FUNDAMENTALS OF MACHINE LEARNING IN DATA SCIENCE

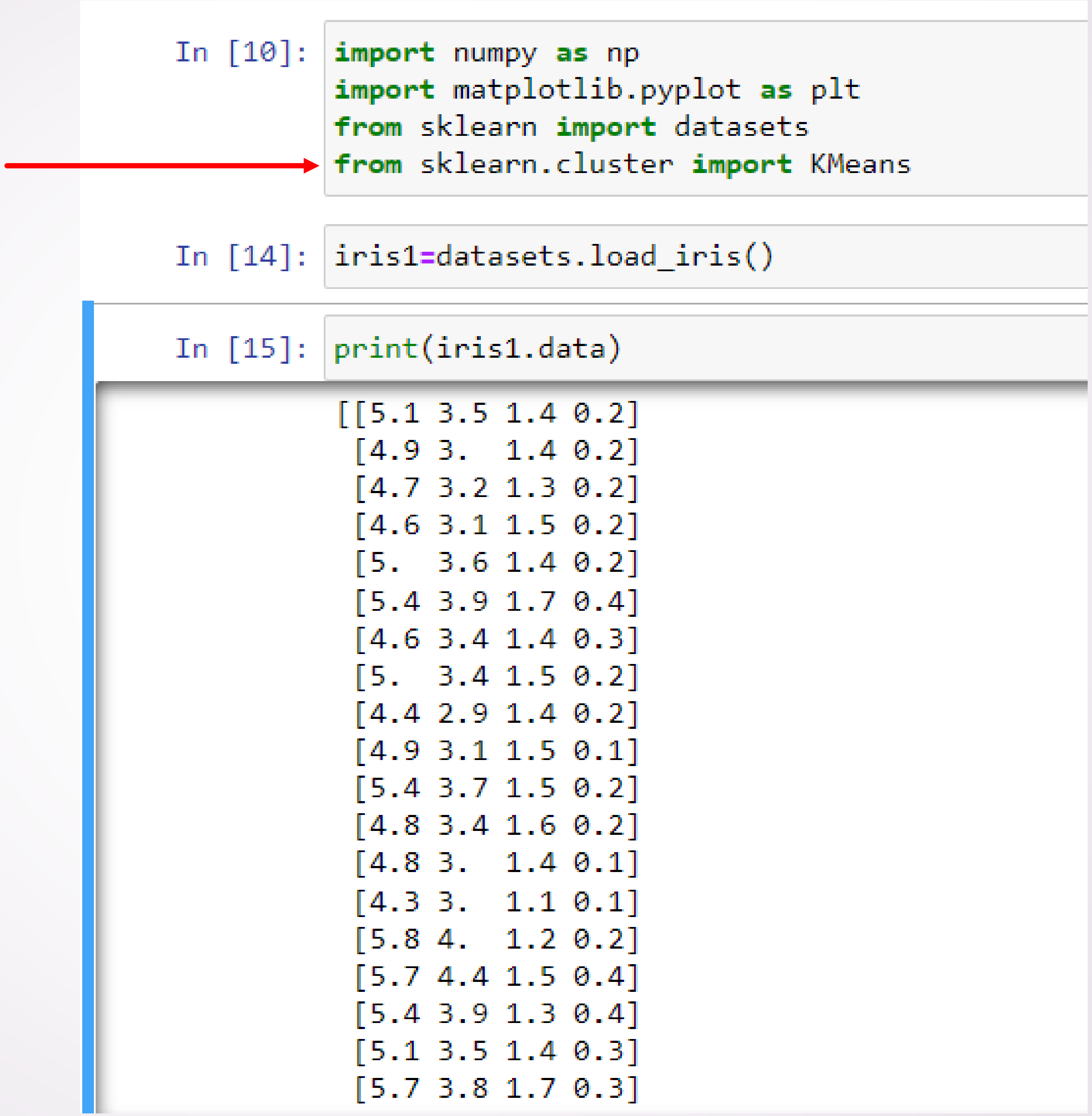
CSIS 3290

UNSUPERVISED LEARNING 1 (KMEANS, DBSCAN)

IN SKLEARN

FATEMEH AHMADI

KMeans



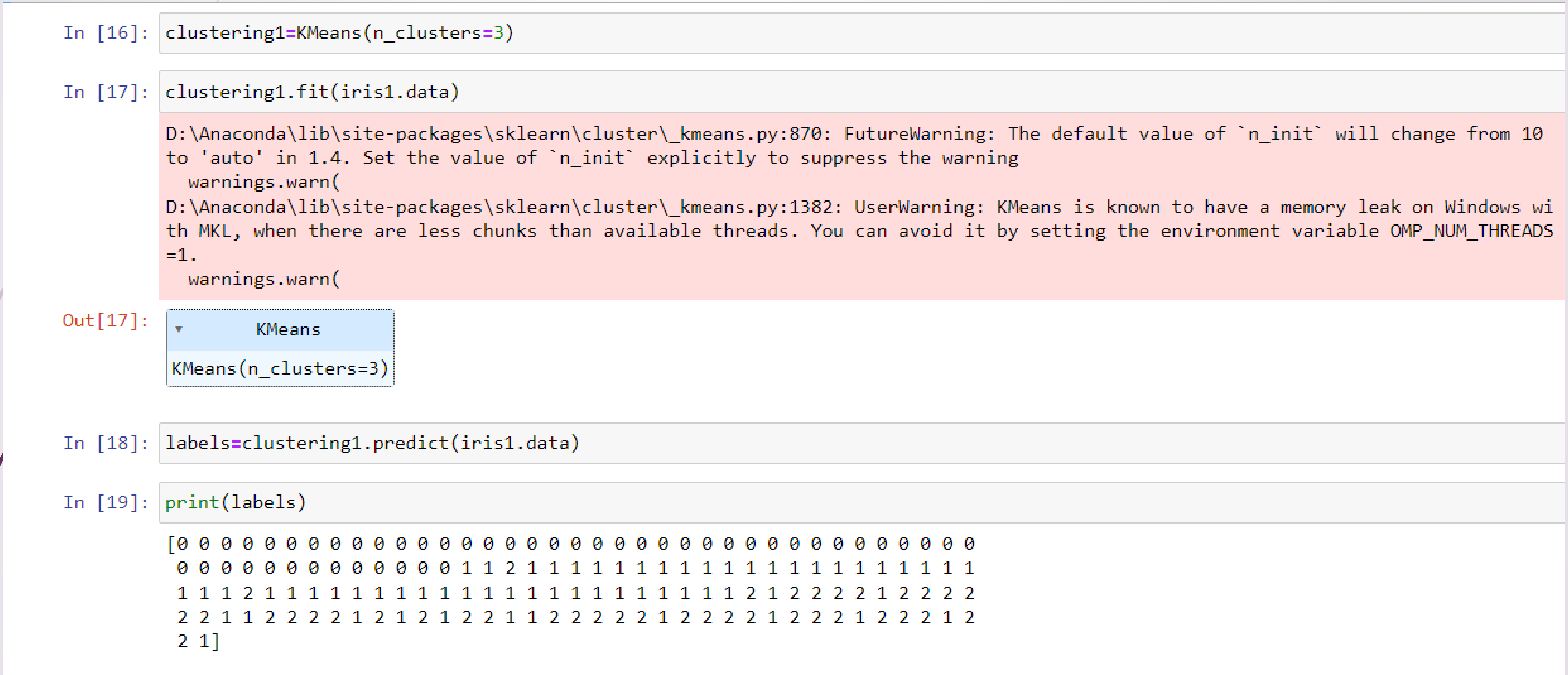
```
In [10]: import numpy as np
         import matplotlib.pyplot as plt
         from sklearn import datasets
         from sklearn.cluster import KMeans
```

```
In [14]: iris1=datasets.load_iris()
```

```
In [15]: print(iris1.data)
```

```
[[5.1 3.5 1.4 0.2]
 [4.9 3.  1.4 0.2]
 [4.7 3.2 1.3 0.2]
 [4.6 3.1 1.5 0.2]
 [5.  3.6 1.4 0.2]
 [5.4 3.9 1.7 0.4]
 [4.6 3.4 1.4 0.3]
 [5.  3.4 1.5 0.2]
 [4.4 2.9 1.4 0.2]
 [4.9 3.1 1.5 0.1]
 [5.4 3.7 1.5 0.2]
 [4.8 3.4 1.6 0.2]
 [4.8 3.  1.4 0.1]
 [4.3 3.  1.1 0.1]
 [5.8 4.  1.2 0.2]
 [5.7 4.4 1.5 0.4]
 [5.4 3.9 1.3 0.4]
 [5.1 3.5 1.4 0.3]
 [5.7 3.8 1.7 0.3]
```

KMeans



```
In [16]: clustering1=KMeans(n_clusters=3)
```

```
In [17]: clustering1.fit(iris1.data)
```

```
D:\Anaconda\lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10
to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
D:\Anaconda\lib\site-packages\sklearn\cluster\_kmeans.py:1382: UserWarning: KMeans is known to have a memory leak on Windows wi
th MKL, when there are less chunks than available threads. You can avoid it by setting the environment variable OMP_NUM_THREADS
=1.
  warnings.warn(
```

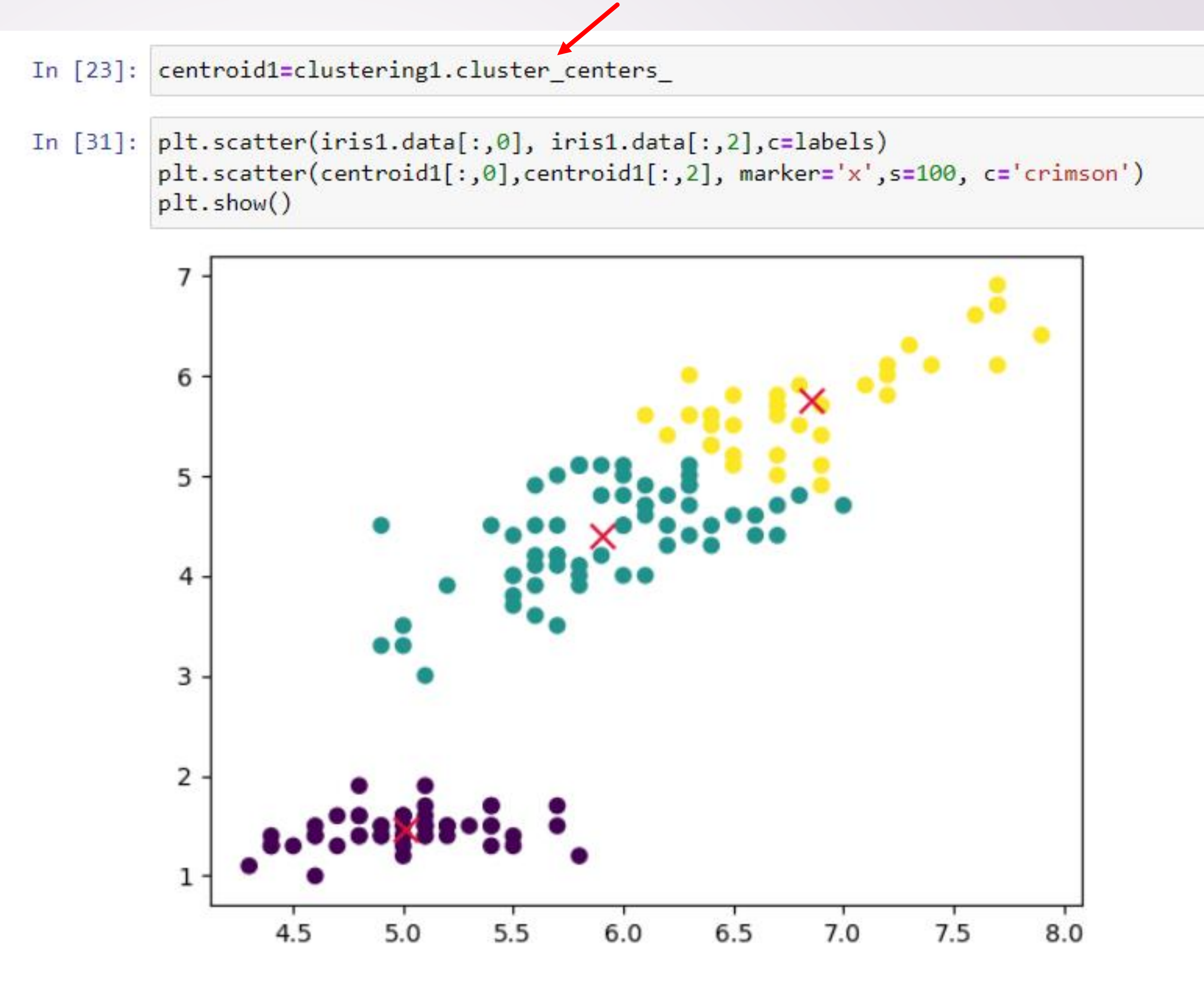
```
Out[17]: ▾ KMeans
         KMeans(n_clusters=3)
```

```
In [18]: labels=clustering1.predict(iris1.data)
```

```
In [19]: print(labels)
```

```
[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 2 2 2 1 2 2 2 2
 2 2 1 1 2 2 2 2 1 2 1 2 1 2 2 1 1 2 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 1 2
 2 1]
```

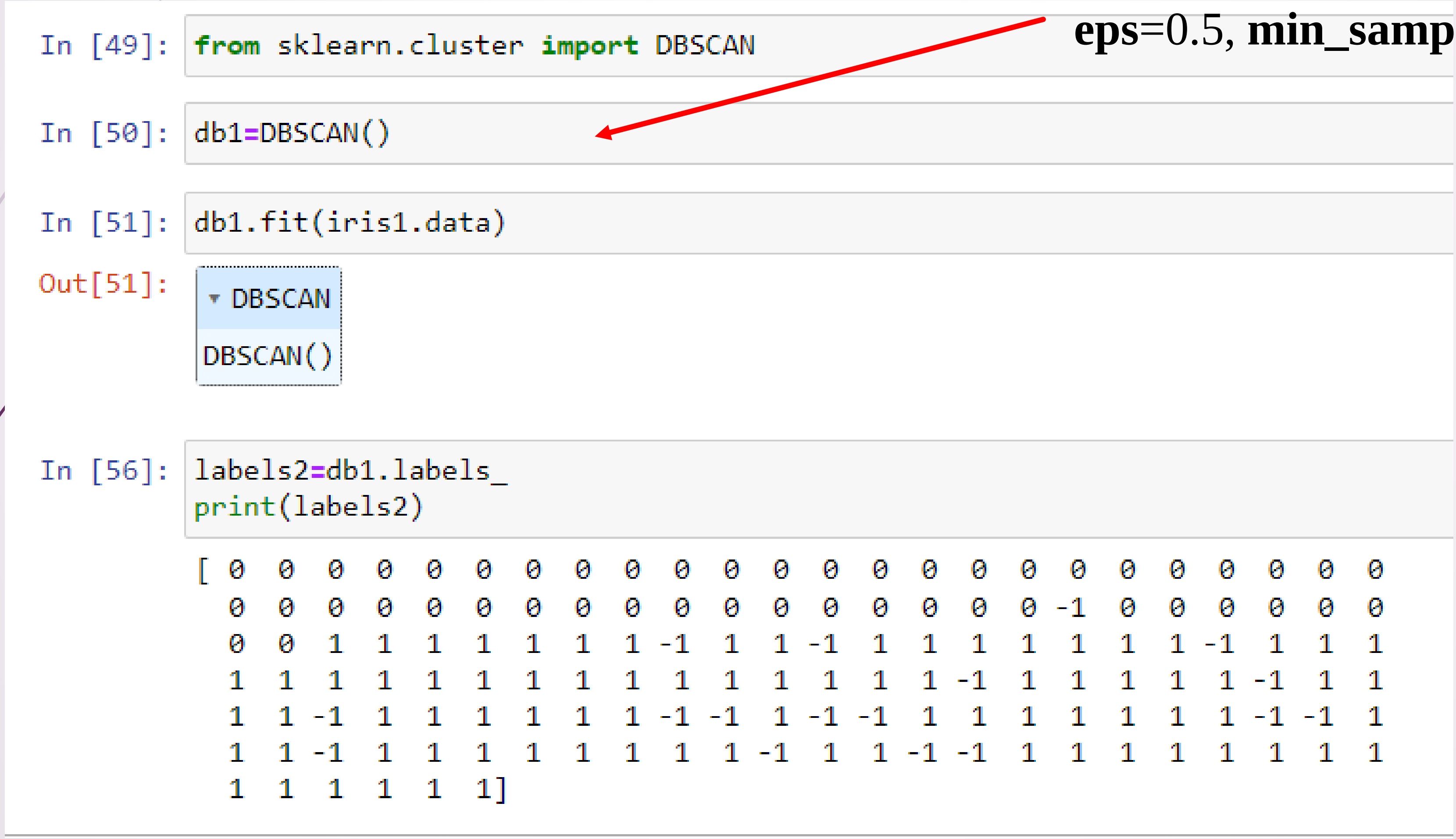
KMeans



```
In [23]: centroid1=clustering1.cluster_centers_
```

```
In [31]: plt.scatter(iris1.data[:,0], iris1.data[:,2],c=labels)
         plt.scatter(centroid1[:,0],centroid1[:,2], marker='x',s=100, c='crimson')
         plt.show()
```

DBSCAN



In [49]:
```
from sklearn.cluster import DBSCAN
```

In [50]:
```
db1=DBSCAN()
```

The default values for **parameters**:
eps=0.5, **min_samples**=5

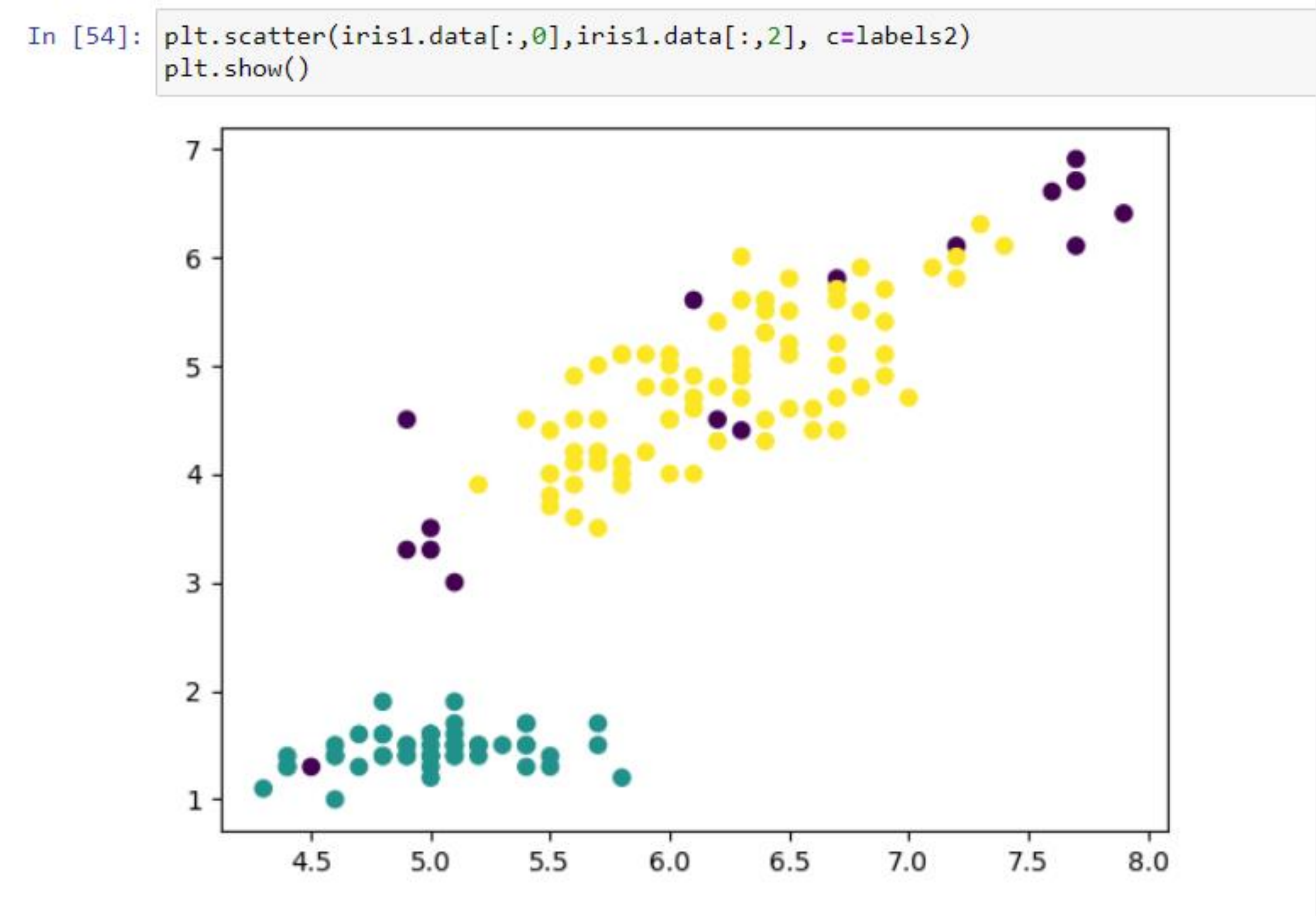
In [51]:
```
db1.fit(iris1.data)
```

Out[51]:
```
▾ DBSCAN
DBSCAN()
```

In [56]:
```
labels2=db1.labels_
print(labels2)
```

```
[ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 -1  0  0  0  0  0  0
  0  0  1  1  1  1  1  1  1 -1  1  1 -1  1  1  1  1  1  1  1 -1  1  1  1
  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1 -1  1  1  1  1  1 -1  1  1
  1  1 -1  1  1  1  1  1  1 -1 -1  1 -1 -1  1  1  1  1  1  1  1 -1 -1  1
  1  1 -1  1  1  1  1  1  1  1  1 -1  1  1 -1 -1  1  1  1  1  1  1  1  1
  1  1  1  1  1  1]
```

DBSCAN

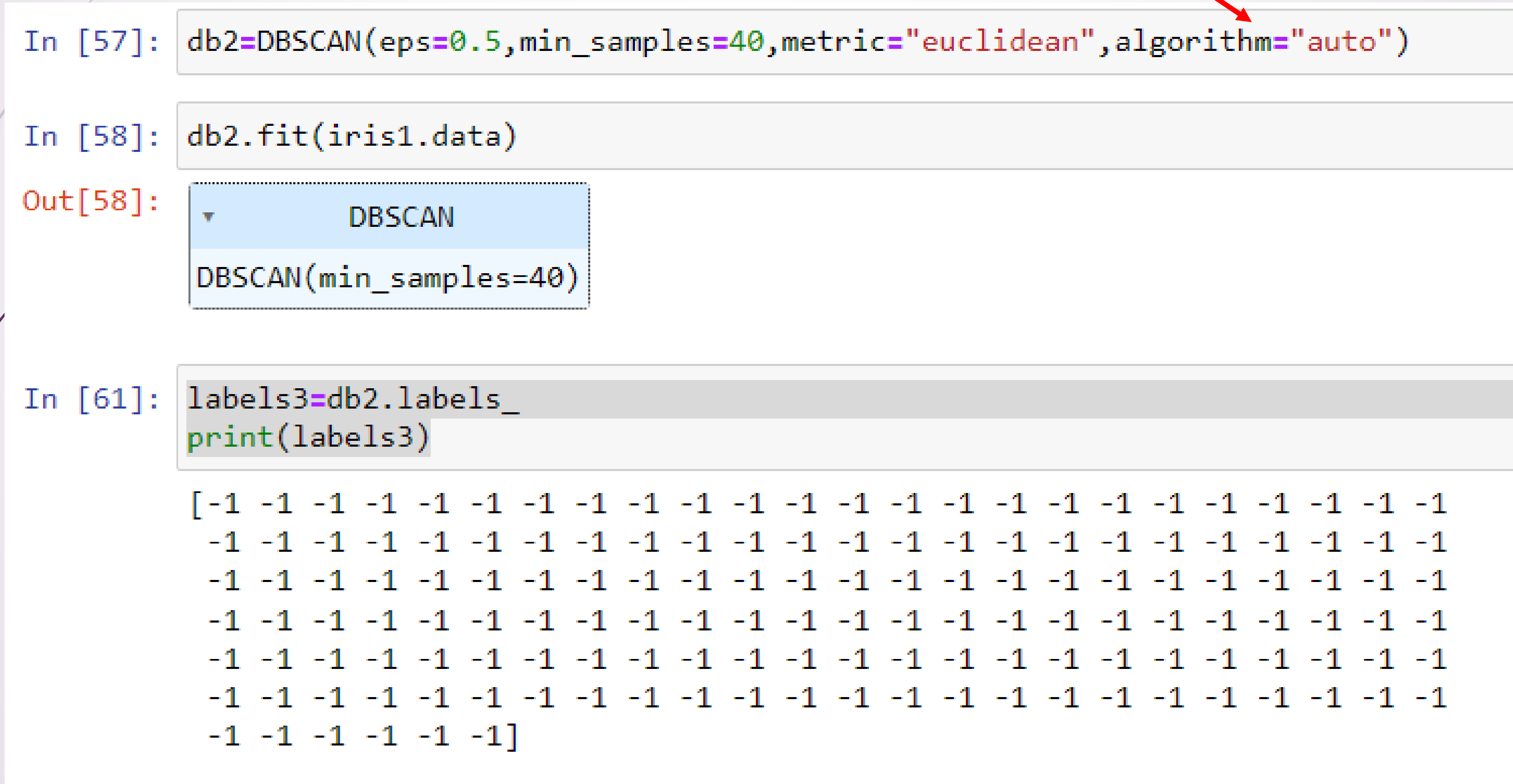


In [54]:

```
plt.scatter(iris1.data[:,0],iris1.data[:,2], c=labels2)
plt.show()
```

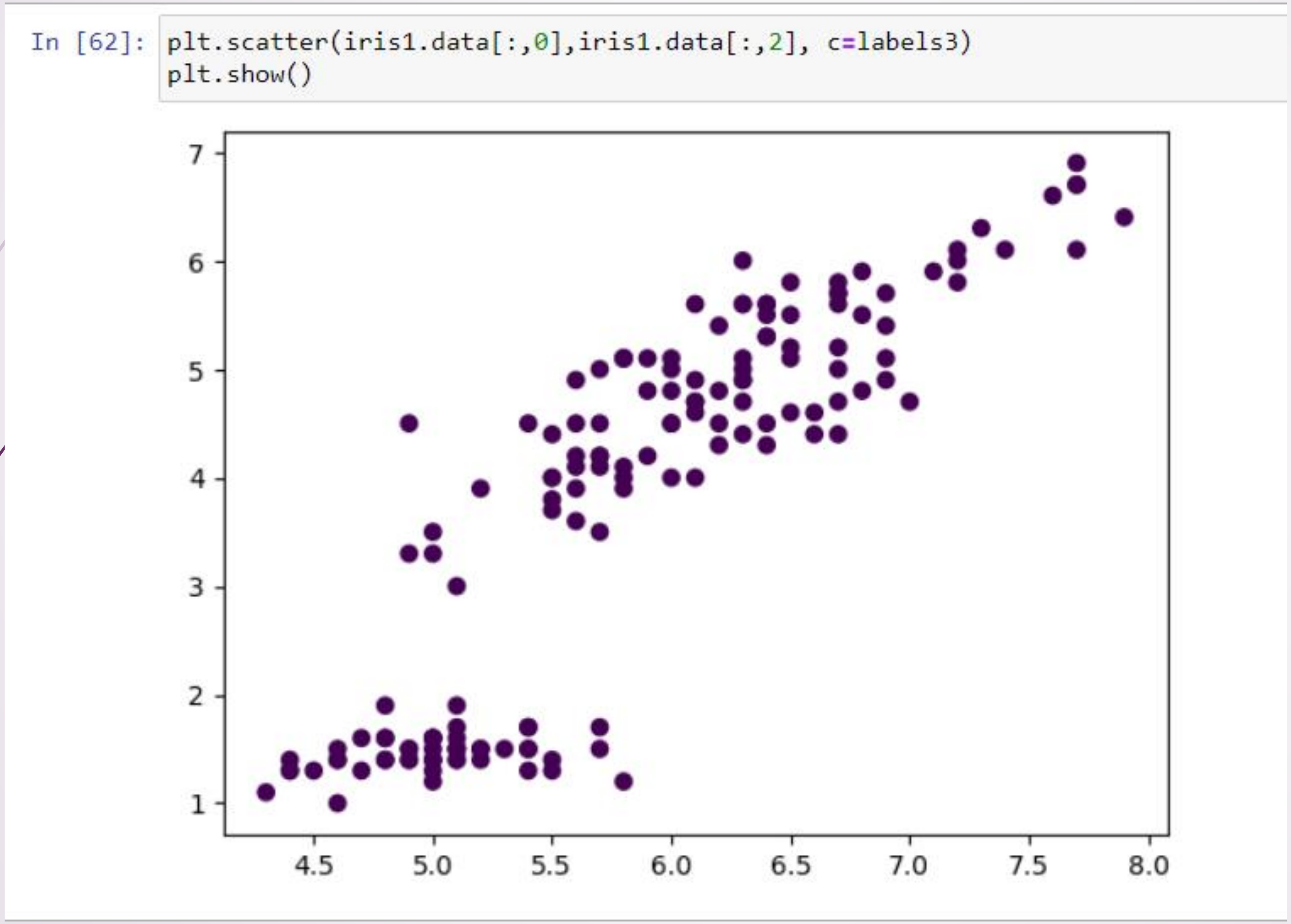
DBSCAN

- **algorithm{‘auto’, ‘ball_tree’, ‘kd_tree’, ‘brute’}, default=’auto’**
- The algorithm to be used by the **NearestNeighbors module** to compute pointwise distances and find nearest neighbors.



```
In [57]: db2=DBSCAN(eps=0.5,min_samples=40,metric="euclidean",algorithm="auto")

In [58]: db2.fit(iris1.data)

Out[58]:
         ▾        DBSCAN
         DBSCAN(min_samples=40)

In [61]: labels3=db2.labels_
         print(labels3)

         [-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
          -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
          -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
          -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
          -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
          -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
          -1 -1 -1 -1 -1 -1]
```

DBSCAN



```
In [62]: plt.scatter(iris1.data[:,0],iris1.data[:,2], c=labels3)
         plt.show()
```