Statistical Learning 2024 Fall

Exam 2

Chenchuan He

---

1. Prove ELBO

Solution

$$\begin{aligned} D_{KL}\left(q_\phi(\mathbf{z}|\mathbf{x})\|p_\theta(\mathbf{z}|\mathbf{x})\right) &= \int_{\mathbf{z}} q_\phi(\mathbf{z}|\mathbf{x}) \log \frac{q_\phi(\mathbf{z}|\mathbf{x})}{p_\theta(\mathbf{z}|\mathbf{x})} d\mathbf{z} \\ &= \int_{\mathbf{z}} q_\phi(\mathbf{z}|\mathbf{x}) \left(\log q_\phi(\mathbf{z}|\mathbf{x}) - \log p_\theta(\mathbf{z}|\mathbf{x})\right) d\mathbf{z} \\ &= \int_{\mathbf{z}} q_\phi(\mathbf{z}|\mathbf{x}) \left(\log q_\phi(\mathbf{z}|\mathbf{x}) - \log p_\theta(\mathbf{z}|\mathbf{x})\right) d\mathbf{z} \end{aligned}$$

while

$$\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] = \int_{\mathbf{z}} q_\phi(\mathbf{z}|\mathbf{x}) \log p_\theta(\mathbf{x}|\mathbf{z}) dz$$

and

$$\begin{aligned} -D_{KL}\left(q_\phi(\mathbf{z}|\mathbf{x})\|p_\theta(\mathbf{z})\right) &= -\int_{\mathbf{z}} q_\phi(\mathbf{z}|\mathbf{x}) \log \frac{q_\phi(\mathbf{z}|\mathbf{x})}{p_\theta(\mathbf{z})} d\mathbf{z} \\ &= -\int_{\mathbf{z}} q_\phi(\mathbf{z}|\mathbf{x}) \log q_\phi(\mathbf{z}|\mathbf{x}) d\mathbf{z} + \int_{\mathbf{z}} q_\phi(\mathbf{z}|\mathbf{x}) \log p_\theta(\mathbf{z}) d\mathbf{z} \end{aligned}$$

Add the three terms above together, we get

$$\begin{aligned} & D_{KL}\left(q_\phi(\mathbf{z}|\mathbf{x})\|p_\theta(\mathbf{z}|\mathbf{x})\right) + \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] - D_{KL}\left(q_\phi(\mathbf{z}|\mathbf{x})\|p_\theta(\mathbf{z})\right) \\ &= \int_{\mathbf{z}} q_\phi(\mathbf{z}|\mathbf{x}) \log q_\phi(\mathbf{z}|\mathbf{x}) d\mathbf{z} - \int_{\mathbf{z}} q_\phi(\mathbf{z}|\mathbf{x}) \log p_\theta(\mathbf{z}|\mathbf{x}) d\mathbf{z} \\ &\quad + \int_{\mathbf{z}} q_\phi(\mathbf{z}|\mathbf{x}) \log p_\theta(\mathbf{x}|\mathbf{z}) d\mathbf{z} \\ &\quad - \int_{\mathbf{z}} q_\phi(\mathbf{z}|\mathbf{x}) \log q_\phi(\mathbf{z}|\mathbf{x}) d\mathbf{z} + \int_{\mathbf{z}} q_\phi(\mathbf{z}|\mathbf{x}) \log p_\theta(\mathbf{z}) d\mathbf{z} \\ &= -\int_{\mathbf{z}} q_\phi(\mathbf{z}|\mathbf{x}) \log p_\theta(\mathbf{z}|\mathbf{x}) d\mathbf{z} \\ &\quad + \int_{\mathbf{z}} q_\phi(\mathbf{z}|\mathbf{x}) \log p_\theta(\mathbf{x}|\mathbf{z}) d\mathbf{z} + \int_{\mathbf{z}} q_\phi(\mathbf{z}|\mathbf{x}) \log p_\theta(\mathbf{z}) d\mathbf{z} \\ &= \log p_\theta(\mathbf{x}) \int_{\mathbf{z}} q_\phi(\mathbf{z}|\mathbf{x}) d\mathbf{z} \\ &= \log p_\theta(\mathbf{x}) \end{aligned}$$