Machine Learning 1 - Homework Week 6

Minh Ngo 10897402[1]

October 16, 2015

[1] University of Amsterdam
minh.ngole@student.uva.nl

Task 3

$$p(x|\lambda) = \frac{1}{x!}\lambda^x \exp(-\lambda),\ x \in \mathcal{N},\ \lambda > 0 \tag{1}$$

$$p(x_n) = \sum_{k=1}^{K} \pi_k p(x_n|\lambda_k) \tag{2}$$

$$\sum_k \pi_k = 1 \tag{3}$$

(a) Likelihood:

$$\begin{aligned} p(\mathbf{x}) &= \prod_{n=1}^{N} p(x_n) = \prod_{n=1}^{N} \sum_{k=1}^{K} \pi_k p(x_n|\lambda_k) \\ &= \prod_{n=1}^{N} \sum_{k=1}^{K} \pi_k \frac{1}{x_n!} \lambda_k^{x_n} \exp(-\lambda_k) \end{aligned} \tag{4}$$

(b) Log-likelihood:

$$\log p(\mathbf{x}) = \sum_{n=1}^{N} \log \sum_{k=1}^{K} \pi_k \frac{1}{x_n!} \lambda_k^{x_n} \exp(-\lambda_k) \tag{5}$$

(c)

$$r_{nk} = \frac{\pi_k p(x_n|\lambda_k)}{\sum_{j=1}^{K} \pi_j p(x_n|\lambda_j)} = \frac{\pi_k \frac{1}{x_n!} \lambda_k^{x_n} \exp(-\lambda_k)}{\sum_{j=1}^{K} \pi_j \frac{1}{x_n!} \lambda_j^{x_n} \exp(-\lambda_j)} = \frac{\pi_k \lambda_k^{x_n} \exp(-\lambda_k)}{\sum_{j=1}^{K} \pi_j \lambda_j^{x_n} \exp(-\lambda_j)} \tag{6}$$

(d)

$$\begin{aligned} \frac{\partial p(x_n|\lambda_k)}{\partial \lambda_k} &= \frac{1}{x_n!}\Big(x_n \lambda_k^{x_n - 1} \exp(-\lambda_k) - \lambda_k^{x_n} \exp(-\lambda_k) \Big) \\ &= p(x_n|\lambda_k)(\frac{x_n}{\lambda_k} - 1) \end{aligned} \tag{7}$$

$$\begin{aligned} \frac{\partial \log p(\mathbf{x})}{\partial \lambda_k} &= \sum_{n=1}^{N} \frac{\pi_k}{\sum_{j=1}^{K} \pi_j p(x_n|\lambda_j)} \frac{1}{x_n!}\Big(x_n \lambda_k^{x_n - 1} \exp(-\lambda_k) - \lambda_k^{x_n} \exp(-\lambda_k) \Big) \\ &= \sum_{n=1}^{N} (\frac{x_n}{\lambda_k} r_{nk} - r_{nk}) = 0 \end{aligned} \tag{8}$$

$$\frac{1}{\lambda_k} \sum_{n=1}^{N} x_n r_{nk} = N_k \Leftrightarrow \lambda_k = \frac{1}{N_k} \sum_{n=1}^{N} r_{nk} x_n \tag{9}$$

(e)

$$\mathcal{L}(\mathbf{x}, \lambda) = \log p(\mathbf{x}) + \lambda(\sum_{j=1}^{K} \pi_j - 1) = 0 \tag{10}$$

$$\frac{\partial \mathcal{L}(\mathbf{x}, \lambda)}{\partial \pi_k} = \sum_{n=1}^{N} \frac{p(x_n|\lambda_k)}{\sum_{j=1}^{K} \pi_j p(x_n|\lambda_j)} + \lambda = 0 \tag{11}$$

$$\sum_{n=1}^{N} r_{nk} + \lambda\pi_k = 0$$
$$N_k + \lambda\pi_k = 0 \Leftrightarrow \pi_k\lambda = -N_k \tag{12}$$

$$\sum_{j=1}^{K} \pi_k\lambda = \sum_{j=1}^{K}(-N_k) \Leftrightarrow \lambda = -N,\ \pi_k = \frac{N_k}{N} \tag{13}$$

(f) Let $p(\mathbf{x}, \boldsymbol{\pi}, \boldsymbol{\lambda}|a, b, \alpha, K)$ be $p(\mathbf{Z}|\boldsymbol{\theta})$.

$$p(\mathbf{Z}|\boldsymbol{\theta}) = \left(Dir(\boldsymbol{\pi}, \frac{\boldsymbol{\alpha}}{K})\prod_{j=1}^{K}\mathcal{G}(\lambda_j|a,b)\right)\left(\prod_{n=1}^{N}\sum_{j=1}^{K}\pi_j\frac{1}{x_n!}\lambda_j^{x_n}\exp(-\lambda_j)\right) \tag{14}$$

$$\begin{aligned}\log p(\mathbf{Z}|\boldsymbol{\theta}) = \log Dir(\boldsymbol{\pi}, \frac{\boldsymbol{\alpha}}{K}) + \sum_{j=1}^{K}\log\mathcal{G}(\lambda_j|a,b) + \\ + \sum_{n=1}^{N}\log\sum_{j=1}^{K}\pi_j\frac{1}{x_n!}\lambda_j^{x_n}\exp(-\lambda_j)\end{aligned} \tag{15}$$

(g)

$$\begin{aligned}\frac{\partial\mathcal{G}(\lambda_k|a,b)}{\partial\lambda_k} &= \frac{\partial\frac{1}{\Gamma(a)}b^a\lambda_k^{a-1}\exp(-b\lambda_k)}{\partial\lambda_k} \\ &= \frac{1}{\Gamma(a)}b^a((a-1)\lambda_k^{a-2}\exp(-b\lambda_k) - b\lambda_k^{a-1}\exp(-b\lambda_k)) \\ &= \frac{1}{\Gamma(a)}b^a\lambda_k^{a-2}\exp(-b\lambda_k)(a-1-b\lambda_k) \\ &= \mathcal{G}(\lambda_k|a,b)(\frac{a-1}{\lambda_k} - b)\end{aligned} \tag{16}$$

$$
\begin{aligned}
\frac{\partial \log p(\mathbf{Z}|\boldsymbol{\theta})}{\partial \lambda_k} &= \frac{1}{\mathcal{G}(\lambda_k|a,b)}\mathcal{G}(\lambda_k|a,b)(\frac{a-1}{\lambda_k}-b)+\sum_{n=1}^{N}(\frac{x_n}{\lambda_k}r_{nk}-r_{nk}) \\
&= (\frac{a-1}{\lambda_k}-b)+\sum_{n=1}^{N}(\frac{x_n}{\lambda_k}r_{nk}-r_{nk}) \\
&= \frac{a-1}{\lambda_k}-b+\frac{1}{\lambda_k}\sum_{n=1}^{N}x_n r_{nk}-N_k=0
\end{aligned} \tag{17}
$$

$$
\lambda_k(b+N_k)=a-1+\sum_{n=1}^{N}x_n r_{nk} \tag{18}
$$

$$
\lambda_k=\frac{a-1+\sum_{n=1}^{N}x_n r_{nk}}{b+N_k} \tag{19}
$$

(h)

$$
\mathcal{L}(x,\lambda)=\log p(x)+\lambda(\sum_{j=1}^{K}\pi_j-1) \tag{20}
$$

$$
\begin{aligned}
\frac{\partial \log Dir(\boldsymbol{\pi},\frac{\boldsymbol{\alpha}}{K})}{\partial \pi_k} &= \frac{\partial \log\left(C(\frac{\boldsymbol{\alpha}}{K})\prod_{j=1}^{K}\pi_j^{\frac{\alpha}{K}-1}\right)}{\partial \pi_k} \\
&= \frac{\alpha-K}{K\pi_k}
\end{aligned} \tag{21}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}(x,\lambda)}{\partial \pi_k} &= \frac{\alpha-K}{K\pi_k}+\sum_{n=1}^{N}\frac{\frac{1}{x_n!}\lambda_k^{x_n}\exp(-\lambda_k)}{\sum_{j=1}^{K}\pi_j\frac{1}{x_n!}\lambda_j^{x_n}\exp(-\lambda_j)}+\lambda \\
&= \frac{\alpha-K}{K\pi_k}+\sum_{n=1}^{N}\frac{r_{nk}}{\pi_k}+\lambda=0
\end{aligned} \tag{22}
$$

$$\begin{aligned}
\lambda \pi_k &= \frac{K-\alpha}{K} - N_k \\
\lambda &= \sum_{j=1}^{K} (\frac{K-\alpha}{K} - N_j) \\
\lambda &= K - \alpha - N
\end{aligned} \tag{23}$$

$$\pi_k = \frac{K - \alpha - KN_k}{K(K - \alpha - N)} \tag{24}$$

(i) EM algorithm:

1 Initialize $\boldsymbol{\pi}$, $\boldsymbol{\lambda}$

2 Repeat until convergence (check change in log-joint)

i. E-step (for all k, n):

$$r_{nk} = \frac{\pi_k \lambda_k^{x_n} \exp(-\lambda_k)}{\displaystyle\sum_{j=1}^{K} \pi_j \lambda_j^{x_n} \exp(-\lambda_j)} \tag{25}$$

ii. M-step (for all k):

$$\lambda_k = \frac{a - 1 + \displaystyle\sum_{n=1}^{N} x_n r_{nk}}{b + N_k} \tag{26}$$

$$\pi_k = \frac{K - \alpha - KN_k}{K(K - \alpha - N)} \tag{27}$$