CALD 10–702
Statistical Approaches to Learning and Discovery

Assignment 2

February 10, 2003

Due in class on Wednesday, February 19.

Instructions: This assignment includes three problems involving the EM algorithm. Please submit written answers to each question, either in handwritten or printed form. For the problems that involve computing, you do not need to hand in source code, but you should include a description of your implementation.

Problem 1. *EM and Deconvolution*

Suppose that $Y_1 \sim \exp(\theta_1)$ and $Y_2 \sim \exp(\theta_2)$ are independent exponential random variables. Then $X = Y_1 + Y_2$ is distributed according to $\exp(\theta_1) \star \exp(\theta_2)$, where the convolution $f \star g$ of densities f and g is given by $f \star g(x) = \int g(y)\, f(x-y)\, dy$.

(a) Give a formula for the density f_{θ_1,θ_2} of X.

(b) Suppose that $x_1, x_2, \ldots, x_n$ are iid samples from f_{θ_1,θ_2}. Give expressions for the appropriate complete data and incomplete data likelihoods.

(c) Derive the E-step and M-step and give explicit expressions for the parameter updates for the maximum likelihood estimates of θ_1 and θ_2.

(d) Implement the EM algorithm derived above, and run it on the data contained in the file `/afs/cs/academic/class/10702/data/assign2.exponential.dat`, which is a sample of 500 points from f_{θ_1,θ_2}. Report your estimate $(\widehat{\theta}_1, \widehat{\theta}_2)$, and give plots of

1. The density $f_{\widehat{\theta}_1,\widehat{\theta}_2}$ against a histogram of the data.
2. The incomplete data likelihood as a function of iteration.
3. The estimated parameters θ_1 and θ_2 as a function of iteration.

Note: the following integration-by-parts formula may come in handy:

$$\int_0^x y e^{\alpha y}\, dy = \frac{1}{\alpha}\left(x e^{\alpha x} - \int_0^x e^{\alpha y}\, dy \right)$$

Problem 2. *Rate of Convergence*

In this problem you will analyze the rate of convergence of EM in a special case, where the complete data is from a one-dimensional exponential family. Qualitatively, the conclusion is that the convergence is linear in the amount of "missing" information at the MLE. That is, if z denotes the complete data and x is the incomplete (observed) data, then the convergence is linear in the ratio of Fisher informations $I_{z\,|\,x}(\widehat{\theta})/I_z(\widehat{\theta})$.

Let the incomplete data density be $g(x\,|\,\theta)$, and suppose that the complete data z comes from a one-dimensional exponential family with natural parameter θ and sufficient statistic $t(z)$. Thus, the complete data density is written as

$$f(z\,|\,\theta) \;=\; b(z)\,\exp\left(\theta t(z) - \psi(\theta)\right)$$

and the conditional density of z given x is

$$h(z\,|\,x,\theta) \;=\; \frac{f(z\,|\,\theta)}{g(x\,|\,\theta)}$$

(a) Show how the E-step can be carried out, by deriving an expression for the Q function

$$Q(\theta,\theta'\,|\,x) \;=\; \int h(z\,|\,x,\theta')\log f(z\,|\,\theta)\,dz$$

in terms of θ, θ', ψ, b and t.

(b) Derive the M-step by giving an equation for the value of θ that maximizes the Q function for $\theta' = \theta^{(k)}$, in the k-th iteration. (Note: you will only be able to give an *implicit* equation for the update $\theta^{(k+1)}$; in general, this M-step must be computed numerically.)

(c) Show that as $k \longrightarrow \infty$,

$$\frac{\theta^{(k+1)} - \widehat{\theta}}{\theta^{(k)} - \widehat{\theta}} \;=\; \frac{I_{z\,|\,x}(\widehat{\theta})}{I_z(\widehat{\theta})} + o(1)$$

where $\widehat{\theta}$ is the MLE, I_z is the Fisher information of z, and $I_{z\,|\,x}$ is the Fisher information of z conditioned on x. (Hint: use the linear approximation $E_\theta[t] = E_{\widehat{\theta}}[t] + I_z(\widehat{\theta})(\theta - \widehat{\theta}) + o(|\theta - \widehat{\theta}|)$, and the corresponding approximation for $E_\theta[t\,|\,x]$.)

Problem 3. *Verifying the Theory*

Returning to the setup of Problem 1, let $Y_1 \sim \exp(\theta_1)$ and $Y_2 \sim \exp(\theta_2)$ be independent exponential random variables, but *now suppose that* θ_2 *is known*. Again let $X = Y_1 + Y_2$, and suppose that $x_1, x_2, \ldots, x_n$ are iid samples from $f_{\theta_1,\theta_2} = \exp(\theta_1) \star \exp(\theta_2)$. We are now interested in estimating the single unknown parameter θ_1.

(a) Explain how this problem matches the setup of Problem 2, by specifying the density $g(x \,|\, \theta_1)$, the complete data z, and the one-dimensional exponential family model $f(z \,|\, \theta_1)$.

(b) Derive expressions for the Fisher informations $I_z(\theta_1)$ and $I_{z\,|\,\{x_i\}}(\theta_1)$.

(c) Implement the EM algorithm derived in Problem 2 for this case, and run it on the data in `/afs/cs/academic/class/10702/data/assign2.exponential2.dat`, which is a sample of 500 points from $f_{\theta_1,0.05}$ (note that this is a different file from the one used in Problem 1). Report your estimate of the MLE $\widehat{\theta}_1$, and give a plot of $\frac{\theta_1^{(k)}-\widehat{\theta}_1}{\theta_1^{(k-1)}-\widehat{\theta}_1}$ versus k.

(d) Compute $I_z(\widehat{\theta}_1)$ and $I_{z\,|\,\{x_i\}}(\widehat{\theta}_1)$. Does the convergence match what is predicted in Problem 2?

(e) Now repeat steps (c) and (d) on the *first 100 points* in the same data set. How does the convergence rate compare to that obtained on all 500 data points?