

Program 7

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.

Program

```
from sklearn import datasets
from sklearn import metrics
from sklearn.cluster import KMeans
from sklearn.model_selection import train_test_split

iris_dataset = datasets.load_iris()
print()
print("Dataset is loaded successfully\n")

x_train, x_test, y_train, y_test = train_test_split(iris_dataset.data, iris_dataset.target)

print("->Applying K-Means algorithm")
#K-Means Clustering
KMmodel = KMeans(n_clusters=3, n_init=1)
KMmodel.fit(x_train, y_train)
KMmodel.score

print("->Applying EM algorithm")
#EM Clustering
from sklearn.mixture import GaussianMixture
EMmodel = GaussianMixture(n_components=3)
EMmodel.fit(x_train, y_train)
EMmodel.score

print("\nAfter comparing both the algorithms, the Quality of clustering is given below")
print("-----")
print("K-Means clustering:\t", metrics.accuracy_score(y_test, KMmodel.predict(x_test)))
print("EM clustering:\t", metrics.accuracy_score(y_test, EMmodel.predict(x_test)))
print("-----")
```

Result

Dataset is loaded successfully

->Applying K-Means algorithm

->Applying EM algorithm

After comparing both the algorithms, the Quality of clustering is given below

K-Means clustering: 0.02631578947368421

EM clustering: 0.2631578947368421
