

Obtener clusters.

Guillermo Ruiz

En este ejercicio se deberán obtener los clusters del conjunto de datos del archivo `X.npy`. Usar los algoritmos

- k-means
- DBSCAN
- HDBSCAN
- AgglomerativeClustering

Luego evaluar la calidad de los cluster de cada algoritmo con las métricas intra/inter (con 500 pares) y silueta. Modificar los parámetros de los algoritmos para obtener clusters de buena calidad.

```
from sklearn.cluster import KMeans
from sklearn.cluster import DBSCAN
from sklearn.cluster import HDBSCAN
from sklearn.cluster import AgglomerativeClustering
import numpy as np
from sklearn.metrics import silhouette_score
```

```
X = np.load("X.npy")
X.shape
```

(1797, 64)

Resultados

Los resultados deberán ser mostrados usando el siguiente código.

```
import pandas as pd
```

```

res = {"algoritmos": ["k-means", "DBSCAN", "HDBSCAN", "AgglomerativeClustering"],
      "intra/inter": # aquí sus scores
      "silueta": # aquí sus scores
      }
output = pd.DataFrame(res)
output

```

	algoritmos	intra/inter	silueta
0	k-means	0.734233	0.143674
1	DBSCAN	0.908178	-0.043750
2	HDBSCAN	0.917728	0.039615
3	AgglomerativeClustering	0.721954	0.178497

¿Las dos métricas coinciden en el mejor algoritmo de agrupamiento?