2201111783
Jiangyong Huang

Homework 2

November 21, 2022
Statistical Vision

Problem 1

1. Z is defined as $Z = \int \exp\{-\sum_{i=1}^{K} < \lambda_i, H_i(I) >\} dI$. So we have:

$$\frac{\partial \log Z}{\partial \lambda_i} = \frac{1}{Z} \int -H_i(I) \exp\{-\sum_{i=1}^{K} < \lambda_i, H_i(I) >\} dI = -E_p[H_i(I)]$$

2. Take second derivative w.r.t. λ_j, we have:

$$\begin{aligned}
-\frac{\partial^2 \log Z}{\partial \lambda_i \partial \lambda_j} &= \frac{\partial}{\partial \lambda_j} \left[\frac{1}{Z} \int H_i(I) \exp\{- \sum_{i=1}^{K} < \lambda_i, H_i(I) >\} dI \right] \\
&= \frac{\int -H_i(I) H_j(I) \exp\{- \sum_{i=1}^{K} < \lambda_i, H_i(I) >\} dI}{Z} + E_p[H_i(I)] E_p[H_j(I)] \\
&= -E_p[H_i(I) H_j(I)] + h_i h_j = -E_p[(H_i(I) - h_i)(H_j(I) - h_j)]
\end{aligned}$$

3. Given the partition functions of the two model p, p_+ are Z, Z_+ respectively. We can derive that:

$$\begin{aligned}
KL(f||p) - KL(f||p_+) &= \int f \log \frac{f}{p} - \int f \log \frac{f}{p_+} \\
&= \int f \log \frac{p_+}{p} \\
&= E_f[\log p_+] - E_f[\log p] \\
&= E_f \left[-\sum_{i=1}^{K} < \lambda_i^*, H_i > - < \lambda_+, H_+ > - \log Z_+ \right] \\
&\quad - E_f \left[-\sum_{i=1}^{K} < \lambda_i, H_i > - \log Z \right] \\
&= \left[-\sum_{i=1}^{K} < \lambda_i^*, E_f[H_i] > - < \lambda_+, E_f[H_+] > - \log Z_+ \right] \\
&\quad - \left[-\sum_{i=1}^{K} < \lambda_i, E_f[H_i] > - \log Z \right] \\
&= \left[-\sum_{i=1}^{K} < \lambda_i^*, E_{p_+}[H_i] > - < \lambda_+, E_{p_+}[H_+] > - \log Z_+ \right] \\
&\quad - \left[-\sum_{i=1}^{K} < \lambda_i, E_{p_+}[H_i] > - \log Z \right] \\
&= E_{p_+}[\log p_+] - E_{p_+}[\log p] = KL(p_+||p)
\end{aligned}$$

Problem 2

1. The constraints $E_p[H_i] = h_i$ means $\int pH_i - h_i = 0$. Incorporating Lagrange multipliers, we can write the Lagrange function:
$$L(p) = \int p \log \frac{p}{q} + \sum \lambda_i (\int pH_i - h_i) + \lambda_0 (\int p - 1)$$
Using Euler-Lagrange equation, we have:
$$\frac{\delta L}{\delta p} = \log p + 1 + \sum \lambda_i H_i + \lambda_0 = 0$$
Thus, $p(I) = \exp\{-1 - \lambda_0 - \sum \lambda_i H_i(I)\}$.

2. Referring to the derivation in Problem 1, we have $E_p[\log p] = E_f[\log p]$ and $E_p[\log q] = E_f[\log q]$. So it is easy to verify $KL(f||q) - KL(f||p) = KL(p||q)$.

3. When $q(I)$ is a uniform distribution, which means $q(I)$ is a constant over the image space, minimizing $\int p \log \frac{p}{q}$ is equivalent to minimizing $\int p \log p$, which is maximizing $-\int p \log p$. This is exactly maximum entropy principle.

Problem 3

1. z_k is defined as $z_k = \int q_{k-1} \exp(-\beta_k h_k)$. So we have:
$$\frac{\partial \log z_k}{\partial \beta_k} = \frac{1}{z_k} \int -h_k q_{k-1} \exp(-\beta_k h_k) = -E_{q_k}[h_k]$$

2. The normalizing function Z_k is formulated as $Z_k = \int q_0 \exp\{-\sum_{i=1}^{k} \beta_i h_i\}$. Similar to Part 2 in Problem 1, the second derivative will produce a covariance-like form:
$$\begin{aligned}
\frac{\partial^2 \log Z_k}{\partial \beta_i \partial \beta_j} &= \frac{\partial}{\partial \beta_j} \left[-\frac{1}{Z_k} \int h_i q_0 \exp\{-\sum_{i=1}^{k} \beta_i h_i\} \right] \\
&= \frac{\int h_i h_j q_0 \exp\{-\sum_{i=1}^{k} \beta_i h_i\}}{Z_k} - E_{q_k}[h_i] E_{q_k}[h_j] \\
&= E_{q_k}[(h_i - h_i^{obs})(h_j - h_j^{obs})]
\end{aligned}$$

3. Similar to Part 3 in Problem 1, transform the KL divergence into terms of expectation:
$$\begin{aligned}
KL(f||q_k) - KL(f||q_{k+1}) &= E_f[\log q_{k+1}] - E_f[\log q_k] \\
&= E_f[\log q_k] - E_f[\beta_{k+1} h_{k+1}] - z_{k+1} - E_f[\log q_k] \\
&= -\beta_{k+1} E_f[h_{k+1}] - z_{k+1}
\end{aligned}$$

On the other hand, the right side can be formulated as:

$$\begin{aligned}KL(q_{k+1}||q_k) &= E_{q_{k+1}}[\log q_{k+1}] - E_{q_{k+1}}[\log q_k] \\ &= E_{q_{k+1}}[\log q_k] - E_{q_{k+1}}[\beta_{k+1}h_{k+1}] - z_{k+1} - E_{q_{k+1}}[\log q_k] \\ &= -\beta_{k+1}E_{q_{k+1}}[h_{k+1}] - z_{k+1}\end{aligned}$$

Given $E_{q_{k+1}}[h_{k+1}] = E_f[h_{k+1}]$, the equality is proved.

Problem 4

1. The cardinality of $\Omega(q)$ is a combination number C_N^{qN}. Hence, $|\Omega(0.2)| = C_N^{0.2N}$ and $|\Omega(0.5)| = C_N^{0.5N}$.
2. For each sequence $S_N \in \Omega(q)$, the probability of its occurrence is:
$$p(S_N) = p^{qN}(1-p)^{(1-q)N}$$
The total probability mass for all sequences is:
$$p\left(\Omega(q)\right) = C_N^{qN}p^{qN}(1-p)^{(1-p)N}$$
3. Using Stirling approximation $N! \sim \sqrt{2\pi N}(\frac{N}{e})^N$, we have:
$$\log N! = N\log N - N + O(\log N)$$
Therefore, we can take logarithm of the probability $p\left(\Omega(q)\right)$ and factorize as follows:
$$\begin{aligned}\log p\left(\Omega(q)\right) &= \log N! - \log(qN)! - \log[(1-q)N]! + qN\log p + (1-q)N\log(1-p) \\ &= (N\log N - N) - (qN\log(qN) - qN) - \big((1-q)N\log[(1-q)N] - (1-q)N\big) \\ &\quad + O(\log N) + qN\log p + (1-q)N\log(1-p) \\ &= \big(N\log N - qN\log N - (1-q)N\log N\big) - \big(N - qN - (1-q)N\big) \\ &\quad - qN\log q - (1-q)N\log(1-q) + qN\log p + (1-q)N\log(1-p) + O(\log N) \\ &= N\big[q\log\frac{p}{q} + (1-q)\log\frac{1-p}{1-q}\big] + O(\log N)\end{aligned}$$
Consider the coefficient of N as a function $c(q)$ w.r.t. q, that is,
$$c(q) = q\log\frac{p}{q} + (1-q)\log\frac{1-p}{1-q}$$
Take derivative w.r.t. q, we have:
$$\frac{\partial c}{\partial q} = \log\frac{p(1-q)}{q(1-p)}$$
So, $c(q)$ reaches the maximum at $q = p$, which is $c(p) = 0$. Consequently, when $N \to \infty$, $\log p\big(\Omega(q)\big)$ goes to $-\infty$ and thus $p\big(\Omega(q)\big)$ goes to 0 at anywhere except $q = p$. This indicates that only sequences from the type p, i.e. set $\Omega(p)$, can be observed.