Statistics 5114: The EM Algorithm

For each homework assignment, turn in at the beginning of class on the indicated due date. Late assignments will only be accepted with special permission. Write each problem up *very* neatly (LaTeX is preferred). Show all of your work.

Consider the p-dimensional Gaussian Mixture Model:

$$x_i \sim \sum_{k}^{K} \pi_k p(x|\mu_k, \Sigma_k, C_k), \qquad \text{for } i = 1, \ldots, N, \tag{1}$$

where $p(x|\mu_k, \Sigma_k) = \frac{1}{\pi^{p/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)'\Sigma_k^{-1}(x-\mu_k)}$.

Recall that the EM algorithm for fitting model (1) iterates over the following updates:

For $t = 1, \ldots, T$

1. $\pi_{i,k}^{(t)} = p(x_i \in C_k|\mu_k^{(t-1)}, \Sigma_k^{(t-1)})$
2. $\mu_k^{(t)} = \sum_{i=1}^N \pi_{i,k}^{(t)} x_i / \sum_{i=1}^N \pi_{i,k}^{(t)}$
3. $\Sigma_k^{(t)} = \sum_{i=1}^N \pi_{i,k}^{(t)} (x_i - \mu_k^{(t)})(x_i - \mu_k^{(t)})' / \sum_{i=1}^N \pi_{i,k}^{(t)}$

part a

Given $\pi_{i,k}$s, show that the M.L.E for μ_ks are given by $\sum_{i=1}^N \pi_{i,k} x_i / \sum_{i=1}^N \pi_{i,k}$.

part b

Given μ_ks and $\pi_{i,k}$s, show that the M.L.E for Σ_ks are given by $\sum_{i=1}^N \pi_{i,k}(x_i - \mu_k)(x_i - \mu_k)' / \sum_{i=1}^N \pi_{i,k}$ (Given μ_ks).

part c

Implement the EM algorithm. For various settings of μs, Σs, and πs, show how well the EM algorithm works in identifying the model described by model (1). The parameter settings are for you to pick, however, your simulation study should be thorough. Note: the question of how well the method works is a rather open ended question. You'll need to choose a metric for answering this question as well.