Assignment 8

Josh Barber

barb4630@mylaurier.ca

215854630

1. Fit the 2 estimators with X and evaluate their performance by [v\_measure\_score](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.v_measure_score.html). Write down the code and compare its performance. (1 point)

from sklearn import cluster

from sklearn.metrics.cluster import v\_measure\_score

from sklearn.cluster import KMeans

from sklearn import datasets

from sklearn.preprocessing import StandardScaler

iris = datasets.load\_iris()

X = iris.data

labels\_true = iris.target

#X = StandardScaler().fit\_transform(X)

estimators = [

("k\_means\_iris\_3", KMeans(n\_clusters=3)),

(f"DBSCAN\_0.5", cluster.DBSCAN(eps=0.5))

]

for name, algorithm in estimators:

algorithm.fit(X)

if hasattr(algorithm, "predict"):

labels\_pred = algorithm.predict(X)

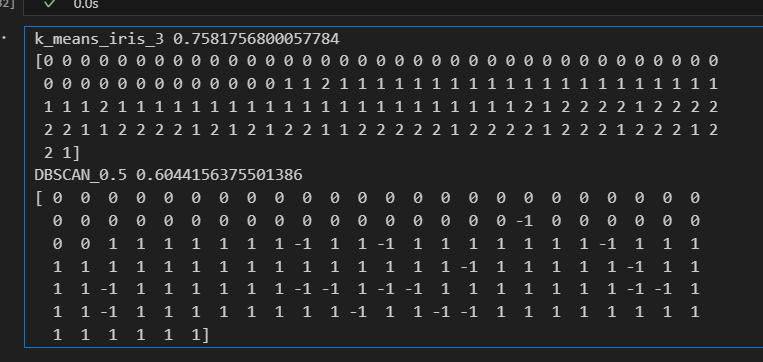
else:

labels\_pred = algorithm.fit\_predict(X)

print(name, v\_measure\_score(labels\_true, labels\_pred))

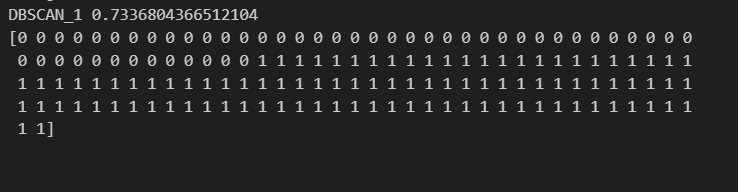
print(algorithm.labels\_)

1. After fitting the 2 estimators, print the labels of the DBSCAN estimator. If your DBSCAN estimator is stored in variable **est**, you can get the labels by running **print(est.labels\_)**. You will see many -1 in the list. What are those data points with the -1 label? If you change **eps** to 1, do you still see it? Can you explain the change? (2 points)

K means (3 clusters) and DBSCAN (eps=0.5) v\_measure\_score and labels

DBSCAN labels noisy data with –1, so all the data points with a label of –1 are considered noisy data and are not put into a cluster.

If we increase the epsilon value to 1, we can see that there is no longer any noisy data. Epsilon is the radius value around a point x, to help determine clusters. Increasing the epsilon value allowed all data points to be clustered.

DBSCAN (eps=1) v\_measure\_score and labels