CSCI-567: Assignment #5

Due on Monday, November 16, 2015

Saket Choudhary
2170058637

Contents

Problem 1 **3**
Problem 1: (a) 3
Problem 1: (b) 4

Problem 3 **6**

Problem 1

Problem 1: (a)

To find $\nabla_{y_t} L$:

$$\begin{aligned}
\nabla_{y_t} L &= \frac{\partial}{\partial y_t}\frac{1}{2}\sum_{i=1}^{N}(y_i - \hat{y}_i)^T(y_i - \hat{y}_i) \\
&= \frac{\partial}{\partial y_t}\frac{1}{2}\sum_{i=1}^{N}(y_i^T y_i - 2y_i^T\hat{y}_i + \hat{y_t}^2) \\
&= \frac{1}{2}(2y_t - 2\hat{y}_t) \\
&= y_t - \hat{y}_t
\end{aligned}$$

$$\boxed{\nabla_{y_t} L = y_t - \hat{y}_t}$$

Problem 1 continued on next page. . .

Problem 1: (b)

To find $\nabla_{y_t} L$:

$$\nabla_{s_t} L = \sum_{k=1}^{T} \frac{\partial L}{\partial y_k} \times \frac{\partial y_k}{\partial s_k} \times \frac{\partial s_k}{\partial s_t}$$

Let's define $z_k = W_{IH}x_t + W_{HH}s_{t-1}$
Thus,

$$\begin{aligned} z_k &= W_{IH}x_k + W_{HH}s_{k-1} \\ s_k &= \sigma(z_k) \\ y_k &= W_{HO}s_k \end{aligned}$$

Thus,

$$\frac{\partial y_k}{\partial s_k} = W_{HO} \tag{1}$$

$$\frac{\partial s_k}{\partial z_k} = \sigma(z_k)(1 - \sigma(z_k)) \tag{2}$$

$$\frac{\partial z_k}{\partial W_{IH}} = x_k \tag{3}$$

$$\frac{\partial y_k}{\partial W_{HH}} = y_{k-1} \tag{4}$$

$$\frac{\partial z_k}{\partial s_{k-1}} = W_{HH} \tag{5}$$

$$\frac{\partial s_k}{\partial s_{k-1}} = \frac{\partial s_k}{\partial z_k}\frac{\partial z_k}{\partial s_{k-1}} = \sigma(z_k)(1 - \sigma(z_k))W_{HH} \tag{6}$$

Let's now consider $\frac{\partial s_k}{\partial s_t}$:
s_k depends on $s_{k-1}, s_{k-2}, \dots s_1$. And hence:

$$\frac{\partial s_k}{\partial s_t} = 0 \; \forall \; k < t$$

For $k \geq t$:

$$\frac{\partial s_k}{\partial s_t} = \frac{\partial s_k}{\partial s_{k-1}} \times \frac{\partial s_{k-1}}{\partial s_{k-2}} \times \frac{\partial s_{k-2}}{\partial s_{k-3}} \times \dots \frac{\partial s_{k-(k-t)+1}}{\partial s_{k-(k-t)}}$$

Thus, consider a special case of $t = T$:

$$\begin{aligned} \nabla_{s_T} L &= \sum_{k=T}^{T} \frac{\partial L}{\partial y_k} \times \frac{\partial y_k}{\partial s_k} \times \frac{\partial s_k}{\partial s_t} \\ &= \frac{\partial L}{\partial y_T} \times \frac{\partial y_T}{\partial s_T} \\ &= (y_T - \hat{y_T})W_{HO} \end{aligned}$$

Thus,

$$\boxed{\nabla_{S_T} L = (y_T - \hat{y_T})W_{HO}}$$

Problem 1 [Problem 1: (b)] continued on next page...

Let's consider $\nabla_{s_t} L$ and $\nabla_{s_{t+1}} L$:

$$\nabla_{s_{t+1}} L = \sum_{k=t+1}^{T} \frac{\partial L}{\partial y_k} \times \frac{\partial y_k}{\partial s_k} \times \frac{\partial s_k}{\partial s_t}$$
$$\nabla_{s_t} L = \sum_{k=t}^{T} \frac{\partial L}{\partial y_k} \times \frac{\partial y_k}{\partial s_k} \times \frac{\partial s_k}{\partial s_t}$$
$$\implies \nabla_{s_t} L = \nabla_{s_{t+1}} L + \frac{\partial L}{\partial y_t} \times \frac{\partial y_t}{\partial s_t} \times \frac{\partial s_t}{\partial s_t}$$
$$\implies \nabla_{s_t} L = \nabla_{s_{t+1}} L + (y_y - \hat{y}_t) W_{HO}$$

Thus,

$$\boxed{\nabla_{s_t} L = \nabla_{s_{t+1}} L + (y_y - \hat{y}_t) W_{HO}}$$

Problem 3

Given:

$$p(x_i) = \begin{cases} \pi + (1-\pi)e^{-\lambda} & x_i = 0 \\ (1-\pi)\frac{\lambda^{x_i}e^{-\lambda}}{x_i!} & x_i > 0 \end{cases}$$

Alternatively:

$$X_i = \begin{cases} 0 & \text{probability} = \pi + (1-\pi)e^{-\lambda} \\ x_i & \text{probability} = (1-\pi)\frac{\lambda^{x_i}e^{-\lambda}}{x_i!} \end{cases}$$

We define a *latent* variable Z_i for all cases where $X_i = 0$. It is latent because when we observed $X_i = 0$ we do not know if it came out of the 'Poisson' distribution or it came out the 'degenerate' distribution(which has a probability of 1 at point 0.). we cannot observe the following. So X_i comes out of a mixture of a degenerate distribution as follows:

$$Z_i = \begin{cases} 1 & X_i \text{ is from the degenerate distrbution} \\ 0 & \text{otherwise} \end{cases}$$

$$p(X_i = 0, Z_i = 1) = p(Z_i = 1) \times p(X_i = 0|Z_i = 1) = \pi \times 1$$
$$P(X_i = 0, Z_i = 0) = p(Z_i = 0) \times p(X_i = 0|Z_i = 0) = (1-\pi)e^{-\lambda} \times 1$$

$$L(Complete) = \prod_{x_i=0} \pi^{Z_i} \times ((1-\pi)e^{-\lambda})^{1-Z_i} \times \prod_{x_i>0} (1-\pi)e^{\frac{\lambda_i^x e^{-\lambda}}{x_i!}} \tag{7}$$

$$\log L = \sum_{I(x_i=0)} z_i \log(\pi) + (1-z_i)\big(\log(1-\pi) - \lambda\big) + \sum_{I(x_i>0)} \big(\log(1-\pi) + x_i \log(\lambda_i) - \lambda - \log(x_i!)\big) \tag{8}$$

E step:

$$Q(\theta, \theta_0) = \sum_{I(x_i=0)} E_{P(Z|X)}[z_i] \log(\pi) + (1 - E_{P(Z|X)}[z_i])\big(\log(1-\pi) - \lambda\big)$$
$$+ \sum_{I(x_i>0)} \big(\log(1-\pi) + x_i \log(\lambda_i) - \lambda - \log(x_i!)\big)$$

$$\begin{aligned} E_{P(Z|X_i)}[z_i] &= 0 \times p(Z_i = 0|X) + 1 \times p(Z_i = 1|X_i = 0) \\ &= \frac{p(X_i = 0|Z_i = 1)p(Z_i = 1)}{p(X_i = 0|Z_i = 0)p(Z_i = 0) + p(X_i = 0|Z_i = 1)p(Z_i = 1)} \\ &= \frac{\pi_0}{\pi_0 + (1-\pi_0)e^{-\lambda_0}} \end{aligned}$$

Hence,

$$Q(\theta, \theta_0) = \sum_{I(x_i=0)} \frac{\pi_0}{\pi_0 + (1-\pi_0)e^{-\lambda_0}} \log(\pi) + (\frac{(1-\pi_0)e^{-\lambda_0}}{\pi_0 + (1-\pi_0)e^{-\lambda_0}})\big(\log(1-\pi) - \lambda\big)$$
$$+ \sum_{I(x_i>0)} \big(\log(1-\pi) + x_i \log(\lambda) - \lambda - \log(x_i!)\big)$$

M step:

$$
\begin{aligned}
\frac{\partial Q}{\partial \lambda} &= 0 \\
&= \sum_{I(x_i=0)} (1-E[z_i])(-1) + \sum_{I(x_i>0)} (\frac{x_i}{\lambda} - 1) = 0 \\
\implies \hat{\lambda} &= \frac{\sum_{I(x_i>0)} x_i}{n - \sum_{I(x_i=0)} E[z_i]} \\
\hat{\lambda} &= \frac{\sum_{I(x_i>0)} x_i}{n - \sum_{I(x_i=0)} \hat{z_i}} \\
\text{where } \hat{z} &= \frac{\pi_0}{\pi_0 + (1-\pi_0)e^{-\lambda_0}}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial Q}{\partial \pi} &= 0 \\
&= \sum_{I(x_i=0)} (\frac{E[z_i]}{\pi} - \frac{1-E[z_i]}{1-\pi}) - \sum_{I(x_i>0)} \frac{1}{1-\pi} = 0 \\
&= \sum_{I(x_i=0)} (\frac{E[z_i]}{\pi} + \frac{E[z_i]}{1-\pi}) - \frac{n}{1-\pi} = 0 \\
\implies \hat{\pi} &= \sum_{I(x_i=0)} \frac{\hat{z_i}}{n}
\end{aligned}
$$