



Tarefa Prática - Testando os métodos de agrupamento com sementes de trigo.

Usando os mesmos dados de sementes de trigo da Tarefa 2, realizaremos a comparação de métodos de agrupamento de k-Means e hierárquico usando o pacote [scikit-learn](#).

1. Obtenha os dados de propriedades de sementes de três variedades diferentes de trigo do site [OpenML](#). Os dados possuem 210 instâncias com 8 atributos cada.
2. Pré-processamento de dados:
 - Verifique as correlações entre atributos e elimine os atributos dependente (por exemplo, use somente 4 atributos que tem correlação média menor).
 - Faça a normalização dos atributos independentes.
3. Realize agrupamento k-Means (usando [sklearn.cluster.KMeans](#)) supondo 3 grupos e pontos centrais iniciais aleatórios. Discute como melhorar a escolha dos centroids iniciais. Faça um agrupamento supondo 4 grupos. Calcule a coesão e separação de agrupamentos em 3 e 4 grupos e determine qual número de grupo é correto.
4. Realize agrupamento hierárquico (usando [sklearn.cluster.AgglomerativeClustering](#)) e produza o dendrograma dele. Para dicas, veja o tutorial [SciPy Hierarchical Clustering and Dendrogram](#) e [Hierarchical Clustering](#).