



Integrantes: Layon Walker Tobias Pedro e Marcos Vinícius Matias do Nascimento

RA: 202311072 , 202311003

GCC128 - Inteligencia Artificial

Relatório Técnico Clusterização K-means

O objetivo deste projeto foi realizar a implementação do algoritmo de clusterização K-means a partir do zero em Python, utilizando o dataset Iris como base experimental. A lógica central foi encapsulada na classe `KMeans_Do_Zero`, desenvolvida para ser robusta ao espelhar o comportamento da biblioteca Scikit-learn, executando o algoritmo 10 vezes (`n_init=10`) com diferentes centróides iniciais e selecionando o melhor resultado com base na menor inércia. Após o desenvolvimento, foram conduzidos experimentos para comparar a solução própria com a implementação padrão do Scikit-learn. A análise dos resultados, conforme exibido na saída do terminal, mostrou que a configuração com $k=3$ foi a melhor, alcançando um `silhouette_score` de 0.5528, enquanto $k=5$ obteve um score de 0.4863.

Um ponto interessante foi o tempo de execução: para $k=3$, a implementação própria foi mais rápida (0.0240s vs 0.0460s), mas para $k=5$, o Scikit-learn se mostrou mais eficiente (0.0335s vs 0.0401s). Isso se deve à estratégia de inicialização `k-means++` do Scikit-learn, que, embora tenha um custo inicial maior, converge mais rapidamente em cenários mais complexos. Para validar visualmente a eficácia da clusterização, a técnica de PCA foi aplicada sobre o melhor resultado ($k=3$), reduzindo a dimensionalidade para dois componentes. O gráfico gerado ilustra claramente três clusters distintos e bem separados, com os centróides (marcados em "X") corretamente posicionados, confirmando a alta qualidade do agrupamento indicado pelo score de silhueta.

Em conclusão, o algoritmo foi implementado com sucesso, produzindo uma melhora significativa no entendimento do conteúdo necessário para implementação do zero e do desempenho comparativo com a biblioteca padrão.