

06

A estabilidade dos clusters

Transcrição

Nosso próximo passo é validar a estabilidade do cluster. A ideia é que dividamos a base de dados e executemos o `kmeans` para cada uma dessas divisões a fim de obtermos resultados similares.

Utilizaremos uma função no Numpy para realizar essa divisão, a `np.array_split()`, por 3 conjuntos. Atribuiremos nomes para cada um deles, executaremos o algorítimo para esses conjuntos além de coletar o valor das métricas.

```
set1, set2, set3 = np.array_split(values, 3)
w1, dbs1, calinski1 = clustering_algorithm(5, set1)
w2, dbs2, calinski2 = clustering_algorithm(5, set2)
w3, dbs3, calinski3 = clustering_algorithm(5, set3)
print(s1, dbs1, calinski1)
print(s2, dbs2, calinski2)
print(s3, dbs3, calinski3)
```

Termos valores parecidos nas três colunas e com as métricas diferentes, isso indica que nosso cluster é válido. Feito isso, podemos partir para interpretação dos clusters, já que obtemos bons resultados e temos dados confiáveis.