ISTD 50.570 (Machine Learning): Assignment 3

Grading Policy and Due Date

- You are required to submit 1) a report that summarizes your experimental results and findings, based on each of the following question asked; 2) your implementation of the algorithms.
- Submit your assignment report and code to the Dropbox folder shared with you.
- This assignment is an individual assignment. Discussions amongst yourselves are allowed and encouraged, but you should write your own code and report.
- Submit your assignment to the Dropbox folder by 11:59 PM on Friday 15 April 2016. **You are strongly suggested to start this assignment early.** Late submissions will be heavily penalized (20% deduction per day).

Task 1: Mixture of Gaussians & EM

Download the file `em-data.zip` from the course web page and unzip it.

1. (10 pts) Assume the data comes from a mixture of $k = 2$ spherical Gaussian distributions. Now you would like to estimate the parameters for the two Gaussian distributions. Discuss what are the model parameters to be estimated, and what are the hidden variables for the mixture model.
2. (10 pts) Now, use the hard EM algorithm discussed in class to estimate the model parameters. Discuss how you initialize the model and when to terminate the parameter estimation process. Plot the value of the (log) joint likelihood associated with the data as a function of the number of EM iterations. Discuss the behavior of the curve. Think of a way to visualize the final result (*e.g.*, you may use graph to show (partial) membership of each individual data point).

   (Hint: to overcome potential numerical overflow issues, you might find the logsumexp trick useful. See this page: `https://en.wikipedia.org/wiki/LogSumExp`)
3. (10 pts) Now, instead of using the hard EM algorithm, let us use the soft EM algorithm, and repeat the above question.

Task 2: Theorem Proofing

(10 pts) Being able to work out neat derivation steps is a very critical skill in machine learning research. We have not yet got a chance to work on derivations in the previous assignments. In this assignment, let us work on the proof of the following theorem as an exercise.

Theorem *For any vectors* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N \in \mathbf{R}^m$*, any positive scalars* $\lambda_0, \lambda_1, \ldots, \lambda_N \in \mathbf{R}^+$*, let* $\mathbf{v}_0 = (\sum_{i=1}^N \lambda_i \mathbf{v}_i)/\lambda_0$*, then the following always holds:*

$$\lambda_0 ||\mathbf{v}_0||^2 + \sum_{i=1}^N \lambda_i ||\mathbf{v}_i||^2$$
$$= \sum_{i=1}^N \eta_{0,i} ||\mathbf{v}_i + \mathbf{v}_0||^2 + \sum_{1 \leq j < k \leq N} \eta_{j,k} ||\mathbf{v}_j - \mathbf{v}_k||^2$$

where

$$\eta_{i,j} = \frac{\lambda_i \lambda_j}{\sum_{l=0}^N \lambda_l}$$

for all $0 \leq i < j \leq N$*.*

Note that it is encouraged for you to work out a short proof rather than a long proof. Bonus points (up to 5 points) will be given to proofs with neat/short steps.