Recitation: Some Notes for Assignment 2

October 12, 2004

Some Notes about EM

Note1: Page 4 of Paper *, equation (3), we could write in a clearer way:
(* You might need feel helpful for the paper : Jeff A. Bilmes, "A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models")

$Q(\theta, \theta^{old}) = E_{P(Y|X,\theta^{old})}[log(L(\theta|X,Y))] = E_{P(Y|X,\theta^{old})}[\sum_{i=1}^{N} log(\alpha_{y_i} p_{y_i}(x_i|\theta_{y_i}))]$

$\because$ *note: Y is a vector $(y_1, y_2, ..., y_N)$; And y_i only depends on x_i

$\therefore Q(\theta, \theta^{old}) = \sum_{i=1}^{N}\{E_{P(y_i|x_i,\theta^{old})}[log(\alpha_{y_i} p_{y_i}(x_i|\theta_{y_i}))]\}$
$= \sum_{i=1}^{N}\{\sum_{y_i=1}^{M}[log(\alpha_{y_i} p_{y_i}(x_i|\theta_{y_i})) * p(y_i|x_i, \theta^{old})]\}$

$\because$ Then we could use l to substitute y_i

$\therefore Q(\theta, \theta^{old}) = \sum_{i=1}^{N}\{\sum_{l=1}^{M}[log(\alpha_l p_l(x_i|\theta_l)) * p(l|x_i, \theta^{old})]\}$

Note2: Simple way to think EM:

a. Include " Hidden/Missing Variable " into your model

b. Derive the Expected Value of the " Hidden/Missing Variable " based on the inputs X and your old parameter θ^{old}

c. Derive new *theta* from results of "b"

Question2. EM

For the following questions, please give clear step by step derivation.

2.1 Suppose that the p.d.f. of a random variable X has a 2-component mixture form:

$$p_\alpha(x) = \alpha * p_1(x) + (1 - \alpha) * p_2(x) \tag{1}$$

One component is the density model $p_1(x)$ and the other component is the density model $p_2(x)$. We know both $p_1(x)$ and $p_2(x)$. We do not know α. Given that $\{ x_1, x_2, ..., x_n \}$ are iid samples from the distribution of X, please give an EM algorithm for estimating α. (Describe the E-step and M-step clearly in your answer).

Hint:

The same framework as the Reference paper and the first page of this note.

2.2 Suppose that $Y_1 \sim exp(1/\theta_1)$ and $Y_2 \sim exp(1/\theta_2)$, and $\theta_1 \neq \theta_2$. Y_1 and Y_2 are independent. Let $X = Y_1 + Y_2$ denote the sum of Y_1 and Y_2, Given that $\{\ x_1, x_2, ..., x_n\ \}$ are iid samples from the distribution of X.

- Derive an expression for the density of X in terms of θ_1 and θ_2
(Hint1: The density of Y_1 is $f_{\theta_1}(y) = \theta_1 e^{-\theta_1 y}$, similarly for Y_2)
(Hint2: You could first derive CDF of X, $F(x) = P(Y_1 + Y_2 < x) = \int_0^x \int_0^{x-y_1} f_{\theta_1}(y_1) f_{\theta_2}(y_2) dy_2 dy_1$)

Hint: This step's result may be useful for the next question.

- Derive the E-step and M-step, and give explicit expressions for the parameter updates in the EM process for computing the MLE of θ_1 and θ_2.

Hint:
Y_1 is a hidden variable.

Way1: The same framework as the Reference paper and the first page of this note.

Way2: Use the simple thinking style of EM as in the first page of this note.

$p(y_{i,1}|x_i, \theta^{old}) = \frac{p(x_i, y_{i,1})|\theta^{old}}{p(x_i|\theta^{old})}$

Question4. Regression

Linear regression models a real-valued output Y given an input vector X as

$$Y|X \sim Normal(\mu(X), \sigma^2)$$

where the mean is a linear function of the input: $\mu(X) = \beta^T X = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p$

Logistic regression models a binary output Y by

$$Y|X \sim Bernoulli(\theta(X))$$

where the Bernoulli parameter is related to $\beta^T X$ by the logit transformation

$$logit(\theta(X)) \equiv \log\left(\frac{\theta(X)}{1-\theta(x)}\right) = \beta^T X$$

Given data $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$, for each of the two regression models above , show that at the MLE $\hat{\beta}$

$$\sum_{i=1}^{n} x_i * y_i = \sum_{i=1}^{n} x_i * E[\text{ Y } |X = x_i, \beta = \hat{\beta}]$$

Hint: Maximum Likelihood Estimation

One Extra Note: Actually the above equation shows that we achieve a good regression estimation for the data.