Dejar a la máquina clasificar sola

Ha llegado el turno de la máquina. ¿Sabrá clasificar las cerezas y los albaricoques correctamente?

Para eso vamos a utilizar el algoritmo K-Mean parametrizándolo de la siguiente manera:

from sklearn.cluster import KMeans

modelo=KMeans(**n\_clusters=2**)

Le hemos indicado la cantidad de  clústeres que es preciso determinar, que en nuestro caso son dos (Cereza o Albaricoque).

Una vez hecha la parametrización, pasamos a la fase de aprendizaje:

modelo.fit(frutas)

A continuación avanzamos hasta la fase de predicciones con una visualización gráfica de estas y los centroides de los dos clústeres:

#Predicciones

predicciones\_kmeans = modelo.predict(frutas)

#Visualización de la clusterización

plt.scatter(frutas.DIAMETRO, frutas.PESO, c=predicciones\_kmeans,

s=50, cmap='viridis')

plt.xlabel("DIAMETRO")

plt.ylabel("PESO")

#Visualización de los centroides

centers = modelo.cluster\_centers\_

plt.scatter(centers[:, 0], centers[:, 1], c='black', s=200,

alpha=0.5)

plt.show()

Gráfico

Descripción generada automáticamente

*Resultado de la clasificación*

Según este gráfico, comprobamos que la máquina ha conseguido clasificar correctamente nuestras observaciones. Como habíamos previsto, en la parte inferior del gráfico se encuentran las cerezas, y los albaricoques a la derecha.

Los colores del gráfico se han generado mediante la instrucción cmap (color map) del módulo matplotlib; su documentación y lista de colores disponibles se encuentran en esta dirección: <https://matplotlib.org/examples/color/colormaps_reference.html>

El modelo nos parece pertinente y podemos guardarlo para volver a utilizarlo después.

Por eso, es conveniente importar el módulo joblib a nuestro proyecto Python, crear un directorio nuevo llamado modelos y utilizar las siguientes líneas de programa:

from joblib import dump

dump(modelo,'modelos/kmean.joblib')

Una vez guardado, no olvide incluir este programa como comentarios para no sobrescribir su copia de seguridad durante posibles pruebas posteriores:

#from joblib import dump

#dump(modelo,'modelos/kmean.joblib')