

Ex No. 6	Implementation of Expectation-Maximization (EM) Algorithm
Date:	

### Aim

To implement EM algorithm to cluster the data in IRIS data set.

### Data Set

iris.csv (<https://archive.ics.uci.edu/ml/machine-learning-databases/iris/>)

### Definition

### EM Algorithm

*Expectation-Maximization algorithm* can be used for the latent variables (variables that are not directly observable and are actually inferred from the values of the other observed variables) too in order to predict their values with the condition that the general form of probability distribution governing those latent variables is known to us. This algorithm is actually at the base of many unsupervised clustering algorithms in the field of machine learning.

## Procedure

Open PyCharm Community Edition.

Go to File menu → New Project → Specify the project name → Press “Create” button.

Right Click on Project name → New → Python File → Specify the file name → Press Enter.

Type the following codes. Right click on file name or coding window → Select “Run” to view the result.

## EM.py

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

names = ['Sepal_Length', 'Sepal_Width', 'Petal_Length', 'Petal_Width', 'Class']

dataset = pd.read_csv("C:/Users/2mca2/Downloads/iris.csv", names=names)

X = dataset.iloc[:, :-1]

label = {'Iris-setosa': 0, 'Iris-versicolor': 1, 'Iris-virginica': 2}

y = [label[c] for c in dataset.iloc[:, -1]]

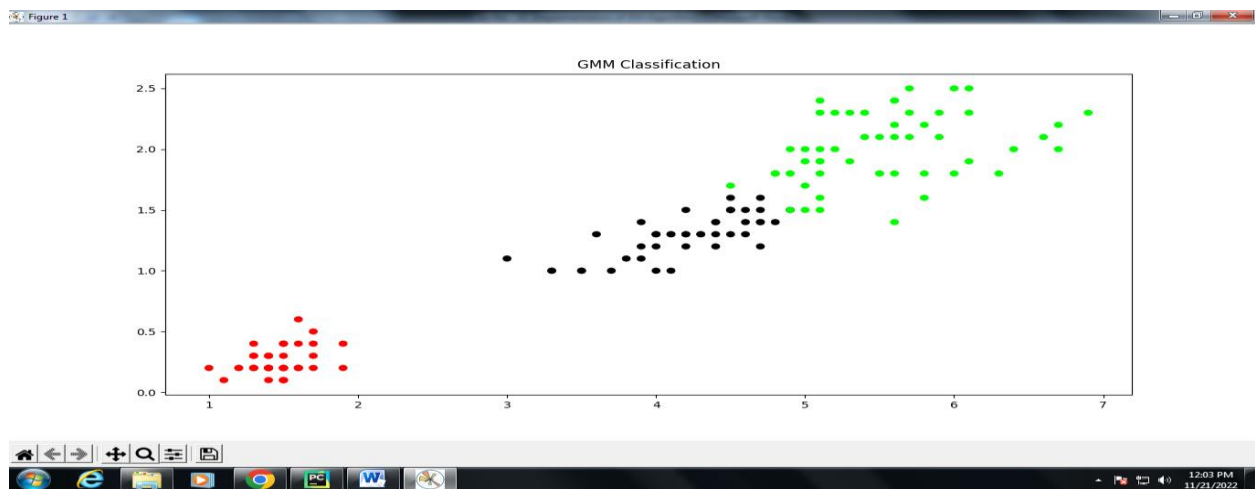
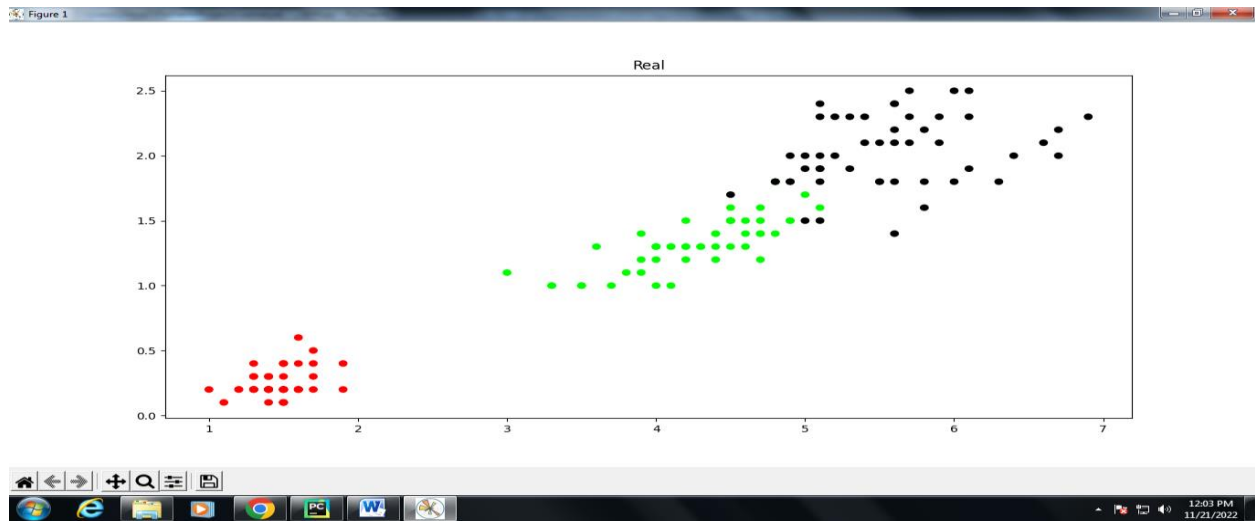
plt.figure(figsize=(14,7))
colormap=np.array(['red','lime','black'])

# REAL PLOT

plt.title('Real')
plt.scatter(X.Petal_Length,X.Petal_Width,c=colormap[y])
plt.show()
# GMM PLOT
gmm=GaussianMixture(n_components=3, random_state=0).fit(X)
y_cluster_gmm=gmm.predict(X)

plt.title('GMM Classification')
plt.scatter(X.Petal_Length,X.Petal_Width,c=colormap[y_cluster_gmm])
plt.show()
print("The accuracy score of EM: ", metrics.accuracy_score(y, y_cluster_gmm))
print("The Confusion matrix of EM:\n ", metrics.confusion_matrix(y, y_cluster_gmm))
```

## Output



C:\Users\2mca2\PycharmProjects\sumaiya\venv\Scripts\python.exe

C:/Users/2mca2/PycharmProjects/sumaiya/EM.py

The accuracy score of EM: 0.36666666666666664

The Confusion matrix of EM:

[[50 0 0]

[ 0 5 45]

[ 0 50 0]]

Process finished with exit code 0

## **Result**

Thus, EM algorithm has been implemented to cluster the data in IRIS dataset.