Stat 461-561: Quiz 2

Friday 29th January 2008

• **Exercise 1**. Let $X_i \overset{\text{i.i.d.}}{\sim} g(x)$ and assume that we want to model these data using the parametrized family of probability density functions (pdf) $\{f(x|\theta); \theta \in \Theta\}$. Let θ_n be the Maximum Likelihood Estimate (MLE) for n observations; that is

$$\theta_n = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \log f(X_i|\theta)$$

Under 'suitable' regularity assumptions, we have

$$\sqrt{n}(\theta_n - \theta^*) \xrightarrow{\text{D}} \mathcal{N}(0, \sigma^2)$$

Question 1.1: [**1 point**] Give the expression of the Kullback-Leibler divergence minimized in $\theta = \theta^*$.

Question 1.2: [**1 point**] *Establish* the expression of σ^2.

See lecture notes.

• **Exercise 2**. Consider n (positive) observations $X_1, ..., X_n$ where $X_i \overset{\text{i.i.d.}}{\sim} g(x)$ with

$$g(x) = \begin{cases} \pi\lambda_1 \exp(-\lambda_1 x) + (1-\pi)\lambda_2 \exp(-\lambda_2 x) & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

with $0 < \pi < 1$, $\lambda_1 > 0$ and $\lambda_2 > 0$. That is we have modeled the pdf $g(x)$ of the data by a mixture of two exponential distributions.

Given $X_1, ..., X_n$, we are interested in estimating $\theta = (\pi, \lambda_1, \lambda_2)$ using the Expectation-Maximization (EM) algorithm.

Question 2.1: [**1 point**] Which set of latent variables can you introduce to implement the EM algorithm? Describe the associated statistical model.

We can associate to each observation a latent variable $Z_i \in \{1, 0\}$ such that

$$p(z_i = 1|\theta) = 1 - p(z_i = 0|\theta) = \pi$$

and

$$p(x_i, z_i|\theta) = \begin{cases} \pi\lambda_1 \exp(-\lambda_1 x) & \text{if } z_i = 1 \\ (1-\pi)\lambda_2 \exp(-\lambda_2 x) & \text{if } z_i = 2. \end{cases}.$$

Then the complete log-likelihood is given by

$$p(x_{1:n}, z_{1:n}|\theta) = \prod_{i=1}^{n} [\pi\lambda_1 \exp(-\lambda_1 x_i)]^{z_i} [(1-\pi)\lambda_2 \exp(-\lambda_2 x_i)]^{1-z_i}$$

Question 2.2: [**3 points**] Establish the EM recursion giving the expression of the parameter estimate $\theta^{(k)}$ at iteration k given $\theta^{(k-1)}$ at iteration $k-1$.

We have

$$
\begin{aligned}
Q\left(\theta, \theta^{(k)}\right) &= \sum \log p\left(x_{1:n}, z_{1:n} \middle| \theta\right) . p\left(z_{1:n} \middle| \theta^{(k)}, x_{1:n}\right) \\
&= \sum_{z_{1:n} \in \{0,1\}^n} \sum_{i=1}^{n} [z_i \left(\log \pi + \log \lambda_1 - \lambda_1 x_i\right) \\
&\quad + \left(1 - z_i\right)\left(\log \pi + \log \lambda_2 - \lambda_2 x_i\right)] p\left(x_{1:n} \middle| \theta^{(k)}, z_{1:n}\right) \\
&= \left(\log \pi + \log \lambda_1\right) \sum_{i=1}^{n} p\left(z_i = 1 \middle| \theta^{(k)}, x_i\right) - \lambda_1 \sum_{i=1}^{n} x_i p\left(z_i = 1 \middle| \theta^{(k)}, x_i\right) \\
&\quad + \left(\log\left(1 - \pi\right) + \log \lambda_2\right) \sum_{i=1}^{n} p\left(z_i = 0 \middle| \theta^{(k)}, x_i\right) - \lambda_2 \sum_{i=1}^{n} x_i p\left(z_i = 0 \middle| \theta^{(k)}, x_i\right).
\end{aligned}
$$

We obtain

$$
\pi = \frac{\sum_{i=1}^{n} p\left(z_i = 1 \middle| \theta^{(k)}, x_i\right)}{n}
$$

and

$$
\begin{aligned}
\lambda_1 &= \frac{\sum_{i=1}^{n} p\left(z_i = 1 \middle| \theta^{(k)}, x_i\right)}{\sum_{i=1}^{n} x_i p\left(z_i = 1 \middle| \theta^{(k)}, x_i\right)}, \\
\lambda_2 &= \frac{\sum_{i=1}^{n} p\left(z_i = 0 \middle| \theta^{(k)}, x_i\right)}{\sum_{i=1}^{n} x_i p\left(z_i = 0 \middle| \theta^{(k)}, x_i\right)}
\end{aligned}
$$

where

$$
\begin{aligned}
p\left(z_i = 1 \middle| \theta^{(k)}, x_i\right) &= 1 - p\left(z_i = 0 \middle| \theta^{(k)}, x_i\right) \\
&= \frac{\pi^{(k)} \lambda_1^{(k)} \exp\left(-\lambda_1^{(k)} x_i\right)}{\pi^{(k)} \lambda_1^{(k)} \exp\left(-\lambda_1^{(k)} x_i\right) + \left(1 - \pi^{(k)}\right) \lambda_2^{(k)} \exp\left(-\lambda_2^{(k)} x_i\right)}.
\end{aligned}
$$

- **Exercise 3**. Consider the following simple polynomial regression model

$$
y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_k x^k + +\varepsilon,\ \varepsilon \sim \mathcal{N}\left(0, \sigma^2\right).
$$

Question 3.1: **[2 points]** Assuming k is known, establish the Maximum likelihood estimate of $\left(\widehat{\beta}_0, \widehat{\beta}_1, ..., \widehat{\beta}_k, \widehat{\sigma}^2\right)$ given n observations $\{x_i, y_i\}_{i=1,\ldots,n}$.

We can rewrite the observations as

$$
\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\epsilon}
$$

where

$$
\begin{aligned}
\mathbf{y} &= \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x_1 & \cdots & x_1^k \\ \\ 1 & x_1 & \cdots & x_n^k \end{pmatrix}, \\
\beta &= (\beta_0, \beta_1, ..., \beta_k)^{\mathrm{T}}
\end{aligned}
$$

and $\boldsymbol{\epsilon} \sim \mathcal{N}\left(0, \sigma^2 I_n\right)$. So the log-likelihood is

$$
l\left(\beta, \sigma^2\right) = -\frac{n}{2}\log\left(2\pi\right) - \frac{n}{2}\log\left(\sigma^2\right) - \frac{\left(\mathbf{y} - \mathbf{X}\beta\right)^{\mathrm{T}}\left(\mathbf{y} - \mathbf{X}\beta\right)}{2\sigma^2}
$$

and

$$
\begin{aligned}
\widehat{\beta} &= \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}, \\
\widehat{\sigma}^2 &= \frac{1}{n}\left(\mathbf{y} - \mathbf{X}\widehat{\beta}\right)^{\mathrm{T}}\left(\mathbf{y} - \mathbf{X}\widehat{\beta}\right)
\end{aligned}
$$

Question 3.2: **[2 points]** We now want to determine k through the Akaike Information Criterion. Establish that we have

$$
AIC\left(k\right) = n\left(\log 2\pi + 1\right) + n\log\widehat{\sigma}^2 + 2\left(k+2\right)
$$

We have

$$
\begin{aligned}
AIC\left(k\right) &= -2\text{log likelihood at MLE } + \ 2 \ \text{(number of parameters)} \\
&= n\log\left(2\pi\right) + n\log\left(\widehat{\sigma}^2\right) + \frac{\left(\mathbf{y} - \mathbf{X}\widehat{\beta}\right)^{\mathrm{T}}\left(\mathbf{y} - \mathbf{X}\widehat{\beta}\right)}{\widehat{\sigma}^2} \\
&\quad +2 \ \left(k+1\right) \\
&= n\left(\log\left(2\pi\right) + 1\right) + n\log\left(\widehat{\sigma}^2\right) + 2\left(k+2\right).
\end{aligned}
$$