MathDNN HW 10

2018-13260 차재현

1 Problem 1

First,

$$\begin{aligned}
&\nabla_\phi \mathbb{E}_{Z\sim q_\phi(Z)}\left[\log\left(\frac{h(Z)}{q_\phi(Z)}\right)\right] \\
&= \nabla_\phi \int_z q_\phi(z)\log\left(\frac{h(z)}{q_\phi(z)}\right)dz \\
&= \int_z \nabla_\phi\left(q_\phi(z)\log\left(\frac{h(z)}{q_\phi(z)}\right)\right)dz \\
&= \int_z \nabla_\phi(q_\phi(z))\log\left(\frac{h(z)}{q_\phi(z)}\right)dz + \int_z q_\phi(z)\nabla_\phi\log\left(\frac{h(z)}{q_\phi(z)}\right)dz.
\end{aligned}$$

Take a closer look on the first term of the last expression. We have

$$\begin{aligned}
\int_z \nabla_\phi(q_\phi(z))\log\left(\frac{h(z)}{q_\phi(z)}\right)dz &= \int_z \frac{\nabla_\phi(q_\phi(z))}{q_\phi(z)}\log\left(\frac{h(z)}{q_\phi(z)}\right)q_\phi(z)dz \\
&= \mathbb{E}_{Z\sim q_\phi(Z)}\left[\nabla_\phi \log q_\phi(z)\log\left(\frac{h(z)}{q_\phi(z)}\right)\right]
\end{aligned}$$

Next we have

$$\int_z q_\phi(z)\nabla_\phi\log\left(\frac{h(z)}{q_\phi(z)}\right)dz = \int_z q_\phi(z)\frac{-q'_\phi(z)}{q_\phi(z)}dz = -\int_z q'_\phi(z)dz. \tag{1}$$

Since $q_\phi(z)$ is a nonnegative function defined on $\mathbb{R}$ and integrable, for any $\epsilon > 0$ there is compact set U containing 0 satisfying $q_\phi(z) < \epsilon$ for all $z \in \mathbb{R} - U$. Therefore (1) is less than 2ϵ, this means (1) is equal to zero.

2 Problem 2

By the definition of C, we can write $x = (a, t)$ for $0 \le t \le 1$. Then $\Pi_C(x, y) = (a - y_1)^2 + (t - y_2)^2$.

i) $y_2 < 0$. The term $(t - y_2)^2$ obtains its minimum at $t = 0$ with minimum y_2^2. In this case, $\min\{\max\{y_2, 0\}, 1\} = 0$.

ii) $0 \le y_2 \le 1$. The term $(t-y_2)^2$ obtains its minimum at $t = y_2$ with minimum 0. In this case, $\min\{\max\{y_2, 0\}, 1\} = y_2$.

iii) $1 \le y_2$. The term $(t-y_2)^2$ obtains its minimum at $t = 1$ with minimum $(y_2-1)^2$. In this case, $\min\{\max\{y_2, 0\}, 1\} = 1$.

Therefore $t = \min\{\max\{y_2, 0\}, 1\}$ for any y_2, equivalently, for any y.

3 Problem 4

(a) Since f_1 is linear, $\partial f_1(x)/\partial x = A$. So we get

$$\log \left| \frac{\partial f_1(x)}{\partial x} \right| = \log \left(|\det(P)| |\det(L)| |\det(U + \operatorname{diag}(s))| \right).$$

Now, since P is permutation, $|\det(P)| = 1$. Also, that L is lower triangular with diagonal entries 1 gives $\det(L) = 1$. Finally, $U + \operatorname{diag}(s)$ is upper triangular with diagonal entries are $s_1, \cdots, s_C$. Therefore $|\det(U + \operatorname{diag}(s))| = \prod_i |s_i|$. Summing up, we conclude

$$\log \left(|\det(P)| |\det(L)| |\det(U + \operatorname{diag}(s))| \right) = \sum_{i \le C} \log |s_i|.$$

(b) Let X has two distinct reshape format X_1, X_2. Similarly, $h(X)$ has two distinct reshape format Y_1, Y_2. Since both X_1, X_2 (resp. Y_1, Y_2) is just an array of entries in X (resp. $h(X)$), there is a permutation matrix P_X (resp. P_Y) satisfying $X_2 = P_X X_1$ (resp. $Y_2 = P_Y Y_1$). We have $Y_2 = P_Y h(P_X^{-1} X_2)$. Therefore

$$\frac{\partial Y_2}{\partial X_2} = \frac{\partial Y_2}{\partial Y_1} \frac{\partial Y_1}{\partial X_1} \frac{\partial X_1}{\partial X_2} = P_Y \frac{\partial Y_1}{\partial X_1} P_X^{-1}.$$

Now, use $|\det(P_X)| = |\det(P_Y)| = 1$, we get

$$\left| \frac{\partial Y_2}{\partial X_2} \right| = \left| P_Y \frac{\partial Y_1}{\partial X_1} P_X^{-1} \right| = \left| \frac{\partial Y_1}{\partial X_1} \right|,$$

which means that $|\partial h(X)/\partial X|$ is independent of the order of reshape.

(c) From (b), we can reshape X and $f_2(X)$ so that the linear map (since convolutional with no bias is linear) f_2 is equivalent to the block diagonal matrix

$$\begin{pmatrix} A & \cdots & \mathbf{0} \\ \vdots & & \vdots \\ \mathbf{0} & \cdots & A \end{pmatrix},$$

where A appears mn times. To be more specific, this can be done by reshaping $X'_{c+C*\mu+m*\eta} = X_{c,\mu,\eta}$ and same for $f_2(X)$. Then, (a), (b) and elementary linear algebra regarding to block diagonal matrix gives

$$\log\left|\frac{\partial f_2(X|P,L,U,s)}{\partial x}\right| = \log|A^{mn}| = mn\log|A| = mn\sum_i \log|s_i|.$$

(d) Using block matrix again, we have

$$\frac{\partial Z}{\partial X} = \begin{pmatrix} \frac{\partial X_{1:C}}{\partial X_{1:C}} & \mathbf{0} \\ \frac{\partial f_2(X_{C+1:2C}|P,L,U,s)}{\partial X_{1:C}} & \frac{\partial f_2(X_{C+1:2C}|P,L,U,s)}{\partial X_{C+1:2C}} \end{pmatrix},$$

and its determinant becomes

$$\det(\frac{\partial Z}{\partial X}) = \det\left(\frac{\partial X_{1:C}}{\partial X_{1:C}}\right)\det\left(\frac{\partial f_2(X_{C+1:2C}|P,L,U,s)}{\partial X_{C+1:2C}}\right).$$

First part becomes 1, and second part is already calculated at (c). Therefore, taking absolute value and log, we get

$$\log\left|\frac{\partial Z}{\partial X}\right| = mn\sum_i \log|s_i|.$$