



# Métricas de avaliação de Clusters

|           |  |
|-----------|--|
| ☰ Ciclo   | Ciclo 05: Aprendizado não-supervisionado |
| # Aula    | 36                                       |
| 🕒 Created | @February 24, 2023 8:50 AM               |
| ☑ Done    | ☑  |
| ☑ Ready   | ☑  |

## Objetivo da Aula:

- ☐ Avaliação de um agrupamento
- ☐ O processo de construção da Silhouette Score
- ☐ Resumo
- ☐ Próxima aula

## Conteúdo:

### ▼ 1. Avaliação de um agrupamento

#### ▼ 1.1 Coesão

A coesão mede o quanto os pontos em um cluster estão próximos uns dos outros.

Ela é medida pela soma do erro das distâncias quadráticas entre cada ponto do cluster e o seu centroide dentro do próprio cluster. Essa medida de coesão é conhecida com "Within-Cluster Sum of Squares", do inglês, WCSS

O WCSS é calculada da seguinte forma

$$WCSS = \sum_{i=1}^n (x_i - c_j)^2$$

Quanto menor o WCSS de um cluster, mais compacto ele é, ou seja, os pontos estão mais próximos uns dos outros.

## ▼ 1.2 Separação

A separação mede o quão distante os pontos em um cluster estão de outros clusters. Ela é medida pela distância entre os centróides dos clusters. Quanto maior a distância entre os centróides, mais separados os clusters estão.

Em outras palavras, a separação mede o quão bem definidos os clusters estão uns dos outros. Um baixo valor de separação indica que os clusters podem estar muito próximos uns dos outros, o que pode tornar difícil a sua distinção. Por outro lado, um valor alto de separação sugere que os clusters são bem definidos e distintos uns dos outros.

## ▼ 1.3 Silhouette Score

O Silhouette Score é uma medida de avaliação para medir a qualidade de um agrupamento. Ele é calculado para cada observação e mede o quão bem ela se encaixa no seu cluster atual, em comparação com outros clusters.

O Silhouette Score varia de -1 a 1, onde um valor mais próximo de 1 indica que a observação está bem ajustada ao seu cluster e mal ajustada aos outros clusters, e um valor mais próximo de -1 indica que a observação está mal ajustada ao seu cluster e bem ajustada aos outros clusters. Um valor próximo de 0 indica que a observação está entre dois clusters e pode ser atribuída a qualquer um deles.

O Silhouette Score é calculado da seguinte forma:

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)}$$

O Silhouette Score geralmente é usado em conjunto com outras medidas de avaliação, como o WCSS, para escolher o número ideal de clusters para um conjunto de dados.

## ▼ 2. O processo de construção da Silhouette Score

<https://docs.google.com/presentation/d/199ayVibqY4JjwoqzttRdw67MiTOoHI2YJAUaMZKWnYg/edit?usp=sharing>

## ▼ 3. Resumo

1. Um agrupamento precisa ter os pontos coesos, próximos entre si, e ser distantes uns dos outros, separação.
2. A métricas de Silhueta mede a qualidade do agrupamento.

## ▼ 4. Próxima aula

Elbow Method para encontrar o valor de K