MathDNN Homework 2

Department of Computer Science and Engineering
2021-16988 Jaewan Park

Problem 4

Let $\varphi(x) = -\log(x)$, then $\varphi''(x) = 1/x^2$ is nonnegative for all $x \in \mathbb{R}^+$. Therefore $-\log(x)$ is a convex function. Since $\varphi(x)$ is defined over $\mathbb{R}^+$ which is a convex set (it is obvious that $\forall x_1, x_2 \in \mathbb{R}^+$, $\eta x_1 + (1-\eta)x_2\ (>0) \in \mathbb{R}^+$), we can apply Jensen's inequality to $-\log(x)$. Therefore we obtain

$$\begin{aligned}
D_{KL}(p||q) &= \sum_{i=1}^{N} p_i \log\left(\frac{p_i}{q_i}\right) = \sum_{i=1}^{N} p_i \left(-\log\left(\frac{q_i}{p_i}\right)\right) = \sum_{i=1}^{N} p_i \varphi(\frac{q_i}{p_i}) \\
&= \mathbb{E}\left[\varphi\left(\frac{q_i}{p_i}\right)\right] \\
&\geq \varphi\left(\mathbb{E}\left[\frac{q_i}{p_i}\right]\right) \\
&= -\log\left(\sum_{i=1}^{N} p_i \frac{q_i}{p_i}\right) = -\log\left(\sum_{i=1}^{N} q_i\right) = -\log 1 \\
&= 0
\end{aligned}$$

when we let q_i/p_i a random variable with probability mass p_i. In cases where $p_i = 0$ or $q_i = 0$, the value of $p_i \log(p_i/q_i)$ is considered either 0 or ∞, so the sum maintains nonnegative.

Problem 5

Let $\varphi(x) = -\log(x)$, then $\varphi''(x) = 1/x^2$ is positive for all $x \in \mathbb{R}^+$. Therefore $-\log(x)$ is a strictly convex function. As described in **Problem 4**, $\mathbb{R}^+$ is a convex set, and if we let $X = q_i/p_i$ a random variable X is non constant. Therefore we can apply the strict Jensen's inequality to $-\log(x)$. Therefore we obtain

$$\begin{aligned}
D_{KL}(p||q) &= \sum_{i=1}^{N} p_i \log\left(\frac{p_i}{q_i}\right) = \sum_{i=1}^{N} p_i \left(-\log\left(\frac{q_i}{p_i}\right)\right) = \sum_{i=1}^{N} p_i \varphi(\frac{q_i}{p_i}) \\
&= \mathbb{E}\left[\varphi\left(\frac{q_i}{p_i}\right)\right] \\
&> \varphi\left(\mathbb{E}\left[\frac{q_i}{p_i}\right]\right) \\
&= -\log\left(\sum_{i=1}^{N} p_i \frac{q_i}{p_i}\right) = -\log\left(\sum_{i=1}^{N} q_i\right) = -\log 1 \\
&= 0.
\end{aligned}$$

For cases where $p_i = 0$ or $q_i = 0$ the inequality stil remains. If $p_i > 0$ and $q_i = 0$, $p_i \log\left(\frac{p_i}{0}\right) = \infty$, so the sum still maintains positive. If $p_i = 0$ and $q_i = 0$,$p_i \log\left(\frac{p_i}{0}\right) = 0$, but since $p \neq q$, $p_i = q_i = 0$ cannot be true for all i. Thus the sum maintains positive. If $p_i = 0$ and $q_i > 0$, $p_i \log\left(\frac{p_i}{0}\right) = 0$, but since $\sum q_i = 1 > 0$, $q_i = 0$ cannot be true for all i. Thus the sum maintains positive.

Compared to **Problem 4**, the difference is that $p \neq q$, and this results in $D_{KL}(p||q) = 0$ not being possible. Since $\forall i \; p_i \log\left(\frac{p_i}{q_i}\right) \geq 0$, for $D_{KL}(p||q)$ to be 0, $p_i \log\left(\frac{p_i}{q_i}\right) = 0$ should be satisfied at all i. In that $\sum p_i = \sum q_i = 1$, this can only be satisfied when $p = q$. Thus supposing $p \neq q$ eliminates the equality condition from **Problem 4**, resulting in a strict inequality to satisfy.

Problem 6

(a)

$$
\begin{aligned}
\frac{\partial f_\theta(x)}{\partial u_i} &= \frac{\partial}{\partial u_i}\sum_{j=1}^{p} u_j \sigma(a_j x + b_j) \\
&= \sum_{j=1}^{p} \frac{\partial}{\partial u_i} u_j \sigma(a_j x + b_j) = \frac{\partial}{\partial u_i} u_i \sigma(a_i x + b_i) \\
&= \sigma(a_i x + b_i)
\end{aligned}
$$

Therefore

$$
\nabla_u f_\theta(x) = \left(\frac{\partial f_\theta(x)}{\partial u_1}, \cdots, \frac{\partial f_\theta(x)}{\partial u_p}\right) = (\sigma(a_1 x + b_1), \cdots, \sigma(a_p x + b_p)) = \sigma(ax + b).
$$

(b)

$$
\begin{aligned}
\frac{\partial f_\theta(x)}{\partial b_i} &= \frac{\partial}{\partial b_i}\sum_{j=1}^{p} u_j \sigma(a_j x + b_j) \\
&= \sum_{j=1}^{p} \frac{\partial}{\partial b_i} u_j \sigma(a_j x + b_j) = \frac{\partial}{\partial b_i} u_i \sigma(a_i x + b_i) \\
&= u_i \sigma'(a_i x + b_i)
\end{aligned}
$$

Therefore

$$
\nabla_b f_\theta(x) = \left(\frac{\partial f_\theta(x)}{\partial b_1}, \cdots, \frac{\partial f_\theta(x)}{\partial b_p}\right) = (u_1 \sigma'(a_1 x + b_1), \cdots, u_p \sigma'(a_p x + b_p)) = \mathrm{diag}(\sigma'(ax + b))u.
$$

(c)

$$\begin{aligned}
\frac{\partial f_\theta(x)}{\partial a_i} &= \frac{\partial}{\partial a_i}\sum_{j=1}^{p} u_j\sigma(a_jx+b_j) \\
&= \sum_{j=1}^{p}\frac{\partial}{\partial a_i}u_j\sigma(a_jx+b_j) = \frac{\partial}{\partial a_i}u_i\sigma(a_ix+b_i) \\
&= u_ix\sigma'(a_ix+b_i)
\end{aligned}$$

Therefore

$$\nabla_a f_\theta(x) = \left(\frac{\partial f_\theta(x)}{\partial a_1},\cdots,\frac{\partial f_\theta(x)}{\partial a_p}\right) = (u_1x\sigma'(a_1x+b_1),\cdots,u_px\sigma'(a_px+b_p)) = \mathrm{diag}(\sigma'(ax+b))ux.$$