

Relatório de Análise de Desempenho: K-means Manual vs. Sklearn

1. Metodologia

Ambas as implementações foram executadas sobre as 4 features numéricas do dataset Iris, desconsiderando a classe alvo (`Species`). Foram realizados experimentos para encontrar 3 e 5 clusters ($K=3$ e $K=5$). A métrica de avaliação utilizada para medir a qualidade da clusterização foi o **Silhouette Score**.

2. Resultados Quantitativos

A tabela abaixo resume os scores obtidos por cada implementação. Um score mais próximo de 1 indica clusters mais bem definidos e separados.

K (Nº de Clusters)	Silhouette Score (Implementação Manual)	Silhouette Score (Sklearn)
K=3	0.5526	0.5526
K=5	0.3723	0.4885

3. Análise Comparativa

- **Para K=3 (Resultado Ótimo):** A implementação manual alcançou um desempenho **idêntico** ao da biblioteca Sklearn. Este resultado valida a corretude da lógica manual, que foi capaz de convergir para a mesma solução ótima encontrada pela ferramenta profissional. Ambas as abordagens identificaram corretamente $K=3$ como a melhor configuração para este dataset.
- **Para K=5 (Resultado Subótimo):** A implementação do Sklearn demonstrou uma **superioridade clara**. O score mais alto (0.4885 vs. 0.3723) se deve principalmente a duas otimizações da biblioteca:
 1. **Inicialização k-means++:** O Sklearn utiliza um método mais inteligente para escolher os centróides iniciais, o que aumenta a chance de uma boa convergência.
 2. **Múltiplas Execuções (`n_init=10`):** A biblioteca executa o algoritmo 10 vezes com diferentes pontos de partida e seleciona o melhor resultado, tornando-a mais robusta contra convergências em mínimos locais (soluções boas, mas não ótimas). A versão manual, por ter rodado apenas uma vez, ficou suscetível a um resultado inferior.
- **Eficiência e Robustez:** Embora não medido, o Sklearn é comprovadamente mais rápido em datasets maiores por ser otimizado em baixo nível (Cython). Sua principal vantagem prática é a robustez, garantida pelo parâmetro `n_init`, que o torna mais confiável para uso geral.

4. Conclusão

A análise demonstra que, para aplicações práticas e cenários mais complexos, a implementação do **Sklearn é a escolha superior**. Sua robustez, velocidade e mecanismos de otimização (como `n_init`) garantem resultados mais consistentes e de maior qualidade, minimizando o risco de soluções subótimas. A construção manual é excelente para o aprendizado, enquanto a biblioteca é a ferramenta adequada para o trabalho profissional.