MA060129, Bayesian ML Homework 3: Theoretical Problems

November 21, 2022

Problem 1 [20 pts]

We consider the VAE model. In order to train it, we usually optimize the $\mathcal{L}$(ELBO) :

$$
\begin{aligned}
\log p(x) \geq \mathbb{E}_{q_\phi(z|x)} \log \frac{p_\theta(x \mid z) p(z)}{q_\phi(z \mid x)} &= \mathcal{L}(x), \text{ where:} \\
p(x) = \int p_\theta(x \mid z) p(z) dz &- \text{ likelihood} \\
p(z) &- \text{ prior over the latent space,} \\
q_\phi(z \mid x) &- \text{ approximation for intractable posterior } p(z \mid x)
\end{aligned}
$$

However, the ELBO is not the only objective to optimize. Let's consider another one:

$$
\mathcal{L}_k = \mathbb{E}_{z_1, \ldots, z_k \sim q_\phi(z|x)} \log \frac{1}{k} \sum_{i=1}^{k} \frac{p_\theta\left(x \mid z_i\right) p\left(z_i\right)}{q_\phi\left(z_i \mid x\right)}
$$

Your task to prove several properties:

1. Prove that $\mathcal{L}_k$ is a lower bound of the marginal log likelihood:

$$
\mathcal{L}_k \leq \log p(x), \ \forall k \in \mathbb{N}
$$

2. Given that:

$$
z = g_\phi(\varepsilon), \ \varepsilon \sim p(\varepsilon)
$$

and denote:

$$
w_i = \frac{p_\theta\left(x \mid z_i\right) p\left(z_i\right)}{q_\phi\left(z_i \mid x\right)}
$$

Compute $\nabla_{\theta,\phi} \mathcal{L}_1$ and $\nabla_{\theta,\phi} \mathcal{L}_k$ using reparametrization trick. The final answer should only contain w and $\nabla_{\theta,\phi} \log w$

Solution

1. By Jensen's inequality

$$
\mathbb{E}_{z_1, \ldots, z_k \sim q_\phi(z|x)} \log \frac{1}{k} \sum_{i=1}^{k} \frac{p_\theta\left(x \mid z_i\right) p\left(z_i\right)}{q_\phi\left(z_i \mid x\right)} \leq \log \frac{1}{k} \sum_{i=1}^{k} \mathbb{E}_{z_i \sim q_\phi(z|x)} \frac{p_\theta\left(x \mid z_i\right) p\left(z_i\right)}{q_\phi\left(z_i \mid x\right)} = \log \mathbb{E}_{q_\phi(z|x)} \frac{p_\theta\left(x \mid z\right) p\left(z\right)}{q_\phi\left(z \mid x\right)} =
$$

$$
= \log \int q_\phi(z \mid x) \frac{p_\theta\left(x \mid z\right) p\left(z\right)}{q_\phi\left(z \mid x\right)} dz = \log p(x)
$$

2.

$$
\nabla_{\theta,\phi} \mathcal{L}_1 = \nabla_{\theta,\phi} \mathbb{E}_{p(\varepsilon)} \log w_1 = \mathbb{E}_{p(\varepsilon)} \nabla_{\theta,\phi} \log w_1
$$

$$
\nabla_{\theta,\phi} \mathcal{L}_k = \nabla_{\theta,\phi} \mathbb{E}_{p(\varepsilon)} \log \frac{1}{k} \sum_{i=1}^{k} w_k = \mathbb{E}_{p(\varepsilon)} \nabla_{\theta,\phi} \log \sum_{i=1}^{k} w_k
$$

Problem 2 [20 pts]

By far, we have assumed that prior over the latent space $p(z)$ is fixed. But we can also use Empirical Bayes approach in order to find the optimal prior distribution. Given the dataset $\{x_n\}_{n=1}^N$, or similarly you are given an empirical distribution $p_e(x) = \sum_{n=1}^N \delta_{x_n}(x)$. Find the prior distribution $p^*(z)$, which maximize the ELBO averaged over this dataset:

$$p^*(z) = \arg\max_{p(z)} \mathbb{E}_{p_e(x)}\mathcal{L}(x), \quad s.t. \int p(z)dz = 1, \forall z : p(z) \geq 0$$

Solution

$$\mathcal{L}(x) = \mathbb{E}_{q(z|x)} \log \frac{p(x \mid z)p(z)}{q(z \mid x)} = \int q(z \mid x) \log \frac{p(x \mid z)p(z)}{q(z \mid x)} dz = -KL\left(q(z \mid x) \mid\mid p(z)\right) + \int q(z \mid x) \log p(x \mid z) dz$$

the second term doesn't depend on $p(z)$ and the first is independent from $p_e(x) \Rightarrow$

$$p^*(z) = \arg\max_{p(z)} \left[-KL\left(q(z \mid x) \mid\mid p(z)\right)\right] = q(z \mid x)$$