COMS 4771 Machine Learning (Spring 2015) Problem Set #5

Jingwei Yang - jy2653@columbia.edu
Discussants: pp2526,yd2300

April 27, 2015

Problem 1

Suppose we have one item i; $\boldsymbol{X}_i = (X_{i,1}, X_{i,2}, ..., X_{i,n})$ and Y_i. Let $\boldsymbol{x}_i = (x_{i,1}, x_{i,2}, ..., x_{i,n}) \in \{0,1\}^n$ be the observed responses for item i. Using Jensen's inequality, we could have

$$\begin{aligned}\ln Pr_\theta(\boldsymbol{X}_i = x_i) &= \ln \sum_{y \in \{0,1\}} q(y) \frac{Pr_\theta(\boldsymbol{X}_i = \boldsymbol{x}_i \wedge Y_i = y)}{q(y)} \\ &\geq \ln \sum_{y \in \{0,1\}} q(y) \ln Pr_\theta(\boldsymbol{X} = \boldsymbol{x}_i \wedge Y_i = y) - \sum_{y \in \{0,1\}} q(y) \ln q(y)\end{aligned}$$

And for each $y \in \{0,1\}$, the "complete log-likelihood" is

$$\begin{aligned}\ln Pr_\theta(\boldsymbol{X}_i = \boldsymbol{x}_i \wedge Y_i = y) = (1-y)[\ln(1-\pi_i) + \sum_{j=1}^{n}(1-x_{i,j}) \ln p_j + x_{i,j} \ln(1-p_j)] \\ +y[\ln \pi_i + \sum_{j=1}^{n} x_{i,j} \ln r_j + (1-x_{i,j}) \ln(1-r_j)]\end{aligned}$$

Suppose $q_i = Pr_\theta(Y_i = 1 | \boldsymbol{X}_i = \boldsymbol{x}_i)$, by independence and Bayes' rule, we have

$$\begin{aligned}q_i &= Pr_\theta(Y_i = 1 | \boldsymbol{X}_i = \boldsymbol{x}_i) \\ &= \frac{\pi_i \prod_{j=1}^n r_j^{x_{i,j}} (1-r_j)^{1-x_{i,j}}}{\pi_i \prod_{j=1}^n r_j^{x_{i,j}} (1-r_j)^{1-x_{i,j}} + (1-\pi_i) \prod_{j=1}^n p_j^{1-x_{i,j}} (1-p_j)^{x_{i,j}}}\end{aligned}$$

When $q(y) = q_i^y (1-q_i)^{1-y}$, we have

$$\sum_{y\in 0,1} q(y)\ln Pr_\theta(\boldsymbol{X}=\boldsymbol{x}\wedge Y=y) = (1-q_i)[\ln(1-\pi_i)+\sum_{j=1}^{n}(1-x_{i,j})\ln p_j + x_{i,j}\ln(1-p_j)]$$
$$+q_i[\ln\pi_i + \sum_{j=1}^{n} x_{i,j}\ln r_j + (1-x_{i,j})\ln(1-r_j)]$$

Therefore, the E step is:

$$q_i = \frac{\pi_i \prod_{j=1}^{n} r_j^{x_{i,j}}(1-r_j)^{1-x_{i,j}}}{\pi_i \prod_{j=1}^{n} r_j^{x_{i,j}}(1-r_j)^{1-x_{i,j}} + (1-\pi_i)\prod_{j=1}^{n} p_j^{1-x_{i,j}}(1-p_j)^{x_{i,j}}}$$

After computing the value of q_i, consider all m, we could get $L_{LB}(\boldsymbol{\theta})$ in following form.

$$L_{LB}(\boldsymbol{\theta}) = \sum_{i=1}^{m}(1-q_i)[\ln(1-\pi_i)+\sum_{j=1}^{n}(1-x_{i,j})\ln p_j + x_{i,j}\ln(1-p_j)]$$
$$+q_i[\ln\pi_i + \sum_{j=1}^{n} x_{i,j}\ln r_j + (1-x_{i,j})\ln(1-r_j)]$$

Take gradient of p_j, we have

$$\nabla_{p_j} L_{LB}(\boldsymbol{\theta}) = \sum_{i=1}^{m}(1-q_i)(\frac{1-x_{i,j}}{p_j} - \frac{x_{i,j}}{1-p_j})$$

To maximize $L_{LB}(\boldsymbol{\theta})$, we assign $\nabla_{p_j} L_{LB}(\boldsymbol{\theta})$ to be 0, thus we have

$$\nabla_{p_j} L_{LB}(\boldsymbol{\theta}) = \sum_{i=1}^{m}(1-q_i)(\frac{1-x_{i,j}}{p_j} - \frac{x_{i,j}}{1-p_j}) = 0$$

$\Longrightarrow$

$$\sum_{i=1}^{m}(1-q_i)((1-x_{i,j})(1-p_j) - x_{i,j}p_j) = 0$$

$\Longrightarrow$

$$\sum_{i=1}^{m}(1-q_i)(1-p_j-x_{i,j}) = 0$$

$\Longrightarrow$

$$\sum_{i=1}^{m}(1-q_i)(1-x_{i,j}) - p_j(1-q_i) = 0$$

$\Longrightarrow$

$$p_j \sum_{i=1}^{m} (1-q_i) = \sum_{i=1}^{m} (1-q_i)(1-x_{i,j})$$

$\Longrightarrow$

$$p_j = \frac{\sum_{i=1}^{m} (1-q_i)(1-x_{i,j})}{\sum_{i=1}^{m} (1-q_i)}$$

Take gradient of r_j, we have

$$\nabla_{r_j} L_{LB}(\boldsymbol{\theta}) = \sum_{i=1}^{m} q_i \left(\frac{x_{i,j}}{r_j} + \frac{x_{i,j}-1}{1-r_j}\right)$$

To maximize $L_{LB}(\boldsymbol{\theta})$, we assign $\nabla_{r_j} L_{LB}(\boldsymbol{\theta})$ to be 0, thus we have

$$\nabla_{p_j} L_{LB}(\boldsymbol{\theta}) = \sum_{i=1}^{m} q_i \left(\frac{x_{i,j}}{r_j} + \frac{x_{i,j}-1}{1-r_j}\right) = 0$$

$\Longrightarrow$

$$\sum_{i=1}^{m} q_i (x_{i,j}(1-r_j) + (x_{i,j}-1) r_j) = 0$$

$\Longrightarrow$

$$\sum_{i=1}^{m} q_i (x_{i,j} - r_j x_{i,j} + r_j x_{i,j} - r_j) = 0$$

$\Longrightarrow$

$$\sum_{i=1}^{m} q_i (x_{i,j} - r_j) = 0$$

$\Longrightarrow$

$$r_j \sum_{i=1}^{m} q_i = \sum_{i=1}^{m} q_i x_{i,j}$$

$\Longrightarrow$

$$r_j = \frac{\sum_{i=1}^{m} q_i x_{i,j}}{\sum_{i=1}^{m} q_i}$$

Therefore, we have following EM algorithm for this problem.
E step :

$$q_i = \frac{\pi_i \prod_{j=1}^{n} r_j^{x_{i,j}} (1-r_j)^{1-x_{i,j}}}{\pi_i \prod_{j=1}^{n} r_j^{x_{i,j}} (1-r_j)^{1-x_{i,j}} + (1-\pi_i) \prod_{j=1}^{n} p_j^{1-x_{i,j}} (1-p_j)^{x_{i,j}}}$$

for all $i \in [m]$

M step :

$$p_j = \frac{\sum_{i=1}^{m}(1-q_i)(1-x_{i,j})}{\sum_{i=1}^{m}(1-q_i)}$$

$$r_j = \frac{\sum_{i=1}^{m} q_i x_{i,j}}{\sum_{i=1}^{m} q_i}$$

Problem 2

(a)
Solution : $p_1 = \frac{1}{15}, p_2 = \frac{1}{15}, p_3 = \frac{1}{15}, p_4 = \frac{1}{5}, p_5 = \frac{3}{10}, p_6 = \frac{3}{10}$

According to the description of problem, we could have following equations.

$$\begin{aligned} p_4 &= 0.2 \\ p_1 + p_2 + p_3 &= 0.2 \\ p_5 + p_6 &= 0.6 \end{aligned}$$

$\Longrightarrow$

$$\begin{aligned} p_5 &= 0.6 - p_6 \\ p_1 &= 0.2 - p_2 - p_3 \end{aligned}$$

Thus, we could replace p_5 and $p_1$in our entropy calculation.

$$\begin{aligned} H(P) = &\sum_{x\in\{1,\ldots,6\}} -P(x)\ln P(x) \\ = &-(0.2 - p_2 - p_3)\ln(0.2 - p_2 - p_3) - p2\ln p2 - p3\ln p3 \\ &- 0.2\ln 0.2 - (0.6 - p_6)\ln(0.6 - p_6) - p_6\ln p_6 \end{aligned}$$

Take gradient of p_6, we have

$$\nabla_{p_6} H(P) = \ln(0.6 - p_6) - \ln p_6$$

To maximize $H(P)$, we assign $\nabla_{p_6} H(P)$ to be 0, thus we have

$$\begin{aligned} p_6 &= 0.3 \\ p_5 &= 0.3 \end{aligned}$$

Also take gradient of p_2, we have

$$\nabla_{p_2} H(P) = \ln(0.2 - p_2 - p_3) - \ln p_2$$

To maximize $H(P)$, we assign $\nabla_{p_2} H(P)$ to be 0, thus we have

$$p_2 = \frac{0.2 - p_3}{2}$$

Replace p_2 in the $H(P)$, we could have

$$\begin{aligned} H(P) = &\sum_{x\in\{1,\ldots,6\}} -P(x)\ln P(x) \\ = &-(\frac{0.2 - p_3}{2})\ln(\frac{0.2 - p_3}{2}) - (\frac{0.2 - p_3}{2})\ln(\frac{0.2 - p_3}{2}) - p_3\ln p_3 \\ &- 0.2\ln 0.2 - (0.6 - p_6)\ln(0.6 - p_6) - p_6\ln p_6 \end{aligned}$$

Take gradient of p_3, we have

$$\nabla_{p_3} H(P) = \ln \frac{0.2 - p_3}{2} - \ln p_3$$

To maximize $H(P)$, we assign $\nabla_{p_3} H(P)$ to be 0, thus we have

$$\begin{aligned} p3 &= \frac{1}{15} \\ p2 &= \frac{1}{15} \\ p1 &= \frac{1}{15} \end{aligned}$$

(b)

Solution : $p_1 = 0.25, p_2 = 0.25, p_3 = 0.125, p_4 = 0.125, p_5 = 0.125, p_6 = 0.125$
According to the problem, we have

$$\begin{aligned} p_1 + p_2 &= 0.5 \\ p_2 + p_4 + p_6 &= 0.5 \\ p_1 + p_3 + p_5 &= 0.5 \end{aligned}$$

$\Longrightarrow$

$$\begin{aligned} p_2 &= 0.5 - p_1 \\ p_3 &= 0.5 - p_5 - p_1 \\ p_4 &= p_1 - p_6 \end{aligned}$$

Thus, we could write the overall entropy $H(P)$ as following form

$$\begin{aligned} H(P) &= \sum_{x \in \{1,\ldots,6\}} -P(x) \ln P(x) \\ &= -p_1 \ln p_1 - (0.5 - p_1) \ln(0.5 - p_1) - (0.5 - p_5 - p_1) \ln(0.5 - p_5 - p_1) \\ &\quad - (p_1 - p_6) \ln(p_1 - p_6) - p_5 \ln p_5 - p_6 \ln p_6 \end{aligned}$$

Take gradient of p_5, we have

$$\nabla_{p_5} H(P) = \ln(0.5 - p_5 - p_1) - \ln p_5$$

To maximize $H(P)$, we assign $\nabla_{p_5} H(P)$ to be 0, thus we have

$$p_5 = \frac{0.5 - p_1}{2}$$

Since we have $p_3 = 0.5 - p_5 - p_1$, we could also have

$$p_3 = \frac{0.5 - p_1}{2}$$

Take gradient of p_6, we have

$$\nabla_{p_6} H(P) = \ln(p_1 - p_6) - \ln p_6$$

To maximize $H(P)$, we assign $\nabla_{p_6} H(P)$ to be 0, thus we have

$$p_6 = \frac{p_1}{2}$$

Since we have $p_2 = 0.5 - p_1$ and $p_2 + p_4 + p_6 = 0.5$, we could also have

$$\begin{aligned} p_2 &= 0.5 - p_1 \\ p_4 &= \frac{p_1}{2} \end{aligned}$$

So far, we could replace p_2, p_3, p_4, p_5, p_6 in the form of p_1, thus we have

$$\begin{aligned} H(P) &= \sum_{x \in \{1,\ldots,6\}} -P(x) \ln P(x) \\ &= -p_1 \ln p_1 - (0.5 - p_1)\ln(0.5 - p_1) - 2 * \frac{0.5 - p_1}{2} \ln(\frac{0.5 - p_1}{2}) - 2 * \frac{p_1}{2} \ln \frac{p_1}{2} \end{aligned}$$

Take gradient of p_1, we have

$$\nabla_{p_6} H(P) = -\ln p_1 + \ln(0.5 - p_1) + \ln(\frac{0.5 - p_1}{2}) - \ln \frac{p_1}{2}$$

To maximize $H(P)$, we assign $\nabla_{p_1} H(P)$ to be 0, thus we have

$$-\ln p_1 + \ln(0.5 - p_1) + \ln(\frac{0.5 - p_1}{2}) - \ln \frac{p_1}{2} = 0$$

$\Longrightarrow$

$$\ln \frac{(0.5 - p_1)^2}{2} = \ln \frac{p_1^2}{2}$$

$\Longrightarrow$

$$p_1 = 0.25$$

Plug the $p_1 = 0.25$ into above equations, we could have

$$\begin{aligned} p_1 &= 0.25 \\ p_2 &= 0.25 \\ p_3 &= 0.125 \\ p_4 &= 0.125 \\ p_5 &= 0.125 \\ p_6 &= 0.125 \end{aligned}$$