Tarefa: Métodos para Calcular o melhor Número de Clusters
Nome: Erich Morais

Como escolher o melhor número ou a quantidade correta de clusters? Para isso existem alguns métodos, porém os mais conhecidos são o Elbow (Cotovelo, veremos o porquê) e o Average Silhouette (Média de Silhueta). Como já sabemos o K-Means é usado para o aprendizado não supervisionado (dados não conhecidos), e separa os dados em k clusters, de acordo com a distância de cada ponto até algo chamado de centroide (protótipo de um cluster).

Elbow: Nesse método é preciso rodar o algoritmo previamente com alguns valores. Assim se calcula a função de custo, sendo ela a soma dos quadrados das distancias internas dos clusters. O melhor número é encontrado quando a adição na quantidade de clusters, não muda significativamente a função de custo (chamada de linha de cotovelo).

Average Silhouette: Esse método mede quão bem um ponto se encaixa dentro do seu cluster. Existe o coeficiente de Silhouette que quando próximo de 1, indica que os pontos estão muitos longes dos pontos de outros clusters, já quando próximo de 0, indica que os pontos estão muito perto ou interseccionando outro cluster. Esse coeficiente é definido pela distancia media de um ponto para todos os outros em seu cluster e a distancia media até os pontos do cluster mais perto.