lab8_kmeans

November 15, 2022

K MEANS CLUSTERING

[ ]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.metrics import confusion_matrix
from sklearn.cluster import KMeans
from sklearn.model_selection import train_test_split
from sklearn import preprocessing
```

[ ]:
```
data=pd.read_csv("C:\Coding\ML_python\machine-learning-lab-main\datasets\iris.
↪csv")
```

[ ]:
```
data.head()
```

[ ]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

[ ]:
```
x=data.iloc[:,1:3]
x.head()
```

[ ]:

	sepal_width	petal_length
0	3.5	1.4
1	3.0	1.4
2	3.2	1.3
3	3.1	1.5
4	3.6	1.4

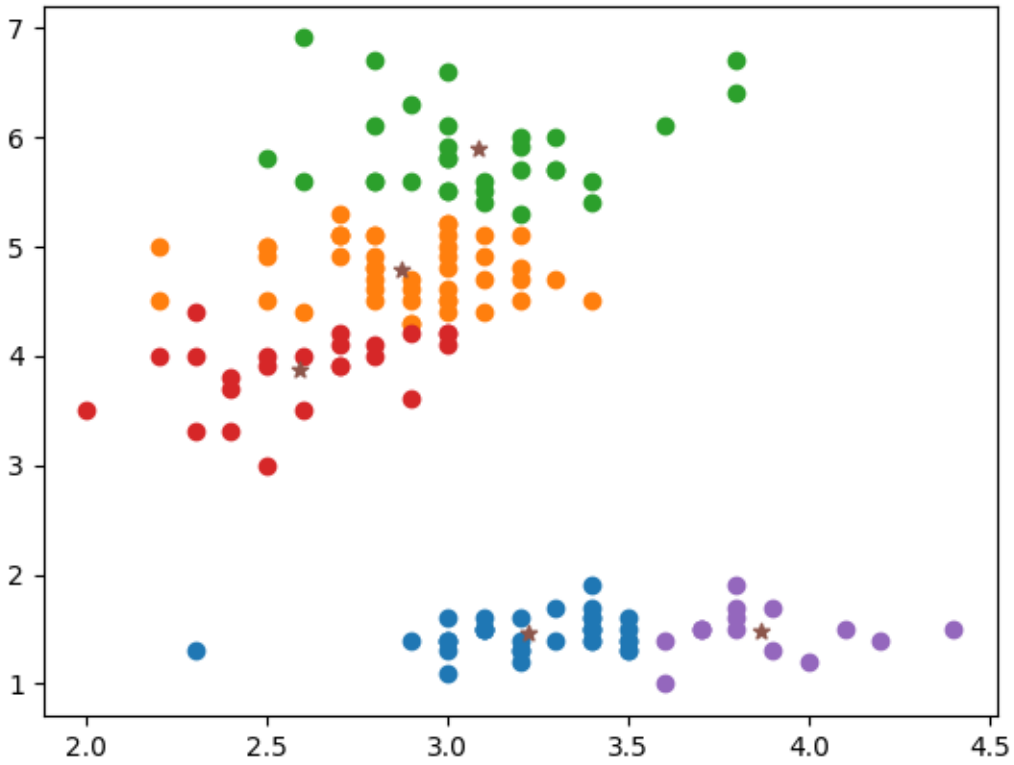
[ ]:
```
#xtrain=x.iloc[:40,:]
#xtest=x.iloc[40:,:]
```

[ ]:
```
k=5
kmeans = KMeans(n_clusters=k, random_state=0).fit(x)
print("cluster centers:")
```

```
print(kmeans.cluster_centers_)
label = kmeans.fit_predict(x)
print("Cluster labels for dataset:\n",label)
for i in range(k):
  cluster = x[label == i]
  plt.scatter(cluster.iloc[:,0],cluster.iloc[:,1])
centers=kmeans.cluster_centers_
cx=[i[0] for i in centers]
cy=[i[1] for i in centers]
plt.scatter(cx,cy,marker='*')
#plt.scatter(x[:][0],)
#print("the testing data belong to clusters:")
#print(kmeans.predict(xtest))
```

```
cluster centers:
[[3.22571429 1.45714286]
 [2.87333333 4.78444444]
 [3.08709677 5.88387097]
 [2.59166667 3.87083333]
 [3.86666667 1.48      ]]
Cluster labels for dataset:
 [0 0 0 0 4 4 0 0 0 0 4 0 0 0 4 4 4 0 4 4 0 4 4 0 0 0 0 0 0 0 0 0 4 4 0 0 0
 0 0 0 0 0 0 0 4 0 4 0 4 0 1 1 1 3 1 1 1 3 1 3 3 3 3 1 3 1 1 3 1 3 1 3 1 1
 1 1 1 1 1 3 3 3 3 1 1 1 1 3 3 3 1 1 3 3 3 3 3 1 3 3 2 1 2 2 2 2 1 2 2 2 1
 1 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 1 1 2 2 1 1 1
 2 1]
```

[ ]: <matplotlib.collections.PathCollection at 0x16a0e1ed0>



K MODES CLUSTERING

[ ]: