Chapter 10 Variational Inference

Yue Yu

10.1 Variational Inference

(10.6) – (10.9)

If we focus on the factor $q_j(\mathbf{Z}_j)$,

$$\begin{aligned}
\mathcal{L}(q) &= \int q(\mathbf{Z}) \ln \left\{ \frac{p(\mathbf{X}, \mathbf{Z})}{q(\mathbf{Z})} \right\} \mathrm{d}\mathbf{Z} \\
&= \int \prod_i q_i \left\{ \ln p(\mathbf{X}, \mathbf{Z}) - \sum_i \ln q_i \right\} \mathbf{dZ} \\
&= \int q_j \Bigg\{ \underbrace{\int \cdots \int \ln p(\mathbf{X}, \mathbf{Z}) \prod_{i \neq j} q_i \, \mathrm{d}\mathbf{Z}_i}_{\mathbb{E}_{i \neq j}[\ln p(\mathbf{X}, \mathbf{Z})]} \Bigg\} \mathrm{d}\mathbf{Z}_j - \int q_j \ln q_j \, \mathrm{d}\mathbf{Z}_j + \text{const} \\
&= -\mathrm{KL}(q_j \| q_j^*) + \text{const}
\end{aligned}$$

where we defined a new distribution q_j^* that satisfies

$$\ln q_j^* = \mathbb{E}_{i \neq j}[\ln p(\mathbf{X}, \mathbf{Z})] + \text{const}.$$

Since KL divergence is non-negative, the lower bound $\mathcal{L}(q)$ reaches the maximum when $\mathrm{KL}(q_j \| q_j^*) = 0$. In other words, when the optimal value is just q_j^*.

(10.17)

$$\begin{aligned}
\mathrm{KL}(p \| q) &= -\int p(\mathbf{Z}) \ln \left\{ \frac{q(\mathbf{Z})}{p(\mathbf{Z})} \right\} \mathrm{d}\mathbf{Z} \\
&= -\int p(\mathbf{Z}) \left[\sum_{i=1}^{M} \ln q_i(\mathbf{Z}_i) \right] \mathrm{d}\mathbf{Z} + \underbrace{\int p(\mathbf{Z}) \ln p(\mathbf{Z}) \, \mathrm{d}\mathbf{Z}}_{\text{const}} \\
&= -\int p(\mathbf{Z}) \left[\ln q_j(\mathbf{Z}_j) + \sum_{i \neq j}^{M} \ln q_i(\mathbf{Z}_i) \right] \mathrm{d}\mathbf{Z} + \text{const} \\
&= -\int p(\mathbf{Z}) \ln q_j(\mathbf{Z}_j) \, \mathrm{d}\mathbf{Z} + \text{const} \\
&= -\int \cdots \int p(\mathbf{Z}) \ln q_j(\mathbf{Z}_j) \, \mathrm{d}\mathbf{Z}_1 \cdots \mathrm{d}\mathbf{Z}_M + \text{const} \\
&= -\int \left[\int p(\mathbf{Z}) \prod_{i \neq j} \mathrm{d}\mathbf{Z}_i \right] \ln q_j(\mathbf{Z}_j) \, \mathrm{d}\mathbf{Z}_j + \text{const} \\
&= -\int F_j(\mathbf{Z}_j) \ln q_j(\mathbf{Z}_j) \, \mathrm{d}\mathbf{Z}_j + \text{const}
\end{aligned}$$

where in the forth step, the sum term is absorbed into constant because we assume $q_i(\mathbf{Z}_i)$ is fixed for $i \neq j$ when focusing on $q_j(\mathbf{Z}_j)$. Also, we define

$$F_j(\mathbf{Z}_j) = \int p(\mathbf{Z}) \prod_{i \neq j} \mathrm{d}\mathbf{Z}_i.$$

Then, the corresponding Lagrangian is given by

$$L(q_j(\mathbf{Z}_j)) = -\int F_j(\mathbf{Z}_j) \ln q_j(\mathbf{Z}_j) \, \mathrm{d}\mathbf{Z}_j + \lambda \left(\int q_j(\mathbf{Z}_j) \, \mathrm{d}\mathbf{Z}_j - 1 \right).$$

Identifying $F_j(\mathbf{Z}_j) \ln q_j(\mathbf{Z}_j)$ in the first term as $G_1(q_j(\mathbf{Z}_j))$, and $q_j(\mathbf{Z}_j)$ as $G_2(q_j(\mathbf{Z}_j))$ in the second term, according to the Euler-Lagrange equation, we have

$$-\frac{F_j(\mathbf{Z}_j)}{q_j(\mathbf{Z}_j)} + \lambda = 0,$$

which implies that

$$\lambda q_j(\mathbf{Z}_j) = F_j(\mathbf{Z}_j). \tag{*}$$

Integrating both sides over $\mathbf{Z}_j$, we have

$$\begin{aligned} \lambda &= \int F_j(\mathbf{Z}_j) \, \mathrm{d}\mathbf{Z}_j \\ &= \int \left[\int p(\mathbf{Z}) \prod_{i \neq j} \mathrm{d}\mathbf{Z}_i \right] \mathrm{d}\mathbf{Z}_j \\ &= \int p(\mathbf{Z}) \, \mathrm{d}\mathbf{Z} \\ &= 1. \end{aligned}$$

Substituting back into (*), we obtain the optimal value for $q_j(\mathbf{Z}_j)$, which is

$$q_j^*(\mathbf{Z}_j) = F_j(\mathbf{Z}_j) = \int p(\mathbf{Z}) \prod_{i \neq j} \mathrm{d}\mathbf{Z}_i.$$