

Introdução

Este relatório apresenta uma análise da complexidade do algoritmo K-Means implementado em Python e orientado a objeto. Esse trabalho é referente ao componente curricular Engenharia de Programas. O K-Means é um algoritmo de clustering amplamente utilizado para particionar um conjunto de dados em K grupos, minimizando a variação dentro dos clusters ao longo de suas iterações. O estudo visa compreender como a complexidade do algoritmo varia com os parâmetros 'n' e 'K'.

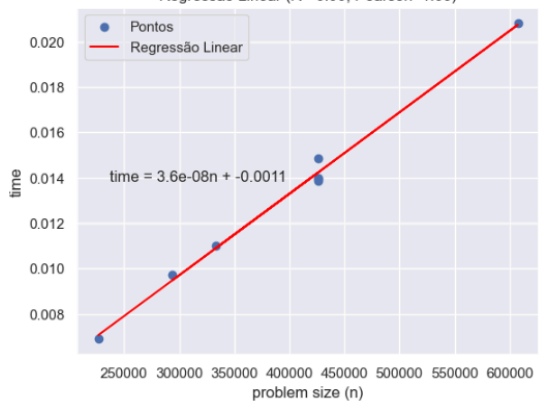
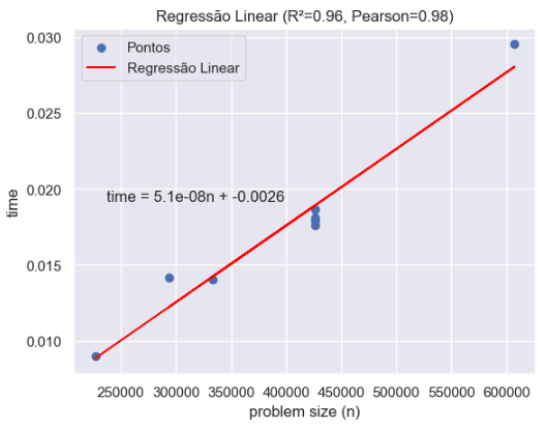
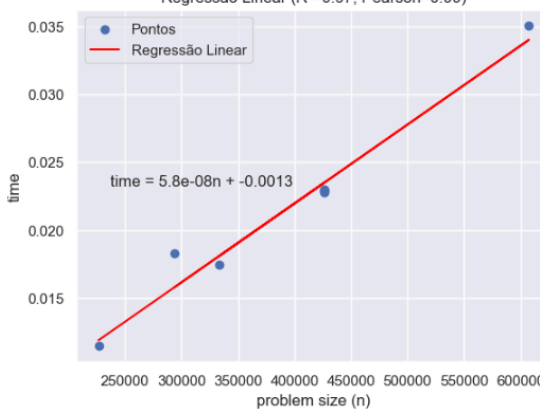
Metodologia

Foi realizada a geração de dados usando 8 imagens escolhidas aleatoriamente e fazendo a clusterização para $K = 2$, $K = 3$, $K=4$ para a análise de suas variáveis como tempo médio de iterações, desvio padrão e variabilidade. Ainda, foram feitas R repetições para obter valores mais estáveis.

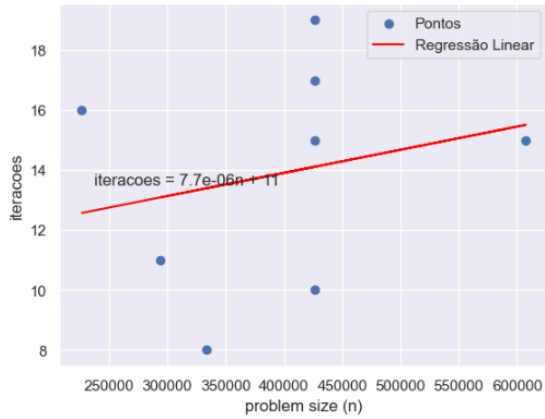
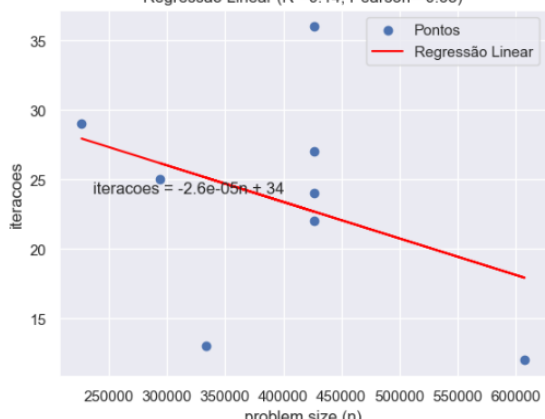
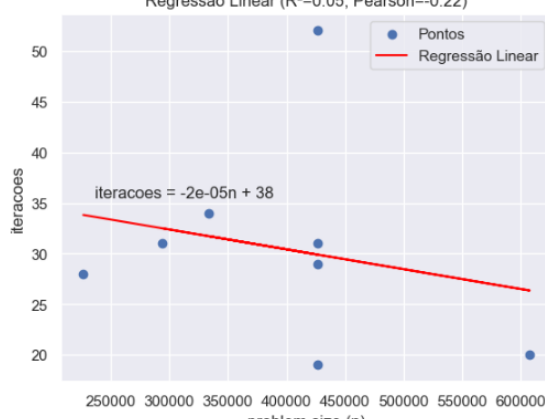
Para cada execução foram armazenados os valores de tempo médio, número de iterações, quantidade de clusters, desvio padrão e variabilidade. Em seguida, foram gerados gráficos para identificar correlações entre variáveis e realizar uma regressão linear entre variáveis para obter uma função que seja possível estimar o tempo médio gasto dado um tamanho n do problema.

Resultado:

Por meio dos resultados dos gráficos dos notebooks para K=2, K=3 e K=4 fica claro a relação entre os valores de tempo (tempo médio para cada iteração) e o tamanho do problema (quantidade de pixels de uma imagem).

	$\text{Time} = 3,8 * 10^{-8} * n$
	$\text{Time} = 5,1 * 10^{-8} * n$
	$\text{Time} = 5,8 * 10^{-8} * n$

Além disso, também foi identificado que não há uma relação entre as variáveis de número de iterações e o tamanho do problema. Pois os pontos no gráfico de dispersão não seguem uma regularidade.

 <p>Regressão Linear ($R^2=0.05$, Pearson=0.23)</p> <p>iteracoes = $7.7e-06n + 11$</p>	<p>Iterações = $7.7 * 10^{-6} * n$</p>
 <p>Regressão Linear ($R^2=0.14$, Pearson=-0.38)</p> <p>iteracoes = $-2.6e-05n + 34$</p>	<p>Iterações = $-2,6 * 10^{-5} * n$</p>
 <p>Regressão Linear ($R^2=0.05$, Pearson=-0.22)</p> <p>iteracoes = $-2e-05n + 38$</p>	<p>Iterações = $-2 * 10^{-5} * n$</p>

