DBSCAN2

July 5, 2023

[ ]:

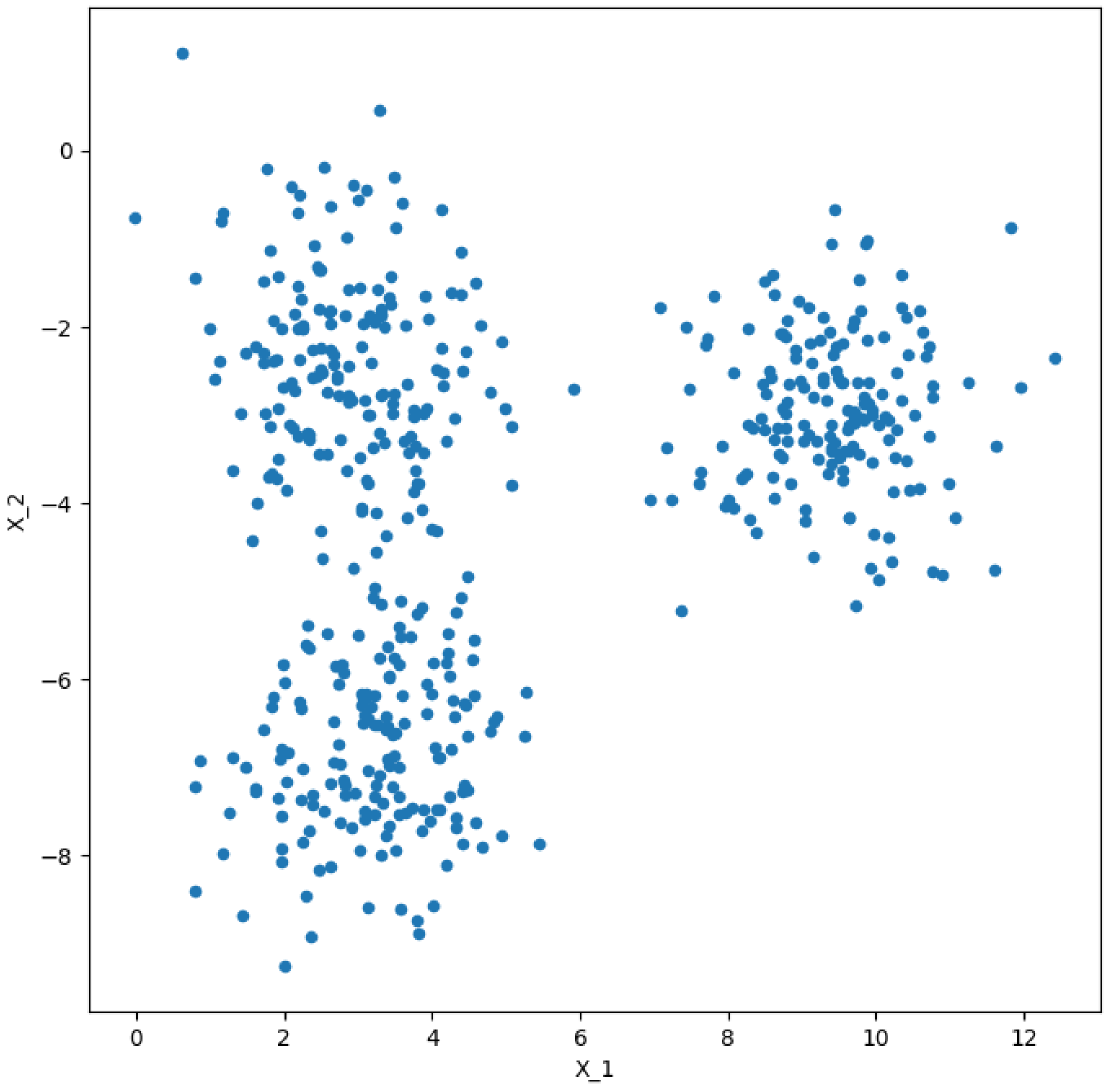
```
# ############################################################
# This code is from following the linked example below
# - source: https://youtu.be/Q7iWANbkFxk
# ############################################################

import numpy as np
from sklearn.datasets import make_blobs
from matplotlib import pyplot as plt
from matplotlib.pyplot import figure
from pandas import DataFrame

# try random states: 20, 25, 30
X, _ = make_blobs(n_samples=500, centers=3, n_features=2, random_state= 30)
```

[ ]:

```
df = DataFrame(dict(x=X[:,0], y=X[:,1]))
fig, ax = plt.subplots(figsize=(8,8))
df.plot(ax=ax, kind='scatter', x='x', y='y')
plt.xlabel('X_1')
plt.ylabel('X_2')
plt.show()
```

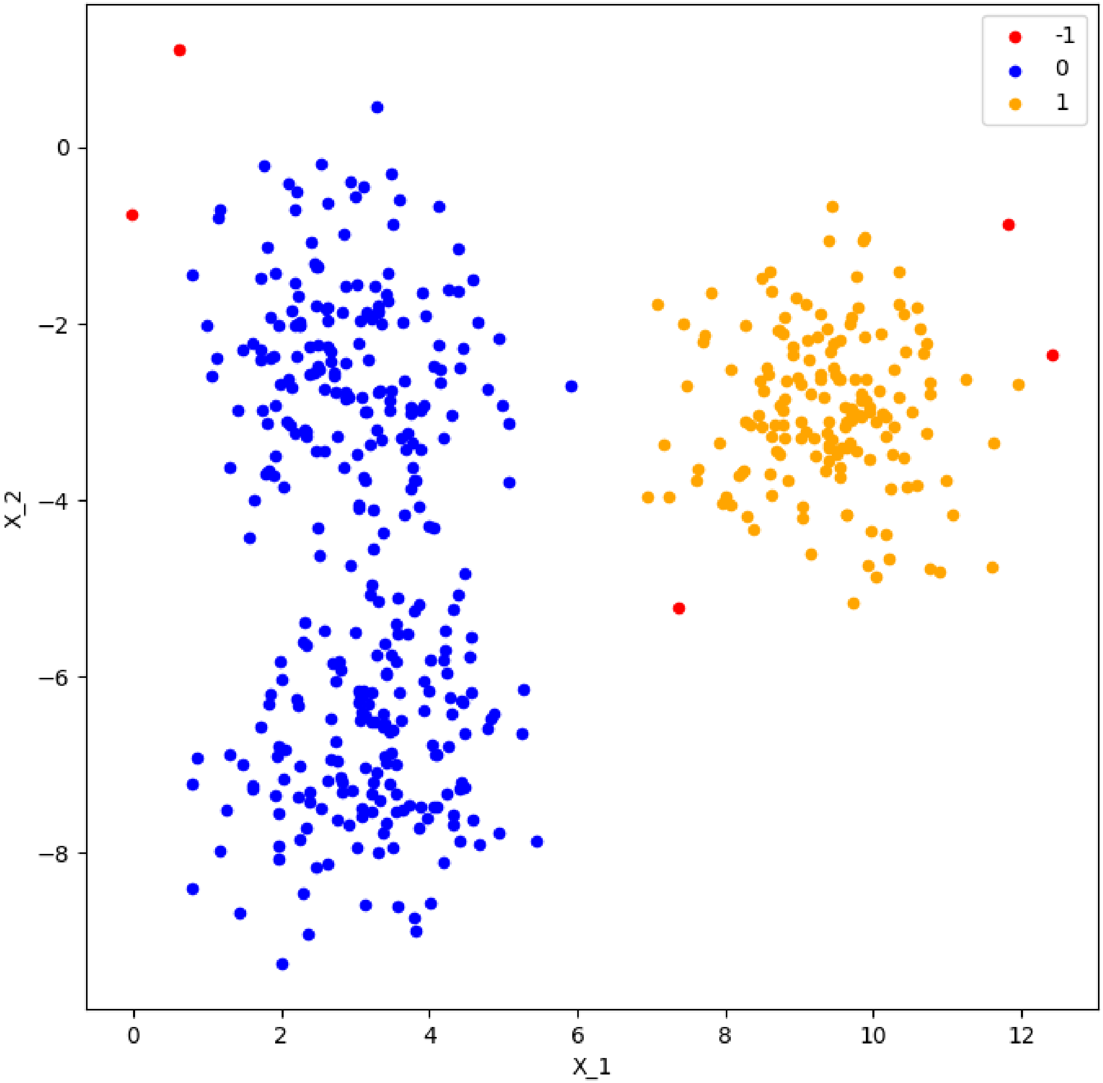


```
[ ]: from sklearn.cluster import DBSCAN
     # two most important paramaters epsilon and z
     # § epsilon = radius of circle around a particular pt. where we draw the lines for neighborhoods
     # z: threshold for least number of pts for a pts neighborhood
     clustering = DBSCAN(eps=1, min_samples=5).fit(X)
     cluster = clustering.labels_
```

```
[ ]: len(set(cluster))
```

```
[ ]: 3
[ ]: def show_clusters(X, cluster):
         df = DataFrame(dict(x=X[:,0], y=X[:,1], label=cluster))
         colors = {-1: 'red', 0:'blue', 1:'orange', 2: 'green', 3:'skyblue'}
         fig, ax = plt.subplots(figsize=(8, 8))
         grouped = df.groupby('label')
         for key, group in grouped:
             group.plot(ax=ax, kind='scatter', x='x', y='y', label=key,␣
       ↪color=colors[key])
         plt.xlabel('X_1')
         plt.ylabel('X_2')
         plt.show()
```

```
[ ]: show_clusters(X, cluster)
```


● -1
● 0
● 1
0
−2
−4
−6
−8
X_2
0
2
4
6
8
10
12
X_1