupos.~~
- Grupos de forma complexa



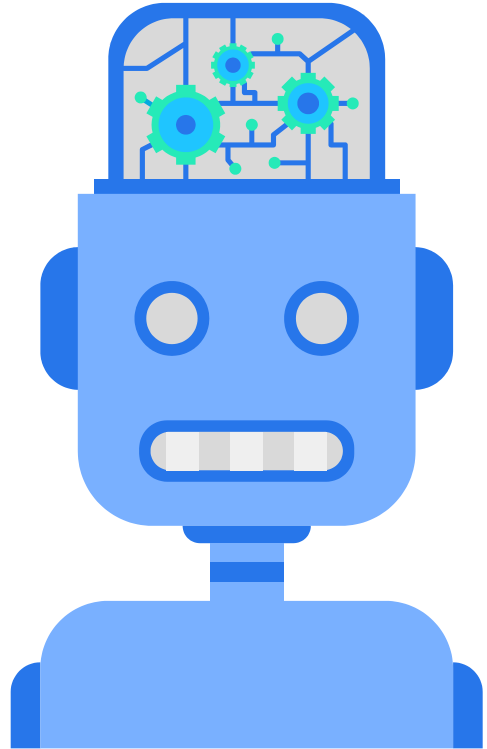


DBSCAN

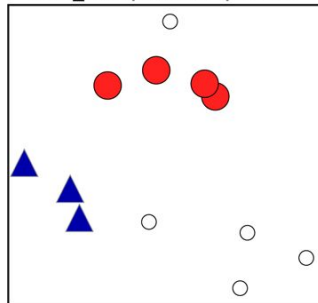


- **Density-based spatial clustering of applications with noise**
- Não depende de definição de número de cluster
- Pega formas complexas
- Procura por regiões “populosas” no espaço de feature

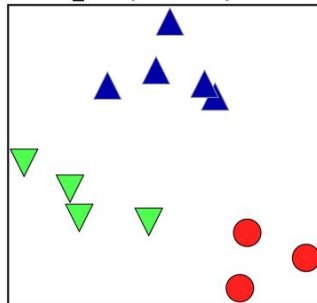
- Um ponto arbitrário é selecionado
- Acha todos os pontos com uma distância menor que **eps**
- Se tem mais pontos que **min_samples** é definido um cluster
- Da mesma forma é olhado os novos pontos vizinhos até não ser possível adicionar um novo ponto.
- Um novo ponto arbitrário é selecionado



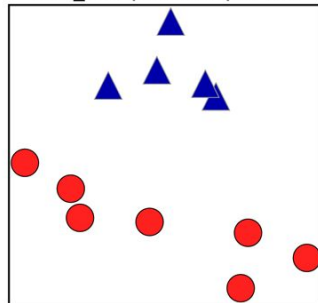
min_samples: 2 eps: 1.0



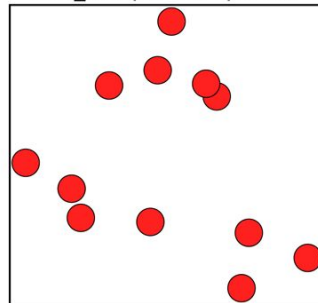
min_samples: 2 eps: 1.5



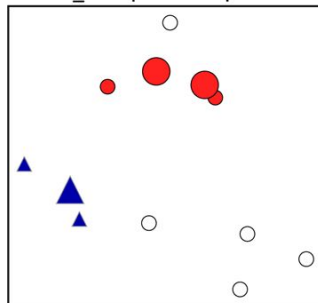
min_samples: 2 eps: 2.0



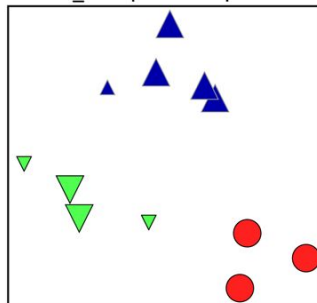
min_samples: 2 eps: 3.0



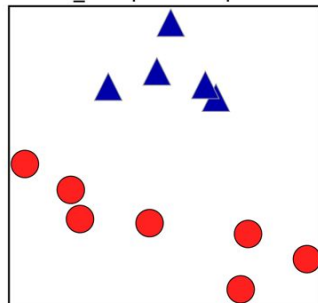
min_samples: 3 eps: 1.0



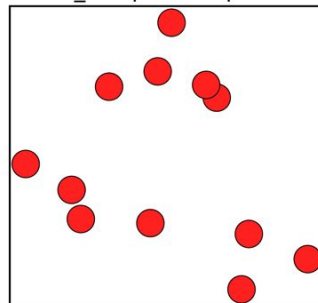
min_samples: 3 eps: 1.5



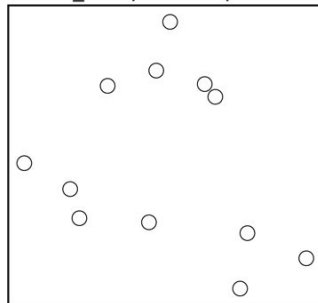
min_samples: 3 eps: 2.0



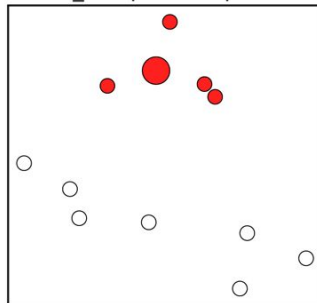
min_samples: 3 eps: 3.0



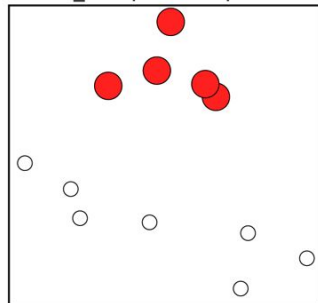
min_samples: 5 eps: 1.0



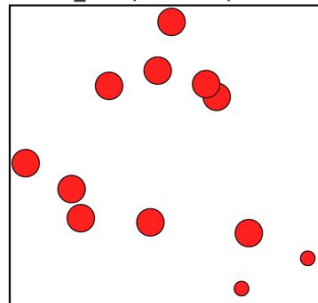
min_samples: 5 eps: 1.5



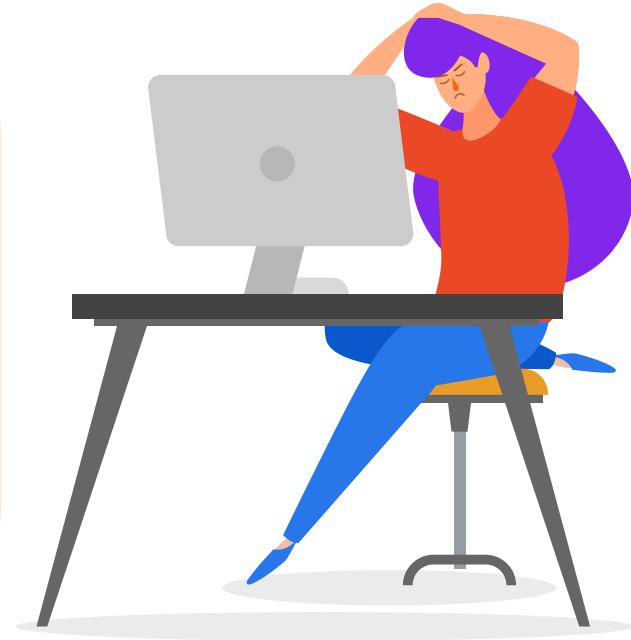
min_samples: 5 eps: 2.0



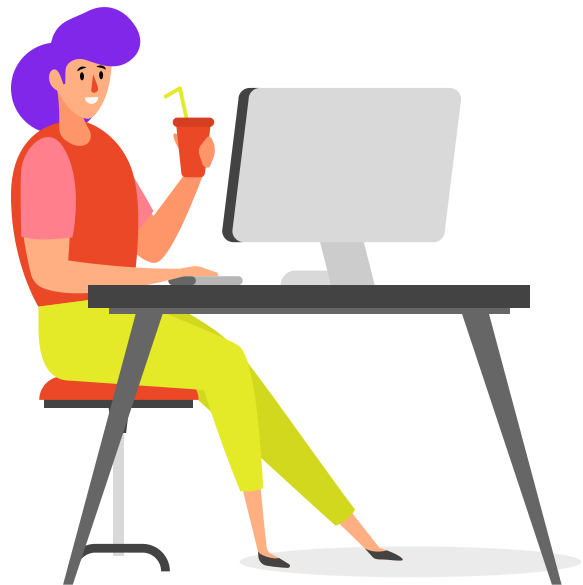
min_samples: 5 eps: 3.0



- ~~Depende da definição do número de grupos.~~
- ~~Grupos de forma complexa~~
- Pontos vizinhos podem cair em clusters diferentes



Métricas de avaliação



Métricas

ARI (Adjusted rand index)

NMI (Normalized mutual information)

Silhouette Score



Hora da prática!!!!