



Instituto Federal de Santa Catarina
Campus Florianópolis

Clusterização

Prof. Glauco Cardozo

glauco.cardozo@ifsc.edu.br

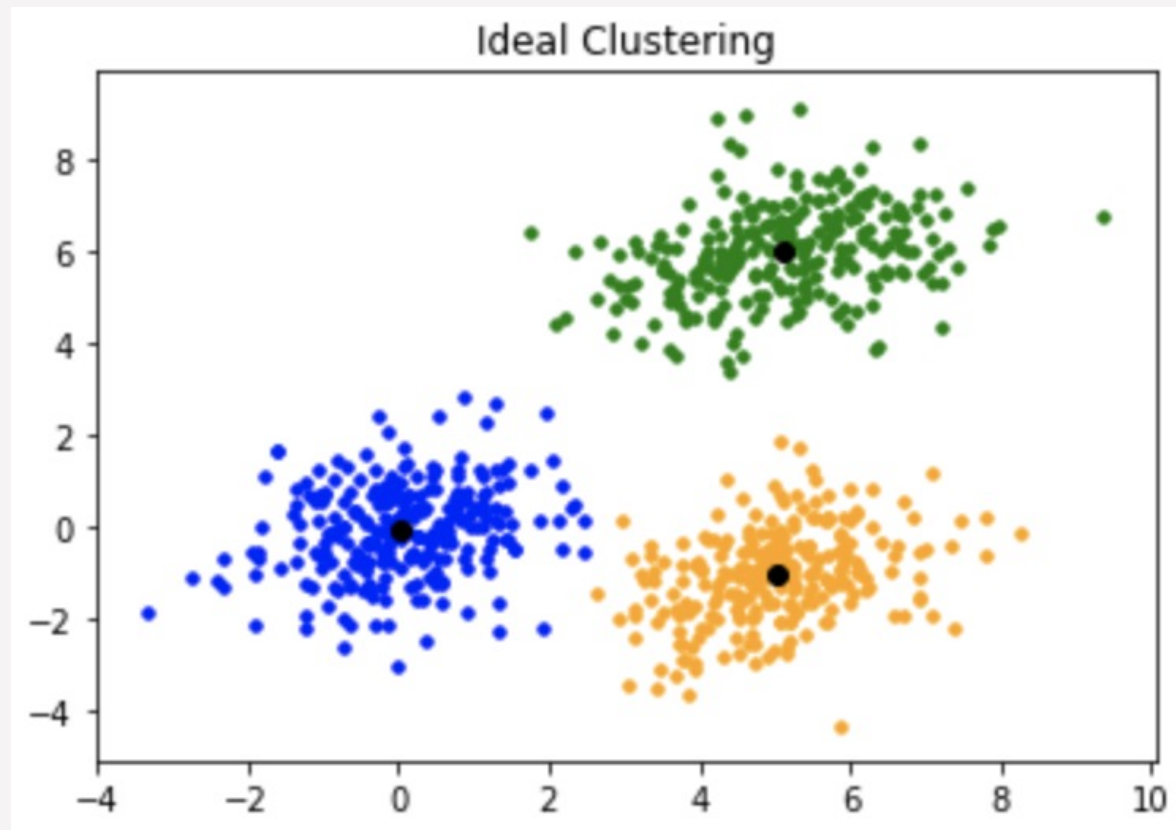


Clusterização

A clusterização é uma técnica de **machine learning não supervisionado** **que visa agrupar os dados em determinados conjuntos distintos entre si**. É muito útil para diversos contextos, como para o marketing e para estudos de mercado. Trata-se de um método descritivo, pois apenas realça as características dos dados de entrada, sem intenção de realizar previsões ativas sobre eles.



Clusterização





Clusterização

Técnicas:

- modelos de conectividade;
- modelos de centróide;
- modelos de distribuição;
- modelos de densidade.



Clusterização

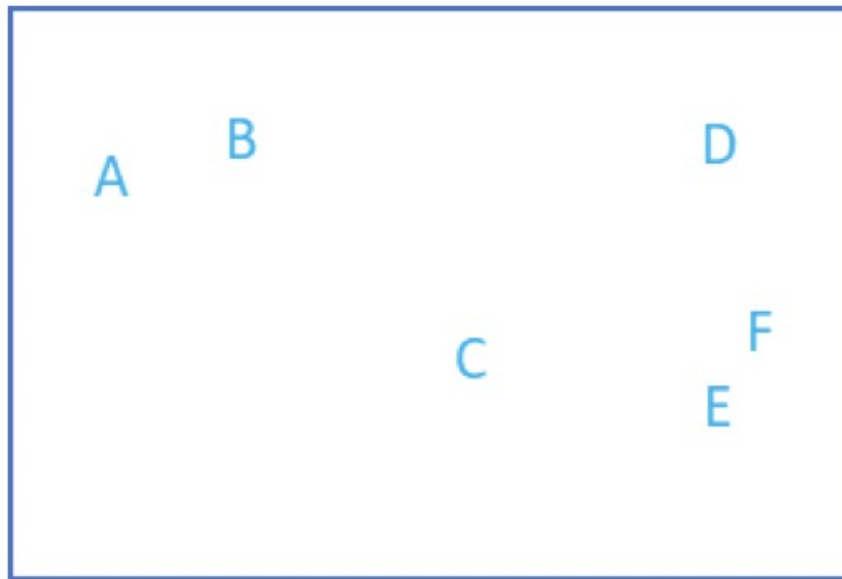
Modelos de conectividade

Os modelos de conectividade, começam demarcando todos os dados como clusters únicos. Então, a partir disso, ele agrupa determinados clusters com base em distância menor, criando uma hierarquia entre todos os grupos.

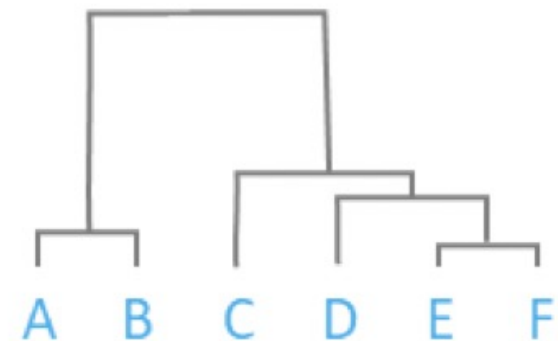


Clusterização

Modelos de conectividade



Dendrogram





Clusterização

Modelos de centroide

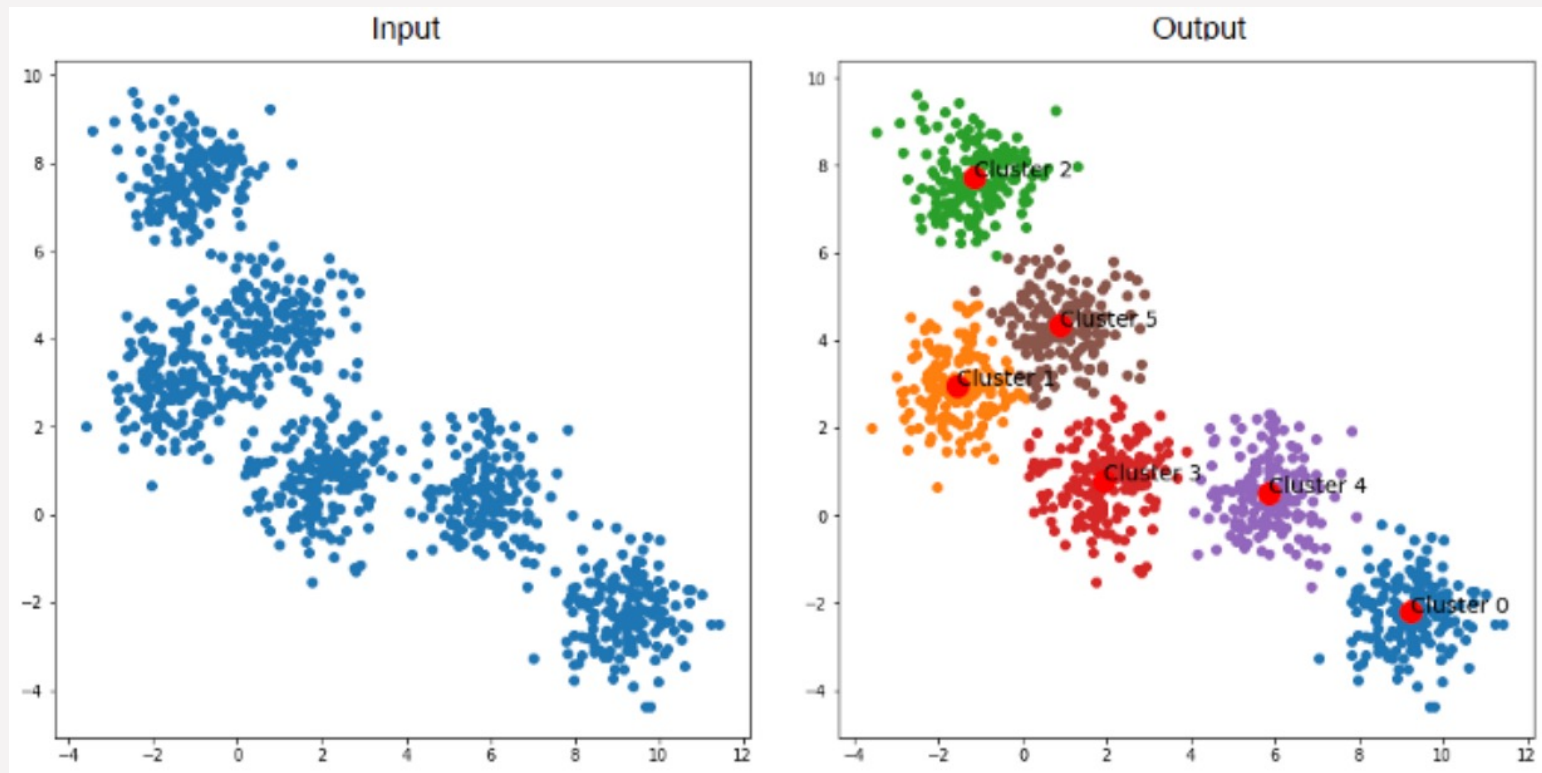
O algoritmo se baseia pela análise de centroides, definindo pontos aleatórios no meio dos dados e agrupando os elementos que estão próximos daqueles. Este processo é ajustado e repetido até encontrar o melhor agrupamento.

Um dos modelos mais famosos é o **K-means**. Um fator importante acerca desse algoritmo é que o número de clusters necessita ser definido previamente, o que requer um conhecimento maior do negócio ou dos dados analisados.



Clusterização

Modelos de centroide





Clusterização

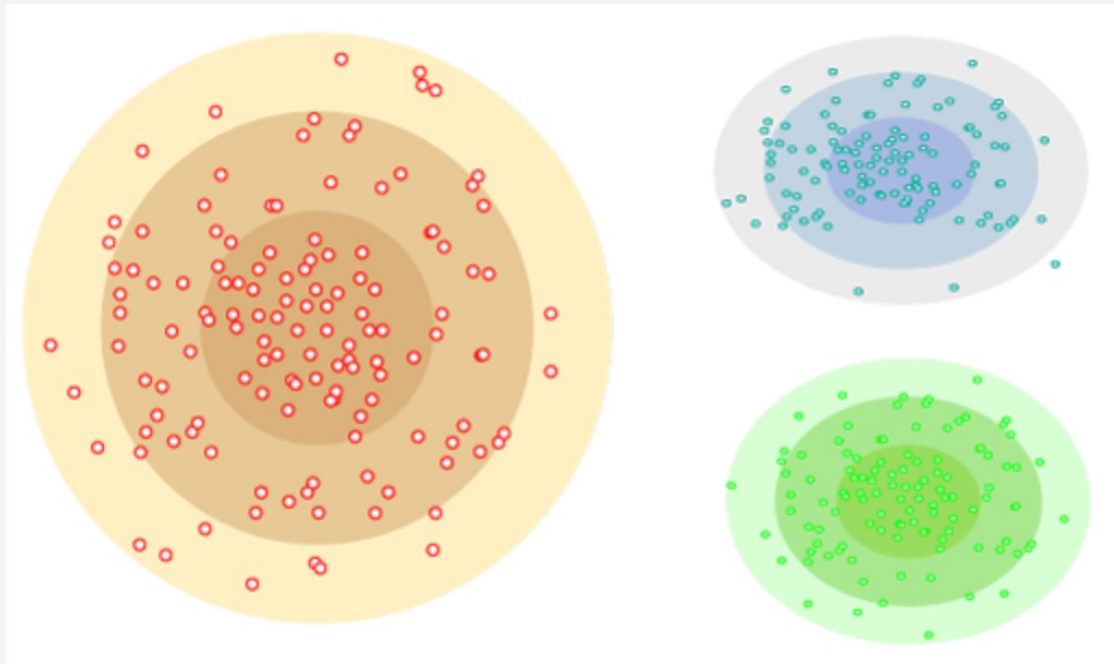
Modelos de distribuição

Os modelos de distribuição trabalham com a probabilidade de um elemento pertencer a um grupo ou não, com base na distância. Desse modo, ele consegue reduzir outliers ao fornecer uma precisão maior para lidar com a incerteza de componentes mais distantes. Então, cabe ao analista determinar o grau de precisão que ela precisa para suas análises.



Clusterização

Modelos de distribuição





Clusterização

Modelos de densidade

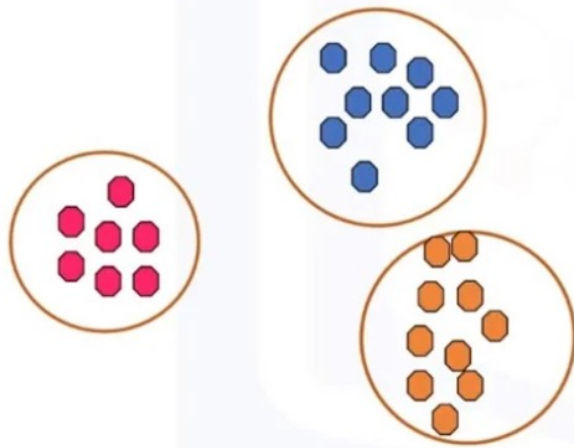
Os modelos de densidade consideram a densidade das regiões do gráfico para avaliar a possibilidade de criar grupos com similares. O **DBScan**, por exemplo, é um famoso exemplo. Ele agrupa elementos com base em um raio definido previamente. O número de clusters é determinado pelo próprio modelo.



Clusterização

Modelos de densidade

- Spherical-shape clusters



- Arbitrary-shape clusters

