

HW2&3_question9_GMM

March 25, 2021

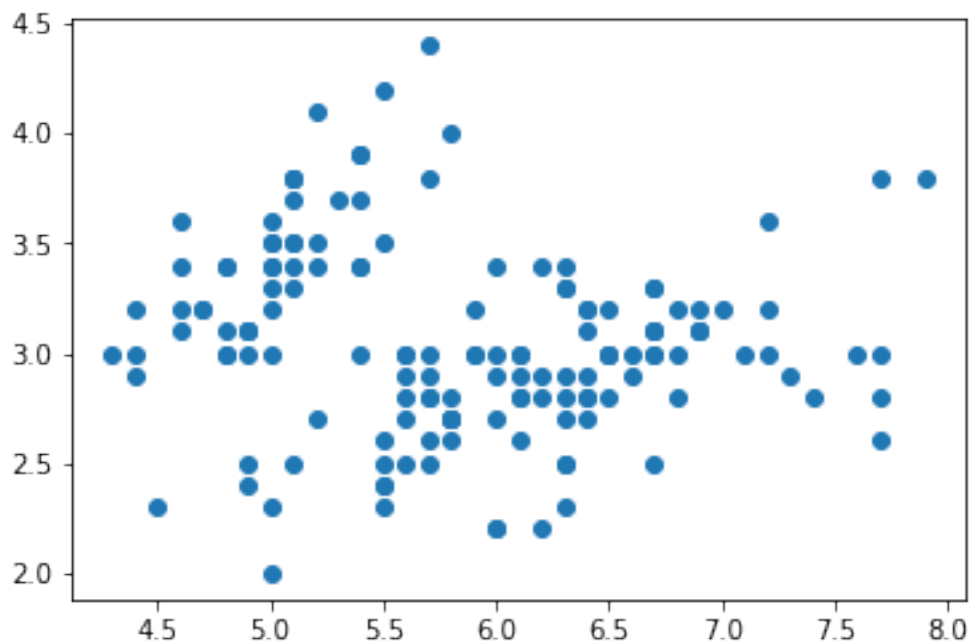
```
[1]: import pandas as pd

# read the iris dataset which is csv format
col_names = ["sepal_length", "sepal_width", "petal_length", "petal_width", "species"]
iris = pd.read_csv("iris.data", header=None, names=col_names)
```

```
[2]: # map iris class name to number
iris_class = {'Iris-setosa':0, 'Iris-versicolor':1, 'Iris-virginica':2}
iris['species_tag'] = [iris_class[i] for i in iris.species]
```

```
[3]: import matplotlib.pyplot as plt
#split data into attributes and target/label
iris_attrs = iris.drop(['species', 'species_tag'], axis=1)
plt.scatter(iris_attrs['sepal_length'], iris_attrs['sepal_width'])
```

```
[3]: <matplotlib.collections.PathCollection at 0x7feb5a0b13d0>
```



```
[4]: from sklearn.model_selection import train_test_split
from sklearn import mixture

# create GMM
gmm = mixture.GaussianMixture(n_components=2)

# fit the model on the training data
gmm.fit(iris_attrs)
# make prediction
labels = gmm.predict(iris_attrs)
iris_attrs['labels'] = labels
iris_0 = iris_attrs[iris_attrs['labels']==0]
iris_1 = iris_attrs[iris_attrs['labels']==1]

# get results of clustering by GMM, x: sepal_length y: sepal_width
plt.scatter(iris_0['sepal_length'], iris_0['sepal_width'], color='r')
plt.scatter(iris_1['sepal_length'], iris_1['sepal_width'], color='b')
```

[4]: <matplotlib.collections.PathCollection at 0x7feb3408b6d0>

