

Eduardo Oliveira Coelho

GitHub: <https://github.com/eduardoocoelho/ai-lists>

1- Os agrupamentos obtidos pela métrica Silhouette sugerem que a clusterização é de boa qualidade, com clusters bem separados e instâncias corretamente atribuídas. O método do cotovelo indica que o número ideal de clusters é 3, com boa separação e coesão entre eles. Em resumo, a clusterização com 3 clusters é apropriada e bem definida.

2-

Silhoutte:

- A Silhouette calcula a coesão e a separação dos pontos dentro e entre clusters.
- Usa distâncias médias entre pontos e clusters.
- Média das Silhouettes mede a qualidade dos clusters (quanto mais próxima de 1, melhor).
- Cálculos:
 - A métrica Silhouette para um ponto 'i' em um cluster 'A' é calculada usando as seguintes equações:
 - Calcular a distância média entre o ponto 'i' e todos os outros pontos no mesmo cluster 'A' como 'a(i)'.
 - Calcular a menor distância média entre o ponto 'i' e todos os pontos em qualquer outro cluster ao qual 'i' não pertence como 'b(i)'.
 - A Silhouette para o ponto 'i' é então calculada como:
 - $Silhouette(i) = (b(i) - a(i)) / (\max\{a(i), b(i)\})$

Elbow:

A métrica Elbow é usada para determinar o número ideal de clusters em análises de clusters. A matemática por trás disso envolve o cálculo da inércia, que é a soma das distâncias quadradas médias entre os pontos e os centróides dos clusters.

No método Elbow, você calcula a inércia para diferentes valores de 'k' e plota esses valores em um gráfico. O ponto no gráfico onde a inércia começa a nivelar é o "cotovelo" e corresponde ao número ótimo de clusters, porque adicionando mais clusters não fornece melhor explicação da variação nos dados.

3- O índice Davies-Bouldin (DBI) é uma métrica de avaliação de clusters que mede a qualidade dos agrupamentos. Ele é calculado com base na relação entre a dispersão intra-cluster e a distância inter-cluster. Um valor menor do DBI indica clusters mais compactos e bem separados, enquanto valores mais altos sugerem sobreposição ou grande distância entre os clusters. O DBI é uma métrica útil para

complementar a avaliação de clusters junto com outras métricas, como Silhouette e método Elbow, fornecendo uma visão abrangente da qualidade dos agrupamentos.

4- A função para visualizar instâncias incorretamente classificadas permite examinar visualmente a atribuição de instâncias a clusters pelo algoritmo K-means. Ela cria dois gráficos de dispersão, um com base nos clusters reais e outro com base nos clusters previstos. Isso facilita a identificação visual das instâncias que foram atribuídas incorretamente, destacadas pela diferença de cores entre os dois gráficos. Essa análise visual auxilia na detecção de problemas de classificação e na tomada de decisões para ajustar o modelo ou explorar abordagens alternativas de clustering.

5-

O código realiza as seguintes etapas:

- **Pré-processamento dos Dados:**
 - Seleciona as colunas 0 a 3 como características de entrada (Entrada).
 - Aplica um escalonamento Min-Max para normalizar as características.
- **Determinação do Número de Clusters Ótimo:**
 - Usa o Silhouette Score e o índice Davies-Bouldin para avaliar diferentes números de clusters.
 - Aplica o método do cotovelo para determinar o número ideal de clusters.
- **Modelo K-Means:**
 - Cria um modelo K-means com o número ideal de 3 clusters.
- **Avaliação do Modelo:**
 - Visualiza as instâncias em dois gráficos de dispersão: um para clusters reais e outro para clusters previstos.
 - Destaca instâncias classificadas incorretamente.
- **Resultados Obtidos:**
 - O código fornece métricas de validação de clusters, sugere 3 clusters como o número ideal e permite a identificação visual das instâncias incorretamente classificadas pelo K-means.