

Relatório – Comparação de Clusterização na Base Iris

Estudante: Estevão Augusto da Fonseca Santos

Professor: Ahmed Ali Abdalla Esmin

Experimentos e Avaliações

Método	K	Silhouette Score	Tempo de execução (s)
KMeans Hardcore	3	0.4553	0.0257
KMeans Hardcore	5	0.3915	0.0273
KMeans Sklearn	3	0.459	0.0395
KMeans Sklearn	5	0.348	0.000

PCA e Visualização

Foi realizada redução de dimensionalidade com PCA para 1 e 2 componentes. Nos plots, os clusters se mantêm visualmente coerentes com as classes originais, e os centróides são representados por “X” preto. A versão 2D permitiu melhor visualização da separação entre clusters, enquanto a 1D mostrou sobreposição em algumas amostras.

Análise Comparativa

- Precisão (Silhouette Score): Ambos os métodos apresentaram valores muito próximos, mostrando que a implementação do zero consegue atingir resultados semelhantes ao sklearn.
- Desempenho (tempo de execução): Implementação manual foi mais rápida que Sklearn no dataset do Iris. Casos em datasets maiores podem variar.
- Robustez: O KMeans Hardcore depende da inicialização aleatória, podendo gerar pequenas variações nos resultados.
- Aplicabilidade: Para aprendizado e compreensão do algoritmo, a implementação do zero é valiosa. Para produção ou datasets grandes, recomenda-se o uso de bibliotecas otimizadas como sklearn.

Conclusão

A clusterização com $k=3$ apresentou melhor desempenho, coerente com as classes originais da base Iris. Em datasets pequenos, a implementação do zero é funcional e possui melhor desempenho se comparado com Sklearn. Devido ao relatório se limitar apenas ao dataset do Iris, não se tem estimativa de desempenho em casos de datasets maiores. A utilização de PCA facilita a visualização e análise dos clusters, sendo uma ferramenta útil em conjunto com qualquer algoritmo de clusterização.