625.664 Computational Statistics

Problem Set 3

Associated Reading: Chapter 2: 2.2.2-2.2.2.2
Chapter 4: Introduction - Section 4.2

Complete the problems either by hand or using the computer and upload your final document to the Blackboard course site. All final submittals are to be in PDF form. Please document any code used to solve the problems and include it with your submission.

1. Problem 2.5. Only do parts (a), (b), (c), and (e).

2. Consider the following mixture of two normal densities:

$$\mathbf{p}(x;\theta) = \pi\phi(x;\mu_1,\sigma_1^2) + (1-\pi)\phi(x;\mu_2,\sigma_2^2)$$

where $\theta = (\pi, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ is the collection of all the parameters driving the model, with π denoting the the mixing proportion (and not the usual number), and ϕ is the standard density for a normal random variable. Direct use of maximum likelihood is made difficult here because of the expression of the likelihood function. So, we wish to simplify this problem by use of the EM algorithm.

Let Z_i be the random variable that denotes the class membership of of an observed x_i in a given sample. Clearly, z_i is the missing data, with $Z_i \in \{0,1\}$ and $Z_i \sim \text{Bernoulli}(\pi)$. The complete-data density for $Y = (X, Z)$ is given by

$$\mathbf{p}(x_i, z_i|\theta) = [\pi\phi(x;\mu_1,\sigma_1^2)]^{z_i}[(1-\pi)\phi(x;\mu_2,\sigma_2^2)]^{(1-z_i)}$$

(a) Show that the complete-data log-likelihood function is

$$l(\theta;x,z) = \log(\pi)\sum_{i=1}^{n} z_i + \sum_{i=1}^{n} z_i \log(\phi(x_i;\mu_1,\sigma_1^2)) + \log(1-\pi)\left(n - \sum_{i=1}^{n} z_i\right)$$
$$+\sum_{i=1}^{n}(1-z_i)\log(\phi(x_i;\mu_2,\sigma_2^2)).$$

For the Expectation step of the EM algorithm, we need $Q(\theta|\theta^{(k)}) = E[l(\theta;y)|x,\theta^{(k)}]$.

(b) Find $Q(\theta|\theta^{(k)})$ in terms of $E[Z_i|x_i,\theta^{(k)}]$, again it is best to leave your answer as the sum of four main terms. Also, if is fine to leave your answer in terms of the function $\phi(x_i;\mu_1^{(k)},(\sigma_1^2)^{(k)})$.

(c) Show that, given $\theta^{(k)}$ and x_i,

$$E[Z_i|x_i,\theta^{(k)}] = P[Z_i = 1|x_i,\theta^{(k)}] = \frac{\pi^{(k)}\phi(x_i;\mu_1^{(k)},(\sigma_1^2)^{(k)})}{\pi^{(k)}\phi(x_i;\mu_1^{(k)},(\sigma_1^2)^{(k)}) + (1-\pi^{(k)})\phi(x_i;\mu_2^{(k)},(\sigma_2^2)^{(k)})}$$

For the Maximization step of the EM algorithm, we need to find the new set of parameters $\theta^{(k+1)}$ that maximizes $Q(\theta|\theta^{(k)})$. This can be done in a straightforward manner by solving

$$\frac{\delta Q(\theta|\theta^{(k)})}{\delta\theta} = 0$$

for each of the parameters in $\theta = (\pi, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$. To simplify your notation let $\eta_i^{(k)} = E[Z_i|x_i, \theta^{(k)}]$ and $\eta^{(k)} = \sum_{i=1}^n \eta_i^{(k)}$.

(d) First show that

$$\pi^{(k+1)} = \frac{\eta^{(k)}}{n} = \frac{1}{n}\sum_{i=1}^n E[Z_i|x_i, \theta^{(k)}].$$

(e) Now, use the fact that

$$\log\phi(x_i;\mu_j,\sigma_j^2) \propto -\frac{1}{2}\log(\sigma_j^2) - \frac{1}{2}(\sigma_j^2)^{-1}(x_i-\mu_j)^2$$

for $j = 1, 2$ to show

$$\mu_1^{(k+1)} = \frac{1}{\eta^{(k)}}\sum_{i=1}^n \eta_i^{(k)} x_i,$$

$$(\sigma_1^2)^{(k+1)} = \frac{1}{\eta^{(k)}}\sum_{i=1}^n \eta_i^{(k)}(x_i - \mu_1^{(k+1)})^2,$$

$$\mu_2^{(k+1)} = \frac{1}{n-\eta^{(k)}}\sum_{i=1}^n (1-\eta_i^{(k)}) x_i,$$

$$(\sigma_2^2)^{(k+1)} = \frac{1}{n-\eta^{(k)}}\sum_{i=1}^n (1-\eta_i^{(k)})(x_i - \mu_2^{(k+1)})^2.$$

(f) Suppose that you have observed the sample $x_1, x_2, \ldots, x_n$ from the above distribution. Write the psuedocode required to estimate θ using the EM algorithm. Be sure to include some form of convergence criteria.