Machine Learning Algorithms: From Math to Code
Assignment for Clustering Analysis

Kang Wang, Yuheng Zhang

1 Problem Set 1

1.1 Ch15_1.m

Implement GMM clustering and apply it with the K-means method to the dataset used in Example 15.3 (generated by the code provided in the example, see Figure 1). Visualize the results by both GMM and K-means methods will all data points color coded, and compare the confusion matrices of both methods to confirm GMM achieves better results than K-means in this particular case.

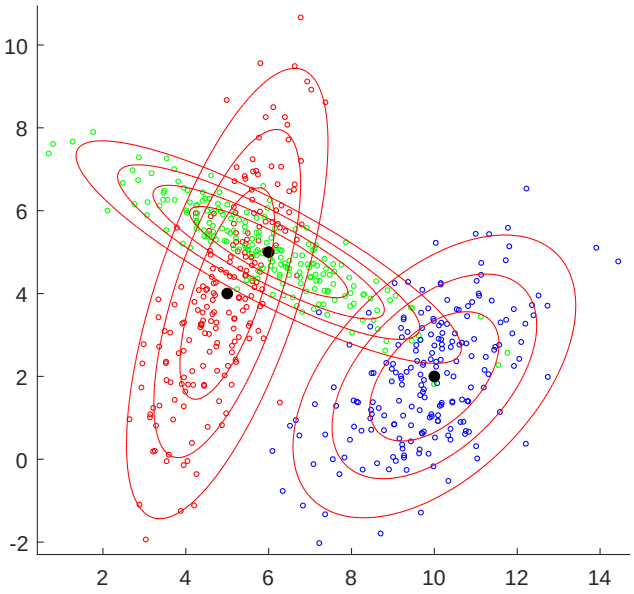

Figure 1: myData0

Complete the missing code and add appropriate comments (to key variables, steps and formulas of the algorithm) for Ch15_1.m, and submit it with a report describing your results in a compressed .zip file on canvas.

After you finish Ch15_1.m, you should get output in command line in the following form, and plots will be shown by matlab.

```
>> Ch15_1
Iteration times = ...
Cm =
...

Percentage error: ...

Iteration times = ...
Cm =
...

Percentage error: ...
```

In the report, you should

1. Include generated plots.
2. Include output in command line.

1.2 Ch15_2.m

Apply the K-means and GMM method to the 4-D Iris dataset.

1. Show your clustering result in confusion matrix;
2. Visualize the color coded data points in 2-D space (Use the PCA method to visualize the first two principal components. The funcion `myPlot` in Ch15_1.m will help you do this).

Write the code Ch15_2.m by yourself through modifying Ch15_1.m. You will need the following function to read in Iris dataset.

```
function [X,Ki]=Irisdata
data_path = 'iris.txt';
data = load(data_path);
X = data(:,1:4);
Ki = [50 50 50];
end
```

After you finish Ch15_2.m, you should get output in command line in the form similar to that of Ch15_1.m, and plots will be shown by matlab.

In the report, you should

1. Include generated plots.
2. Include output in command line.

Notes

- These Two problems should be included in a single report with headings.
- Source codes and report should be compressed into a single .zip file named **Group_xx.zip** and handed on canvas before **next monday midnight, July 24 23:59**.