Assignment Two

Problem:

1) Please build a Gaussian mixture model (GMM) to model the data in file TrainingData_GMM.csv. Note that the data is composed of 4 clusters, and the model should be trained by expectation maximization (EM) algorithm.

2) Based on the GMM learned above, assign each training data point into one of 4 different clusters

Requirements:

Summarize your experimental results in a report. Package your report and code into one file under the name of ***your student no. + your name***.tar (or .zip) and then submit to ftp://172.18.166.144 , with login: students and password: stu123

The report should include:

1）Show how the log-likelihood evolves as the training proceeds

2）The learned mathematical expression for the GMM model after training on the given dataset

3）Randomly select 500 data points from the given dataset and plot them on a 2-dimensional coordinate system. Mark the data points coming from the same cluster (using the results of Problem 2) with the same color.

4）Some analyses on the impacts of initialization on the converged values of EM algorithm

5）Some analyses on the results you obtained

Due: Dec. 29, 2018