```
#!/usr/bin/env python
# coding: utf-8

# # Program 8 : Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-Means  algorithm. Compare the results of these two algorithms and comment on the quality of clustering. You can add Java/Python ML library classes/API in the program.

# In[1]:


from sklearn.mixture import GaussianMixture
from sklearn.cluster import KMeans
import sklearn.metrics as sm
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np


# In[2]:


data=pd.read_csv("8-Kmeans_EM.csv")
print(data.shape)
data.head()
# Getting the values and plotting it
f1 = data['V1'].values
f2 = data['V2'].values
X = np.array(list(zip(f1, f2)))
plt.scatter(f1, f2, c='black', s=7)
plt.title('Dataset')


# In[5]:


gmm = GaussianMixture(n_components=3).fit(X)
labels = gmm.predict(X)
plt.scatter(X[:, 0], X[:, 1], c=labels, s=50, cmap='viridis');
plt.show()
print ("EM predictions")
probs = gmm.predict_proba(X)
print(probs)
print("MEAN:  ",gmm.means_)
print("COVARIANCES: ",gmm.covariances_ )
```

```
# In[6]:


# Number of clusters
k=3
kmeans = KMeans(n_clusters=k)
# Fitting the input data
kmeans = kmeans.fit(X)
# Getting the cluster labels
labels = kmeans.predict(X)
# Centroid values
centroids = kmeans.cluster_centers_
print(centroids)
#plotting the data
plt.title('KMEANS')
plt.scatter(X[:,0], X[:,1], c=labels, cmap='rainbow',s=100)
plt.scatter(kmeans.cluster_centers_[:,0] ,kmeans.cluster_centers_[:,1], marker='*',color='black',s=200)


# In[10]:


gmm = GaussianMixture(n_components=4).fit(X)
labels = gmm.predict(X)
plt.scatter(X[:, 0], X[:, 1], c=labels, s=50, cmap='viridis');
plt.show()
print ("EM predictions")
probs = gmm.predict_proba(X)
print(probs)
print("MEAN: ",gmm.means_ )
print("COVARIANCES : ",gmm.covariances_)


# In[9]:


# Number of clusters
k=4
kmeans = KMeans(n_clusters=k)
# Fitting the input data
kmeans = kmeans.fit(X)
# Getting the cluster labels
labels = kmeans.predict(X)
```

```
# Centroid values
centroids = kmeans.cluster_centers_
print("CENTROIDS")
#plotting the data
plt.title('KMEANS')
plt.scatter(X[:,0], X[:,1], c=labels, cmap='rainbow',s=100)
plt.scatter(kmeans.cluster_centers_[:,0] ,kmeans.cluster_centers_[:,1], marker='*',color='black',s=200)


# In[ ]:
```