# MathDNN HW 11

2018-13260 차재현

## 1 Problem 1

(a) Let $\Omega$ be the image of the random variable $Z$. Since log is concave, Jensen's inequality gives

$$\mathbb{E}_{Z_1,\cdots,Z_K\sim q_\phi(z|x)}\left[\log\frac{1}{K}\sum_{k=1}^{K}\frac{p_\theta(x|Z_k)p_Z(Z_k)}{q_\phi(Z_k|x)}\right]$$

$$\leq \log\mathbb{E}_{Z_1,\cdots,Z_K\sim q_\phi(z|x)}\left[\frac{1}{K}\sum_{k=1}^{K}\frac{p_\theta(x|Z_k)p_Z(Z_k)}{q_\phi(Z_k|x)}\right]$$

$$= \log\frac{1}{K}\sum_{k}\int_{\Omega^{K-1}}\left(\int_\Omega\frac{p_\theta(x|z_k)p_Z(z_k)}{q_\phi(z_k|x)}q_\phi(z_k|x)dz_k\right)\prod_{i\neq k}q_\phi(z_i|x)dz_i$$

$$= \log\frac{1}{K}\sum_{k}\int_{\Omega^{K-1}}\left(\int_\Omega p_\theta(x|z_k)p_Z(z_k)dz_k\right)\prod_{i\neq k}q_\phi(z_i|x)dz_i$$

$$= \log\frac{1}{K}\sum_{k}p_\theta(x) = \log p_\theta(x).$$

(b) Let $\mathscr{K} = \{1,\cdots,K\}$, $S = \{I\subset\mathscr{K} \mid |I| = M\}$. Define a uniform measure $\delta$ on $2^S$ as usual. Let $P_Z$ be a function defined on $S$ by

$$P_Z(I) = \sum_{k\in I}\frac{1}{M}\frac{p_\theta(x|Z_k)p_Z(Z_k)}{q_\phi(Z_k|x)}.$$

Since $\delta$ is a probability measure with symmetric to every $k\in\mathscr{K}$, together with Jensen's inequality we get

$$\mathbb{E}_{Z_1,\cdots,Z_K\sim q_\phi(z|x)}\left[\log\frac{1}{K}\sum_{k=1}^{K}\frac{p_\theta(x|Z_k)p_Z(Z_k)}{q_\phi(Z_k|x)}\right]$$

$$= \mathbb{E}_{Z_1,\cdots,Z_K\sim q_\phi(z|x)}\left[\log\mathbb{E}_{I\sim\delta}\left[P_Z(I)\right]\right]$$

$$\geq \mathbb{E}_{Z_1,\cdots,Z_K\sim q_\phi(z|x)}\left[\mathbb{E}_{I\sim\delta}\log P_Z(I)\right].$$

Now, since summand of last part is integrable for both measures, we can apply Fubini theorem to interchange summation symbol to get

$$\begin{aligned}&\mathbb{E}_{Z_1,\cdots,Z_K\sim q_\phi(z|x)}\left[\mathbb{E}_{I\sim\delta}\log P_Z(I)\right]\\&=\mathbb{E}_{I\sim\delta}\left[\mathbb{E}_{Z_1,\cdots,Z_K\sim q_\phi(z|x)}\log P_Z(I)\right]\\&=\mathbb{E}_{I\sim\delta}\left[\mathbb{E}_{Z_1,\cdots,Z_K\sim q_\phi(z|x)}\log P_Z(I)\right].\end{aligned}$$

Finally, independence condition of $Z$ implies every $\mathbb{E}_{Z_1,\cdots,Z_K\sim q_\phi(z|x)}\left[\log P_Z(I)\right]$ is same, and definition of $\delta$ gives

$$\begin{aligned}&\mathbb{E}_{I\sim\delta}\left[\mathbb{E}_{Z_1,\cdots,Z_K\sim q_\phi(z|x)}\log P_Z(I)\right]\\&=\mathbb{E}_{Z_1,\cdots,Z_M\sim q_\phi(z|x)}\left[\log\frac{1}{M}\sum_{k=1}^{M}\frac{p_\theta(x|Z_k)p_Z(Z_k)}{q_\phi(Z_k|x)}\right]\end{aligned}$$

(c) Given condition holds when inequality proved in (a) becomes equal. According to equality condition of Jensen's inequality, such happens if and only if

$$\frac{p_\theta(x|z)p_Z(z)}{q_\phi(z|x)}=\text{const a.e.}$$

## 2 Problem 2

(a) By the same logic as (a) in problem 1, we have

$$\begin{aligned}\mathrm{VLB}_{\theta,\phi,\lambda}(X_i)&=\mathbb{E}_{Z\sim q_\phi(z|X_i)}\left[\log\frac{p_\theta(X_i|Z)r_\lambda(Z)}{q_\phi(Z|X_i)}\right]\\&\le\log\mathbb{E}_{Z\sim q_\phi(z|X_i)}\left[\frac{p_\theta(X_i|Z)r_\lambda(Z)}{q_\phi(Z|X_i)}\right]\\&=p_\theta(X_i|Z)\end{aligned}$$

(b) It is enough to calculate $\nabla_\theta,\nabla_\phi,\nabla_\lambda$ of given expectation.

i) Differentiate by $\theta$, we get

$$\begin{aligned}\nabla_\theta\,\mathrm{VLB}_{\theta,\phi,\lambda}(X_i)&=\mathbb{E}_{Z\sim q_\phi(z|X_i)}\left[\nabla_\theta\log\frac{p_\theta(X_i|Z)r_\lambda(Z)}{q_\phi(Z|X_i)}\right]\\&=\mathbb{E}_{Z\sim q_\phi(z|X_i)}\left[\frac{\nabla_\theta p_\theta(X_i|Z)}{p_\theta(X_i|Z)}\right]\\&=\mathbb{E}_{Z\sim q_\phi(z|X_i)}\left[\nabla_\theta\log p_\theta(X_i|Z)\right]\end{aligned}$$

ii) Differentiate by $\phi$ with log-derivative trick, we get

$$\begin{aligned}
&\nabla_\phi \,\mathrm{VLB}_{\theta,\phi,\lambda}(X_i) = \nabla_\phi \mathbb{E}_{Z\sim q_\phi(z|X_i)}\left[\log \frac{p_\theta(X_i|Z)r_\lambda(Z)}{q_\phi(Z|X_i)}\right] \\
&= \nabla_\phi \int_z \log\left(\frac{p_\theta(X_i|z)r_\lambda(z)}{q_\phi(z|X_i)}\right) q_\phi(z|X_i)dz \\
&= \int_z \log\left(\frac{p_\theta(X_i|z)r_\lambda(z)}{q_\phi(z|X_i)}\right) \nabla_\phi q_\phi(z|X_i) - \frac{\nabla_\phi q_\phi(z|X_i)}{q_\phi(z|X_i)} q_\phi(z|X_i)dz \\
&= \int_z \left(\log\left(\frac{p_\theta(X_i|z)r_\lambda(z)}{q_\phi(z|X_i)}\right) - 1\right) \frac{\nabla_\phi q_\phi(z|X_i)}{q_\phi(z|X_i)} q_\phi(z|X_i)dz \\
&= \mathbb{E}_{Z\sim q_\phi(z|X_i)}\left[\left(\log\left(\frac{p_\theta(X_i|Z)r_\lambda(Z)}{q_\phi(Z|X_i)}\right) - 1\right) \nabla_\phi \log q_\phi(Z|X_i)\right]
\end{aligned}$$

iii) Differentiate by $\lambda$, we get

$$\begin{aligned}
\nabla_\lambda \,\mathrm{VLB}_{\theta,\phi,\lambda}(X_i) &= \mathbb{E}_{Z\sim q_\phi(z|X_i)}\left[\nabla_\lambda \log \frac{p_\theta(X_i|Z)r_\lambda(Z)}{q_\phi(Z|X_i)}\right] \\
&= \mathbb{E}_{Z\sim q_\phi(z|X_i)}\left[\frac{\nabla_\lambda r_\lambda(Z)}{r_\lambda(Z)}\right] \\
&= \mathbb{E}_{Z\sim q_\phi(z|X_i)}\left[\nabla_\lambda \log r_\lambda(Z)\right]
\end{aligned}$$

(c)

i) Let $p(x)$ be a pdf of standard normal distribution. Using reparameterization trick, we can rewrite $p_\theta$ as

$$p_\theta(x|z) = p\left(\frac{x - f_\theta(z)}{\sigma}\right) = \frac{1}{\sqrt{(2\pi)^k}} \exp\left(-\frac{\|x - f_\theta(z)\|^2}{2\sigma^2}\right).$$

Taking log and differentiate by $\theta$, we get

$$\begin{aligned}
\nabla_\theta \,\mathrm{VLB}_{\theta,\phi,\lambda}(X_i) &= \mathbb{E}_{Z\sim q_\phi(z|X_i)}\left[\nabla_\theta \log p_\theta(X_i|Z)\right] \\
&= \mathbb{E}_{Z\sim q_\phi(z|X_i)}\left[-\nabla_\theta \frac{\|X_i - f_\theta(Z)\|^2}{2\sigma^2}\right] \\
&= \mathbb{E}_{Z\sim q_\phi(z|X_i)}\left[\frac{(X_i - f_\theta(Z))}{\sigma^2} \nabla_\theta f_\theta(Z)\right]
\end{aligned}$$

ii) By the same way as i) with $N(\mu_\phi, \Sigma_\phi) = \mu_\phi + \Sigma_\phi N(\mathbf{0}, I)$, we have

$$\nabla_\phi \,\mathrm{VLB}_{\theta,\phi,\lambda}(X_i) = \mathbb{E}_{Z\sim N(\mathbf{0},I)}\left[\nabla_\phi \log \frac{p_\theta(X_i|\mu_\phi + \Sigma_\phi Z)r_\lambda(\mu_\phi + \Sigma_\phi Z)}{q_\phi(\mu_\phi + \Sigma_\phi Z|X_i)}\right]$$

$$= \mathbb{E}_{Z\sim N(\mathbf{0},I)}[\nabla_\phi \log p_\theta(X_i|\mu_\phi + \Sigma_\phi Z)$$

$$+ \nabla_\phi \log r_\lambda(\mu_\phi + \Sigma_\phi Z)$$

$$- \nabla_\phi \log q_\phi(\mu_\phi + \Sigma_\phi Z|X_i)]$$

Denote $h_\phi = h_\phi(X_i, Z) = \mu_\phi(X_i) + \Sigma_\phi(X_i)Z$ for simplicity. We have

$$\nabla_\phi h_\phi = \nabla_\phi \mu_\phi(X_i) + \nabla_\phi(\Sigma_\phi(X_i)Z).$$

Now, we divide into parts.

- By the same way as i), we have

$$\nabla_\phi \log p_\theta(X_i|\mu_\phi + \Sigma_\phi Z) = -\nabla_\phi \frac{\|X_i - f_\theta(h_\phi)\|^2}{2\sigma^2}$$

$$= \frac{1}{\sigma^2}(X_i - f_\theta(h_\phi))f'_\theta(h_\phi)\nabla_\phi h_\phi$$

- By the same way, we get

$$\nabla_\phi \log r_\lambda(\mu_\phi + \Sigma_\phi Z) = -\nabla_\phi \frac{1}{2}(h_\phi - \lambda_1)^\top \operatorname{diag}(\lambda_2)^{-1}(h_\phi - \lambda_1)$$

$$= (\lambda_1 - f_\theta(h_\phi))^\top \operatorname{diag}(\lambda_2)^{-1}\nabla_\phi h_\phi.$$

- Apply reparameterization trick once more to get

$$q_\phi(h_\phi|X_i) = p(\Sigma_\phi^{-1}(h_\phi - \mu_\phi)) = p(Z).$$

  Note that $p$ is standard normal distribution. Then

$$-\nabla_\phi \log q_\phi(h_\phi|X_i) = -\nabla_\phi \log p(Z) = 0.$$

Summing up, we get desired results.

iii) Similarly, we have

$$r_\lambda(z) = p\left(\sqrt{\operatorname{diag}(\lambda_2)^{-1}}(z - \lambda_1)\right).$$

Here, since $\operatorname{diag}(\lambda_2)$ is variance matrix and diagonal, each diagonal entry is positive, so its square root exists. We choose one consists of positive diagonal entries. We have

$$\nabla_\lambda \operatorname{VLB}_{\theta,\phi,\lambda}(X_i) = \mathbb{E}_{Z\sim q_\phi(z|X_i)}\left[\nabla_\lambda \log r_\lambda(Z)\right]$$

$$= \mathbb{E}_{Z\sim q_\phi(z|X_i)}\left[-\nabla_\lambda \frac{1}{2}(Z-\lambda_1)^\top \operatorname{diag}(\lambda_2)^{-1}(Z-\lambda_1)\right]$$

$$= \mathbb{E}_{Z\sim q_\phi(z|X_i)}\begin{bmatrix}\Lambda_1\\ \Lambda_2\end{bmatrix}$$

where

$$\Lambda_1 = (Z-\lambda_1)^\top \operatorname{diag}(\lambda_2)^{-1} I_k, \qquad \Lambda_2 = \frac{1}{2}(Z-\lambda_1)^\top \operatorname{diag}(\lambda_2)^{-2}(Z-\lambda_1).$$

## 3 Problem 4

Let $p_A = (q_1, q_2, q_3)^\top$ and $p_B = (p_1, p_2, p_3)^\top$. Then our expression becomes

$$\text{minimize}_{p_A\in\Delta^3}\ \text{maximize}_{p_B\in\Delta^3}\ \mathbb{E}_{p_A,p_B}[\text{points in } B]$$

$$= \text{minimize}_{p_A\in\Delta^3}\ \text{maximize}_{p_B\in\Delta^3}\ \mathbb{E}_{p_A,p_B}[p_2(q_1-q_3)+p_3(q_2-q_1)+p_1(q_3-q_2)]$$

$$= \text{minimize}_{p_A\in\Delta^3}\ \text{maximize}_{p_B\in\Delta^3}\ \mathbb{E}_{p_A,p_B}[q_2(p_3-p_1)+q_3(p_1-p_2)+q_1(p_2-p_3)].$$

Let this function $f = f(p_A, p_B)$.
(a) Assume $p_B \neq (1/3, 1/3, 1/3)^\top$ is a solution of minimax problem. In this case, WLOG we can assume $p_3 - p_1 < 0$. Taking $q_2 = 1$ and others 0, expectation becomes negative. Therefore, if $p_A$ is a solution of minimax problem, $f(p_A, p_B) < 0$. In this case, $p_A \neq (1/3, 1/3, 1/3)^\top$ because if any of $p_A, p_B$ is $(1/3, 1/3, 1/3)^\top$, $f = 0$. However, in this case at least one of $(q_1 - q_3), (q_2 - q_1), (q_3 - q_2)$ is positive, and WLOG if we assume $q_2 - q_1 > 0$, we have

$$f(p_A, p_B) < 0 < f(p_A, (0, 0, 1)^\top).$$

This contradicts to the assumption that $p = p_B$ maximizes $f(p_A, p)$, and therefore $p_B = (1/3, 1/3, 1/3)^\top$. By the same argument we also have $p_A p_B = (1/3, 1/3, 1/3)^\top$.

(b) If player $B$ never changes his strategy, i.e., if player $B$ plays purely randomly with given probability, any strategy of $A$ gives the same result. However, if player $B$ "trains" his strategy, (a) gives that player $A$ can find better strategy that can beats $B$.