

Questão 1

https://github.com/Trinalsolid/IA/blob/main/Kmeans/Lendo_e_tratando_arquivo_IRIS.ipynb

Questão 2

no método elbow os dados presentes em $k=3$ ou 4 , não são adequados para representação entre os grupos, já o Silhouette Score sugere $k=2$ com a melhor separação entre os grupos. O problema é que se utilizarmos o $k=2$ poderemos ter um modelo mais simplificado e menos preciso, se utilizarmos o $k=3$ ou 4 teremos um modelo mais complexo mas com redução na coesão dele.

<https://github.com/Trinalsolid/IA/blob/main/Kmeans/Kmeans.ipynb>

Questão 3

a escolha dos centroides no Kmeans é feita de uma forma mais precisa, mesmo que sendo aleatória o algoritmo tende a fazer uma escolha de centroide que esteja mais alinhada com o tamanho do cluster que defino, o que significa que o centroide escolhido para cada um dos clusters é mais bem alinhado com o centro do próprio cluster onde a densidade de dados é maior, do que em regiões mais afastadas que possuem poucos dados.

Questão 4

Questão 5

podemos ter a Índice de Davies-Bouldin (DBI)

O **Davies-Bouldin Index** é uma métrica **interna** que avalia a qualidade dos clusters com base na **dispersão dentro dos clusters** e **distância entre os centróides**.

```
from sklearn.metrics import davies_bouldin_score
```

```
# Suponha que você já tenha:
```

```
# - Entrada: os dados de entrada
```

```
# - saida_kmeans: os rótulos dos clusters definidos pelo KMeans
```

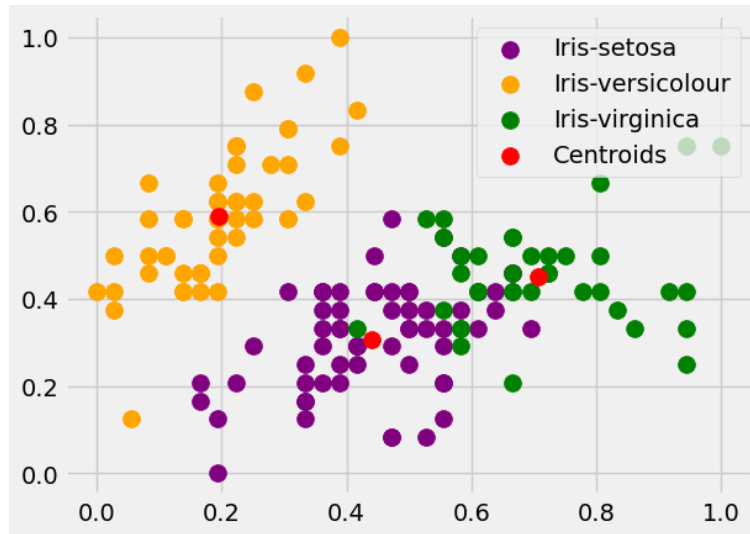
```
dbi = davies_bouldin_score(Entrada, saida_kmeans)
```

```
print(f"Davies-Bouldin Index: {dbi:.3f}")
```

Questão 6

Questão 7

podemos ver que das principais flores as que estão classificadas erradas são iris-verginica e iris-setosa



Questão 8

Leitura e Pré-processamento

- Leitura dos dados da base Iris.
- Remoção da coluna da classe (nome da flor).
- Padronização dos dados (média 0, desvio padrão 1) com StandardScaler.

Treinamento do K-Means

- Aplicação do algoritmo para diferentes valores de k (de 2 a 10).
- Inicialização com k-means++.
- Uso da **distância euclidiana** para medir similaridade entre os pontos.

Avaliação dos Resultados

- **Elbow Method (SSE):** indicou que o melhor valor de k é 3, onde a curva começa a estabilizar.
- **Silhouette Score:** maior valor em $k=2$, mas $k=3$ também teve boa separação.
- **Davies-Bouldin Index:** confirmou que $k = 3$ tem bom equilíbrio entre separação e compactação dos grupos.

Visualização

- Gráfico com os dados agrupados e os centróides destacados.
- Agrupamentos bem formados, especialmente para a classe Setosa.

Conclusão

- O K-Means funcionou bem com $k=3$, refletindo as 3 espécies reais da base.
- As métricas e a visualização confirmam a boa qualidade dos agrupamentos.