# MathDNN HW 2

2018-13260 차재현

## 1 Problem 4

In this problem. $\sum$ denotes $\sum_{i=1}^{n}$.

The meaning of "probability mass function $p \in \mathbb{R}^n$ is $p : \{1, \cdots, n\} \to \mathbb{R}$ satisfying $\sum p(i) = 1$, and the same is analogous for $q$. Note that $\sum p(i) = \sum q(i) = 1$.

**Lemma 1.1.** The function $y = \log x$ is convex in its range.

*Proof.* Let $0 < x < y$. Define $f(t) = \log((1-t)x + ty) - ((1-t)\log x + t \log y)$, where $0 \le t \le 1$. Our assertion is equivalent to $f \ge 0$ on this range.

Obviously $f(0) = f(1) = 0$. Assume $0 < t < 1$. Fixing $x$, and differentiating $f$ w.r.t y gives

$$\frac{\partial f}{\partial y}(t) = \frac{t}{(1-t)x + y} - \frac{t}{y} > 0$$

because $(1-t)x + y < y$ when $0 < t < 1$. Since $f(t) \equiv 0$ when $x = y$ and $y > x$ by assumption, we get desired result. $\square$

First, assume $p(k) = 0$ for some $i$. Our convention $0 \log 0 = 0$ and $0 \log \frac{0}{0} = 0$ reduces the KL-divergence to

$$D_{\mathrm{KL}}(p||q) = \sum_{i \neq k} p(i) \left( -\log \left( \frac{q(i)}{p(i)} \right) \right). \tag{1}$$

Now, let $c = \frac{1}{\sum_{i \neq k} q(i)}$. If $c = 1$, $D_{\mathrm{KL}}(p||q) = \infty$ unless all $p(i) = 0$, which is impossible. So $c > 1$, and we can rewrite (1) as

$$D_{\mathrm{KL}}(p||q) = \sum_{i \neq k} p(i) \left( -\log \left( \frac{cq(i)}{p(i)} \right) \right) + \log c. \tag{2}$$

Both functions $p, cq$ are probability masses with domain $\{1, \cdots, n\} - \{k\}$, so equation (2) is just adding $\log c > 0$ to the KL-divergence of new $p, cq$. Iterating, we can assume $p \neq 0$ anywhere.

Next, when $q(k) = 0$, our hypothesis $p \neq 0$ induces $D_{\mathrm{KL}}(p||q) = \infty$. So from now, assume $q \neq 0$ anywhere.

Since $p$ is probability mass, for any $f : \{1, \cdots, n\} \to \mathbb{R}$ we have $\mathbb{E}_p(f) = \sum f(i)p(i)$. The function $-\log$ is convex by Lemma 1.1, so Jensen's inequality gives

$$D_{\mathrm{KL}}(p||q) = \sum p(i) \left( -\log \left( \frac{q(i)}{p(i)} \right) \right) \geq -\log \left( \sum p(i) \frac{q(i)}{p(i)} \right) = -\log 1 = 0.$$

## 2 Problem 5

In the proof of Problem 4, $\frac{\partial f}{\partial y} > 0$ if $y > x$. Thus $-\log$ is strictly convex. Also, we can assume $p, q \neq 0$ by the same argument.

If the function $p/q$ is constant, $p = q$. Conversely, when $p \neq q$, random variable $p/q$ ($q \neq 0$ guarantees well-definedness) is non-constant. Thus, applying strict Jensen's inequality, we get $D_{\mathrm{KL}}(p||q) > 0$.

## 3 Problem 6

Denote $f_\theta = f$ for simplicity.

(i) Partial derivative with $u_i$ gives

$$\frac{\partial f}{\partial u_i}(x) = \frac{\partial}{\partial u_i} \sum_{j=1}^{p} u_j \sigma(a_j x + b_j) = \sigma(a_i x + b_i),$$

therefore $\nabla_u f(x) = (\sigma(a_1 x + b_1), \cdots, \sigma(a_p x + b_p)) = \sigma(ax + b)$.

(ii) Again, partial derivative with $b_i$ gives

$$\frac{\partial f}{\partial b_i}(x) = \frac{\partial}{\partial b_i} \sum_{j=1}^{p} u_j \sigma(a_j x + b_j) = u_i \sigma'(a_i x + b_i),$$

which gives $\nabla_b f(x) = (u_1 \sigma'(a_1 x + b_1), \cdots, u_p \sigma'(a_p x + b_p))$.

(iii) Use the same way, we get

$$\frac{\partial f}{\partial a_i}(x) = \frac{\partial}{\partial a_i} \sum_{j=1}^{p} u_j \sigma(a_j x + b_j) = u_i x \sigma'(a_i x + b_i),$$

which is just multiplication by $x$ on (ii). Multiplying $x$ on the result of (ii) gives result.