Tarun Krishna
11593040
krishnatarun7@gmail.com

Homework Assignment 5
Machine Learning 1, 17/18

2017-12-30

1 First Exercise

Given:

$$p(x|\lambda) = \frac{1}{x!}\lambda_x \exp(-\lambda)$$
$$p(x_n) = \prod_{k=1}^{K} \pi_k p(x_n|\lambda_k)$$

(a)

$$L = \prod_{n=1}^{N} \left(\sum_{k=1}^{K} \pi_k p(x_n|\lambda_k) \right)$$

(b)

$$\mathcal{L}(x;\lambda) = \sum_{n=1}^{N} \log \left(\sum_{k=1}^{K} \pi_k p(x_n|\lambda_k) \right)$$

(c)

$$r_k(x_n) = \frac{p(r_k)p(x_n|r_k)}{p(x_n)}$$
$$r_{nk} = \frac{\pi_k p(x_n|\lambda_k)}{\sum_{k=1}^{K} \pi_k p(x_n|\lambda_k)}$$

(d)

$$\begin{aligned}
\frac{\partial}{\partial \lambda_k}\mathcal{L}(x;\lambda) &= \frac{\partial}{\partial \lambda_k} \big(\sum_{n=1}^{N} \log \left(\sum_{k=1}^{K} \pi_k \frac{1}{x_n!} \lambda_k^{x_n} \exp(-\lambda_k) \right) \big) \\
&= \sum_{n=1}^{N} \frac{\frac{\partial}{\partial \lambda_k} \left(\sum_{k=1}^{K} \pi_k \frac{1}{x_n!} \lambda_k^{x_n} \exp(-\lambda_k) \right)}{\sum_{k=1}^{K} \pi_k \frac{1}{x_n!} \lambda_k^{x_n} \exp(-\lambda_k)} \\
&= \sum_{n=1}^{N} \frac{\pi_k \frac{1}{x_n!} (\exp(-\lambda_k) x_n \lambda_k^{x_n - 1} - \lambda_k^{x_n} \exp(-\lambda_k))}{\sum_{k=1}^{K} \pi_k \frac{1}{x_n!} \lambda_k^{x_n} \exp(-\lambda_k)} \\
&= \sum_{n=1}^{N} \frac{\pi_k \frac{1}{x_n!} \lambda_k^{x_n} \exp(-\lambda_k)(x_n \lambda_k^{-1} - 1)}{\sum_{k=1}^{K} \pi_k \frac{1}{x_n!} \lambda_k^{x_n} \exp(-\lambda_k)} \\
&= \sum_{n=1}^{N} r_{nk} \big(\frac{x_n}{\lambda_k} - 1 \big)
\end{aligned}$$

Equating it zero

$$\begin{aligned}
0 &= \sum_{n=1}^{N} r_{nk} x_n - \sum_{n=1}^{N} \lambda_k r_{nk} x \\
\lambda_k \sum_{n=1}^{N} r_{nk} &= \sum_{n=1}^{N} r_{nk} x_n \\
\lambda_k &= \frac{\sum_{n=1}^{N} r_{nk} x_n}{\sum_{n=1}^{N} r_{nk}}
\end{aligned}$$

(e)

We need to use langranges's multiplier as we have constrained over of π_k so we use α as our langranges multiplier.So the equation can be written as:

$$\sum_{n=1}^{N} \log\Big(\sum_{k=1}^{K} \pi_k \frac{1}{x_n!} \lambda_k^{x_n} \exp(-\lambda_k)\Big) + \alpha\big(\sum_{n=1}^{N} \pi_k - 1\big) = 0$$
$$\frac{\partial}{\partial \pi_k}\Big(\sum_{n=1}^{N} \log\Big(\sum_{k=1}^{K} \pi_k \frac{1}{x_n!} \lambda_k^{x_n} \exp(-\lambda_k)\Big) + \alpha\big(\sum_{n=1}^{N} \pi_k - 1\big)\Big) = 0$$
$$\sum_{n=1}^{N} \frac{\frac{\lambda_k^{x_n}}{x_n!} \exp(-\lambda_k)}{\sum_{k=1}^{K} \pi_k \frac{1}{x_n!} \lambda_k^{x_n} \exp(-\lambda_k)} + \alpha = 0 \Rightarrow \qquad (1)$$

making use of constraint $\sum_{n=1}^{N} \pi_k = 1$ I can write

$$\frac{1}{\pi_k} \sum_{n=1}^{N} \frac{\pi_k \frac{1}{x_n!} \lambda_k^{x_n} \exp(-\lambda_k)}{\sum_{k=1}^{K} \pi_k \frac{1}{x_n!} \lambda_k^{x_n} \exp(-\lambda_k)} = -\alpha$$
$$\sum_{n=1}^{N} r_{nk} = -\pi_k \alpha$$
$$\pi_k = -\frac{\sum_{n=1}^{N} r_{nk}}{\alpha}$$

Substituting in (1). Multiplying and dividing my π_k as well

$$\sum_{n=1}^{N} \pi_k = 1$$
$$\alpha = -\sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk}$$
$$\alpha = -N$$
$$\pi_k = \frac{\sum_{n=1}^{N} r_{nk}}{N}$$

(f)

$$p(\pi_1 \ldots . \pi_k) = \mathcal{D}(\pi_1 \ldots . \pi k | \alpha/K \ldots . \alpha/K)$$
$$p(\lambda_k | a, b) = \mathcal{G}(\lambda_k | a, b)$$

Denoting

$$p(\pi_1 \ldots . \pi_k | \alpha/K \ldots . \alpha/K) = p(\pi | \frac{\alpha}{K})$$
$$p(\mathbf{X}, \pi, \lambda | a, b, \alpha, K) = p(\lambda | a, b) p(\pi | \alpha, K) p(\mathbf{X})$$
$$= \mathcal{D}(\pi | \frac{\alpha}{K}) \prod_{k=1}^{K} \mathcal{G}(\lambda_k | a, b) \prod_{n=1}^{N} p(x_n)$$
$$ln(p(\pi_1 \ldots . \pi k | \alpha/K \ldots . \alpha/K)) = ln\Big(\mathcal{D}(\pi | \frac{\alpha}{K}) \prod_{k=1}^{K} \mathcal{G}(\lambda_k | a, b) \prod_{n=1}^{N} p(x_n)\Big)$$
$$= ln(C(\frac{\alpha}{K}) \prod_{k=1}^{K} \pi_k^{\frac{\alpha}{K} - 1}) + \sum_{k=1}^{K} ln \frac{1}{\Gamma(a)} b^a \lambda_k^{a-1} e^{-b\lambda_k} + \sum_{n=1}^{N} ln p(\mathbf{x}_n)$$
$$= \sum_{k=1}^{K} (\frac{\alpha}{K} - 1) + \sum_{k=1}^{K} ((a-1) ln \lambda_k - b\lambda_k) + \underbrace{\mathcal{L}}_{\text{likelihood from (b)}} + C$$

(g)

Maximizing λ_k

$$
\begin{aligned}
\frac{\partial}{\partial \lambda_k} \ln(p(\pi_1....\pi k|\alpha/K....\alpha/K)) &= \frac{a-1}{\lambda_k} - b + \sum_{n=1}^{N} r_{nk}\frac{x_n}{\lambda_k} - r_{nk} = 0 \\
b + \sum_{n=1}^{N} r_{nk} &= \frac{1}{\lambda_k}\big(a - 1 + \sum_{n=1}^{N} r_{nk}x_n\big) \\
\lambda_k &= \frac{\big(a - 1 + \sum_{n=1}^{N} r_{nk}x_n\big)}{b + \sum_{n=1}^{N} r_{nk}}
\end{aligned}
$$

(h) Using langranges mulitplier μ

$$
\begin{aligned}
\mathcal{L}(\mathbf{X}, \pi, \lambda a, b, \alpha, K) &= \ln(\mathbf{X}, \pi, \lambda|a, b, \alpha, K) + \mu(\sum_{j=1}^{K} \pi_k - 1) \\
\frac{\partial}{\partial \pi_k}\mathcal{L}(\mathbf{X}, \pi, \lambda a, b, \alpha, \mathcal{K}) &= \frac{1}{\pi_k}(\frac{\alpha}{K} - 1) + (\sum_{n=1}^{N} \frac{1}{\pi_k} r_n k) + \mu = 0 \\
\pi_k \mu &= \left(1 - \frac{\alpha}{K} - N_k\right) \qquad \Rightarrow *
\end{aligned}
$$

Summing over k

$$
\begin{aligned}
\mu \sum_{k}^{K} \pi_k &= \sum_{k}^{K} \left(1 - \frac{\alpha}{K} - N_k\right) \\
\mu &= K - \alpha - \sum_{k}^{K} N_k \\
\mu &= K - \alpha - N \\
\mu &= -(-K + \alpha + N)
\end{aligned}
$$

Plugging a in *

$$
\begin{aligned}
\pi_k &= \frac{\left(1 - \frac{\alpha}{K} - N_k\right)}{-(-K + \alpha + N)} \\
\pi_k &= \frac{\left(-1 + \frac{\alpha}{K} + N_k\right)}{-K + \alpha + N}
\end{aligned}
$$

(i)

1. Initialize π_k, λ_k. Select a reasonable tolerance ϵ

.

2. Evaluate the responsibilities using the current parameter values E-step:$\forall k, \forall n$:

$$
r_n k = \frac{\pi_k \lambda_k^{x_n} \exp\left(-\lambda_k\right)}{\sum_{j=1}^{K} \pi_j^{x_n} \exp\left(-\lambda_j\right)}
$$

3 Re-estimate the parameters using the current responsibilities M-Step : $\forall k$:

$$
\begin{aligned}
\lambda_k^{new} &= \frac{\big(a - 1 + \sum_{n=1}^{N} r_{nk}x_n\big)}{b + \sum_{n=1}^{N} r_{nk}} \\
\pi_k^{new} &= \frac{\left(-1 + \frac{\alpha}{K} + N_k\right)}{-K + \alpha + N}
\end{aligned}
$$

(c) Evalutate the log-joint distribution based on new parameters as calculated in (f)

and check for convergence of either the parameters or the log likelihood. If the convergence criterion is not satisfied return to **E**-Step

2 Second Exercise

$$\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

(a)

$$\hat{x} = x - \bar{x}$$

(b)

$$\frac{1}{N}\sum_{n=1}^{N} \hat{x_n} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \bar{x})$$

$$\begin{aligned}
\frac{1}{N}\sum_{n=1}^{N}(x_n - \bar{x}) &= \frac{1}{N}\left(\sum_{n=1}^{N} x_n - \bar{x}\sum_{n=1}^{N} 1\right) \\
&= \frac{1}{N}\sum_{n=1}^{N} x_n - \frac{1}{N}\bar{x}N \\
&= \bar{x} - \bar{x} \\
&= \mathbf{0}
\end{aligned}$$

(c)

$$\mathbf{S} = \hat{\mathbf{X}}\hat{\mathbf{X}}^T$$

(d)

$$\text{S.shape} = D\text{x}D$$

(e)

$$\mathbf{y}_n = \Lambda_k^{-1/2}\mathbf{U}_k^T\hat{x}_n$$

$\mathbf{U}_k$ is matrix with k eigen-vectors

$$
\begin{aligned}
\mathbf{L} &= \Lambda_k^{-1/2}\mathbf{U}_k^T \\
\frac{1}{N}\sum_{n=1}^{N}\mathbf{y}_n &= \frac{1}{N}\sum_{n=1}^{N}\Lambda_k^{-1/2}\mathbf{U}_k^T\hat{x}_n \\
&= \frac{1}{N}\Lambda_k^{-1/2}\mathbf{U}_k^T\sum_{n=1}^{N}\hat{x}_n \\
&= \Lambda_k^{-1/2}\mathbf{U}_k^T\frac{1}{N}\sum_{n=1}^{N}(x_n-\bar{x}) \\
&= \Lambda_k^{-1/2}\mathbf{U}_k^T\left(\frac{1}{N}\sum_{n=1}^{N}x_n-\frac{1}{N}\bar{x}N\right) \\
&= \Lambda_k^{-1/2}\mathbf{U}_k^T\underbrace{(\hat{x}-\hat{x})}_{\mathbf{0}} \\
&= \mathbf{0} \\
\mathbf{Cov} &= \sum_{n=1}^{N}\mathbf{y}_n\mathbf{y}_n^T \\
&= \Lambda_k^{-1/2}\mathbf{U}_k^T\left(\sum_{n=1}^{N}\hat{x}_n\hat{x}_n^T\right)\mathbf{U}_k\Lambda_k^{-1/2} \\
&= \Lambda_k^{-1/2}\mathbf{U}_k^T\left(\sum_{n=1}^{N}(x_n-\bar{x}_n)(x_n-\bar{x}_n)^T\right)\mathbf{U}_k\Lambda_k^{-1/2} \\
&= \Lambda_k^{-1/2}\mathbf{U}_k^T\mathbf{S}\mathbf{U}_k\Lambda_k^{-1/2} \\
&= \Lambda_k^{-1/2}\Lambda_k\Lambda_k^{-1/2} \\
&= \mathbf{I}_k
\end{aligned}
$$

Well this is called Whitening or Sphering