

Executando o clustering

Transcrição

Em nossa clusterização utilizaremos o algoritmo mais popular do tema: `Kmeans`. Novamente usaremos a função do Scikit learn para coletar esse algoritmo. Precisaremos definir o número de clusters. Inicialmente colocaremos apenas 5.

Em seguida, insiremos o `n_init` como `10`, mas o que isso significa? O `Kmeans` é um algoritmo não-determinístico, isto é, a cada execução dele obteremos um resultado diferente. Então configurando o valor 10, dez vezes seguidas o caminho deve retornar o mesmo valor para garantir sua confiabilidade. Basicamente forçamos o algoritmo a ser executado dez vezes.

Por fim, inserimos o número máximo de interações que o algoritmo irá realizar, que será 300. Feito isso, coletaremos o resultado

```
from sklearn.cluster import Kmeans
kmeans = KMeans(n_clusters=5, n_init=10, max_iter=300)
y_pred = kmeans.fit_predict(values)
```

Nas próximas aulas validaremos esses clusters e interpretaremos seus conteúdos.