EX.NO: 2	Study of Gaussian Mixture Model
DATE: 10.03.2022	

AIM:
To study and implement Gaussian Mixture Model.

Procedure:

1. Load the iris dataset from datasets package. To keep things simple, take only first two columns (i.e sepal length and sepal width respectively).
2. Now plot the dataset.
3. Fit the data as a mixture of 3 Gaussians.
4. Then do the clustering, i.e assign a label to each observation. Also find the number of iterations needed for the log-likelihood function to converge and the converged log-likelihood value.
5. Print the converged log-likelihood value and no. of iterations needed for the model to converge.

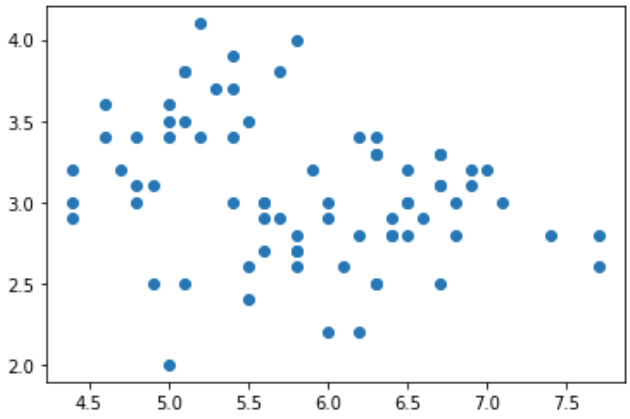
1. Loading the dataset

Program:

```
# imports
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from pandas import DataFrame
from sklearn import datasets
from sklearn.mixture import GaussianMixture

# load the iris dataset
iris = datasets.load_iris()

# select first two columns
X = iris.data[:, :2]

# turn it into a dataframe
d = pd.DataFrame(X)

# plot the data
plt.scatter(d[0], d[1])
```

2. Plotting the dataset

Output:

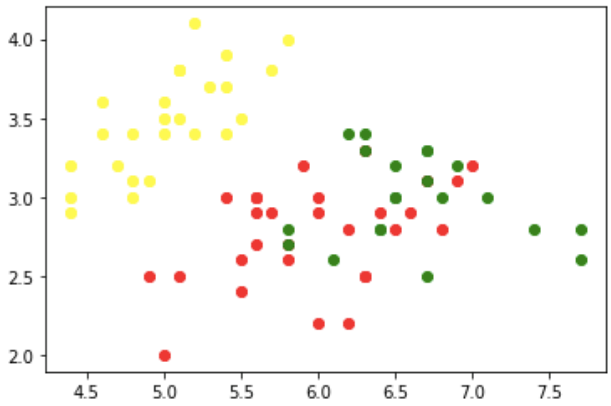


3. Fitting the data as mixture of 3 Guassians

Program:

```
gmm = GaussianMixture(n_components = 3)

# Fit the GMM model for the dataset
# which expresses the dataset as a
# mixture of 3 Gaussian Distribution
gmm.fit(d)

# Assign a label to each sample
labels = gmm.predict(d)
d['labels']= labels
d0 = d[d['labels']== 0]
d1 = d[d['labels']== 1]
d2 = d[d['labels']== 2]

# plot three clusters in same plot
plt.scatter(d0[0], d0[1], c ='r')
plt.scatter(d1[0], d1[1], c ='yellow')
plt.scatter(d2[0], d2[1], c ='g')
```

Output:



4. Finding the log-likelihood

Program:

Print(gmm.lower_bound_)

Output:

```
In [20]: print(gmm.lower_bound_)
         -1.1531810973187941
```

5. Finding converged log-likelihood value and no. of iterations needed for the model to converge.

Program:

print(gmm.n_iter_)

Output:

```
In [21]: print(gmm.n_iter_)
         4
```

RESULT:

Hence, it needed 7 iterations for the log-likelihood to converge. If more iterations are performed, no appreciable change in the log-likelihood value, can be observed.