MFDNN

Homework 10

24/05/23

1. $$\nabla_\phi \mathbb{E}_{z\sim q_\phi(z)}\left[\log\left(\frac{h(z)}{q_\phi(z)}\right)\right] = \nabla_\phi \int_{\mathbb{R}^k} \log\left(\frac{h(z)}{q_\phi(z)}\right) q_\phi(z)\,dz$$

$$= \int_{\mathbb{R}^k} \left(q_\phi(z)\, \nabla_\phi \log\left(\frac{h(z)}{q_\phi(z)}\right) + \log\left(\frac{h(z)}{q_\phi(z)}\right)\nabla_\phi q_\phi(z)\right) dz$$

$$= \int_{\mathbb{R}^k} \left(q_\phi(z)\,\underbrace{\nabla_\phi \log h(z)}_{=0} - q_\phi(z)\nabla_\phi \log q_\phi(z) + \log\left(\frac{h(z)}{q_\phi(z)}\right)\nabla_\phi q_\phi(z)\right) dz$$

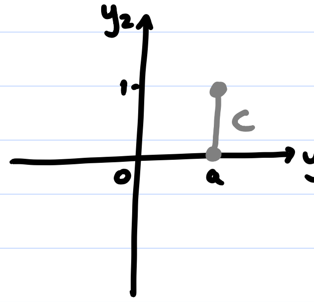
$$= \int_{\mathbb{R}^k} \left(0 - q_\phi(z)\left(\nabla_\phi q_\phi(z)\right)\frac{1}{q_\phi(z)} + \underbrace{\left(\nabla_\phi q_\phi(z)\right)}\log\left(\frac{h(z)}{q_\phi(z)}\right)\right) dz$$

$$= \int_{\mathbb{R}^k} \left(\underbrace{-\nabla_\phi q_\phi(z)}_{=0 \text{ by hint}} + \overbrace{\nabla_\phi \log q_\phi(z)\cdot q_\phi(z)}\log\left(\frac{h(z)}{q_\phi(z)}\right)\right) dz$$

$$= \int_{\mathbb{R}^k} \left(\nabla_\phi \log q_\phi(z)\right) \log\left(\frac{h(z)}{q_\phi(z)}\right) q_\phi(z)\,dz$$

$$= \mathbb{E}_{z\sim q_\phi(z)}\left[\left(\nabla_\phi \log q_\phi(z)\right)\log\left(\frac{h(z)}{q_\phi(z)}\right)\right] \quad \blacksquare$$

2.



$\mathbb{R}^2$ $\qquad \Pi_C(y) = (x_1, x_2)$

Three regions, $Y_1 = \{y \in \mathbb{R}^2 \mid y_2 \geq 1\}$ closest pt is $(a,1) \in C$

$Y_2 = \{y \in \mathbb{R}^2 \mid 0 < y_2 < 1\}$ closest pt is $(a, y_2) \in C$

$Y_3 = \{y \in \mathbb{R}^2 \mid y_2 \leq 0\}$ closest pt is $(a, 0) \in C$

$(Y_1 \cup Y_2 \cup Y_3 = \mathbb{R}^2)$

$$\|x - y\|^2 = \left\|\begin{pmatrix} a \\ x_2\end{pmatrix} - \begin{pmatrix} y_1 \\ y_2\end{pmatrix}\right\|^2 = \underbrace{(a-y_1)^2}_{\text{can't be minimized over } x} + \underbrace{(x_2 - y_2)^2}_{\text{can be minimized}}$$

If $y \in Y_1$, $y_2 \geq 1$ so minimized when $x_2 = 1$

and $\min\{\max\{y_2, 0\}, 1\} = \min\{y_2, 1\} = 1$ ✓

If $y \in Y_2$, $0 < y_2 < 1$ so minimized (to 0) if $x_2 = y_2$

and $\min\{\max\{y_2, 0\}, 1\} = \min\{y_2, 1\} = y_2$ ✓

If $y \in Y_3$, $y_2 \leq 0$ so minimized when $x_2 = 0$

and $\min\{\max\{y_2, 0\}, 1\} = \min\{0, 1\} = 0$ ✓ So $x_2 = \min\{\max\{y_2, 0\}, 1\}$ & $x_1 = a$ $\blacksquare$

4. (a) $\left|\frac{\partial f_1}{\partial \underline{z}}\right| = \left|\det \frac{\partial f_1}{\partial \underline{z}}\right| = \left|\det\left(\frac{\partial}{\partial \underline{z}}(A\underline{z})\right)\right| = |\det A| = |\det(PL(U + \operatorname{diag}(\underline{s})))|$

$$= |\det P \det L \det(U + \operatorname{diag}\underline{s})| = |\det P|\,|\det L|\,|\det(U + \operatorname{diag}(\underline{s}))|$$

$$= 1 \cdot 1 \cdot \left|\prod_{i=1}^{c} s_i\right|$$

since $|\det P| = 1$ by HW8.7

L (lower) triangular $\Rightarrow$ det is diagonal product

U has 0 diagonals. Add s_i. Still diagonal

$\Rightarrow$ det is diagonal product

$$\log\left|\frac{\partial f_1}{\partial \underline{z}}\right| = \log|A| = \log\left|\prod_{i=1}^{c} s_i\right| = \sum_{i=1}^{c} \log|s_i| \quad \blacksquare$$

(b) Reshape operation definitions may differ only by a permutation matrix. That is to say, any reshape operation (call it reshape*) output vector in $\mathbb{R}^{abc}$ is just a permutation of a desired 'true' reshape using a particular P_* unique to the original reshape used. i.e. $X.\text{reshape}^*(abc) = P_*(X.\text{reshape}(abc))$. NB: $*$ is a permutation of length abc

$$\text{So, } \left|\frac{\partial(h(X).\text{reshape}^*(abc))}{\partial(X.\text{reshape}^*(abc))}\right| = \left|\frac{\partial(P_*(\overbrace{h(X).\text{reshape}(abc)}^{h'(\underline{X})}))}{\partial(P_*(\underbrace{X.\text{reshape}(abc)}_{=:\underline{X}'}))}\right|$$

'true' reshape (arbitrary)

any choice of reshape

$$\left(\frac{\partial(A\underline{y})}{\partial(\underline{z})}\right)_{ij} = \frac{\partial(A\underline{y})_i}{\partial z_j} = \frac{\partial}{\partial z_j}\left(\sum_k A_{ik}\, y_k(\underline{z})\right)$$

$$= \sum_k A_{ik}\left(\frac{\partial}{\partial z_j} y_k(\underline{z})\right)$$

$$= \left(A\,\frac{\partial \underline{y}(\underline{z})}{\partial \underline{z}}\right)_{ij}$$

$$\frac{\partial(\underline{y})}{\partial(A\underline{y})}\frac{\partial(A\underline{y})}{\partial(\underline{y})} = I \text{ (chain rule)}$$

$$\frac{\partial(\underline{y})}{\partial(A\underline{y})} = \left(\frac{\partial(A\underline{y})}{\partial(\underline{y})}\right)^{-1}$$

$$= \left|\frac{\partial(P_* h'(\underline{X}))}{\partial(P_* \underline{X}')}\right| = \left|\frac{\partial(P_* h'(\underline{X}))}{\partial(\underline{X}')}\,\frac{\partial(\underline{X}')}{\partial(P_* \underline{X}')}\right|$$

notation clarification

$$= \left|P_*\,\frac{\partial(h'(\underline{X}))}{\partial(\underline{X}')}\,P_*^{-1}\right|$$

by HW 8.7

$$= \left|\det P_* \det \frac{\partial h'}{\partial \underline{x}'} \det P_*^{-1}\right|$$

$$= \left|\det \frac{\partial h'}{\partial \underline{x}'}\right| = \left|\frac{\partial(h(X).\text{reshape}(abc))}{\partial(X.\text{reshape}(abc))}\right|$$

notation

i.e. starting with any reshape operation, reshape*, result depends only on (arbitrary) 'true' reshape so result is always the same & choice of reshape doesn't matter $\blacksquare$

4.(c) $$\left(f_2(X|P,L,U,s)\right)_{\tilde{c},\tilde{i},\tilde{j}} = \sum_{\gamma=1}^{C}\sum_{\alpha=1}^{1}\sum_{\beta=1}^{1} \omega_{\tilde{c},\gamma,1,1}\, X_{\gamma,\tilde{i}+\alpha-1,\tilde{j}+\beta-1}$$

$1 \le \tilde{c} \le C$

$$= \sum_{\gamma=1}^{C} A_{\tilde{c},\gamma}\, X_{\gamma,\tilde{i},\tilde{j}}$$

$$\Rightarrow \underbrace{\underline{vec}(f_2)}_{\in\mathbb{R}^{Cmn}} = (\underbrace{I_{mn}}_{\text{identity } mn\times mn} \underbrace{\otimes}_{\text{Kronecker product}} A)\,\underline{vec}(X)^{\in\mathbb{R}^{Cmn\times 1}}$$

i.e. A applied to each channel of X at each $\tilde{i},\tilde{j}$ position. See vec(...) def. below

$$\begin{pmatrix} (f_2)_{1,1,1} \\ (f_2)_{2,1,1} \\ \vdots \\ (f_2)_{C,1,1} \\ (f_2)_{1,1,2} \\ \vdots \end{pmatrix} = mn \text{ As}\left\{ \begin{pmatrix} \overset{C\times C}{A} & 0 & \cdots & 0 \\ 0 & A & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \underset{C\times C}{A} \end{pmatrix} \right. \cdot \begin{pmatrix} X_{1,1,1} \\ X_{2,1,1} \\ \vdots \\ X_{C,1,1} \\ X_{1,1,2} \\ \vdots \end{pmatrix} \Big\} mn \text{ times}$$

i.e. $I_{mn}\otimes A = \underbrace{\begin{bmatrix} A & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A \end{bmatrix}}_{\text{block matrix}} \in \mathbb{R}^{Cmn\times Cmn}$

Define a reshape operation, $\underline{vec}(\ldots)$, to operate by point/pixel and iterate over channels (i.e. col-major)

$$\underline{vec}(X) = X.\text{reshape}(Cmn) = (X_{1,1,1}, X_{2,1,1}, \ldots, X_{C,1,1},$$
$$X_{1,1,2}, X_{2,1,2}, \ldots, X_{C,1,2},$$
$$\vdots$$
$$X_{1,m,n}, X_{2,m,n}, \ldots, X_{C,m,n})^T$$

$$\left|\frac{\partial f_2(X|P,L,U,s)}{\partial X}\right| = \left|\frac{\partial\left((I_{mn}\otimes A)\underline{vec}(X)\right)}{\partial(\underline{vec}(X))}\right|$$ by part (b) (can use any reshape)

$$= |I_{mn}\otimes A|$$

$$= |I|^C |A|^{mn}$$ by properties of $\otimes$: $I_{mn}\in\mathbb{R}^{mn\times mn}$, $A\in\mathbb{R}^{C\times C}$

$$= 1\cdot|A|^{mn}$$ $\Rightarrow \log\left|\frac{\partial f_2}{\partial X}\right| = \log|A|^{mn} = mn\log|A|$

$= mn\sum_{i=1}^{C}\log|s_i|$ by part (a)

4.(d) $$\log\left|\frac{\partial z}{\partial X}\right| = \log\begin{vmatrix} I_C & 0 \\ * & \frac{\partial}{\partial X_{C+1:2C,:,:}}(f_2)_{C+1:2C,:,:} \end{vmatrix}$$

$z_{1:C,:,:} = X_{1:C,:,:}$ so derivative wrt $X_{1:C,:,:}$ is just identity

not relevant

$$= \log\left|\underbrace{\det(I_C)}_{=1}\det\left(\frac{\partial(f_2)_{C+1:2C,:,:}}{\partial X_{C+1:2C,:,:}}\right)\right|$$ by block lower triangular formula for det

$$= \log\left|\frac{\partial\tilde{f}_2}{\partial\tilde{X}}\right| = mn\sum_{i=1}^{C}\log|s_i|$$ by part (c) since $\tilde{f}_2, \tilde{X}\in\mathbb{R}^{C\times m\times n}$

pdf of X - Gaussian

$$6.(a)\ \nabla_{\mu,\tau}\mathbb{E}_{X\sim\mathcal{N}(\mu,e^{2\tau})}\left[X\sin(X)\right] = \mathbb{E}_{X\sim\mathcal{N}(\mu,e^{2\tau})}\left[X\sin(X)\,\nabla_{\mu,\tau}\log\left(f_{\mu,\tau}(X)\right)\right]$$

$$= \mathbb{E}_{X\sim\mathcal{N}(\mu,e^{2\tau})}\left[X\sin(X)\,\nabla_{\mu,\tau}\left(-\frac{1}{2}\left(\frac{X-\mu}{e^{\tau}}\right)^2 - \log\left(e^{\tau}\sqrt{2\pi}\right)\right)\right]$$

$\hookrightarrow -\frac{1}{2}(X-\mu)^2e^{-2\tau}$

$$= \mathbb{E}_{X\sim\mathcal{N}(\mu,e^{2\tau})}\left[X\sin(X)\left(\frac{X-\mu}{e^{2\tau}},\ \frac{(X-\mu)^2}{e^{2\tau+1}} - 1\right)\right]$$

$$\approx \frac{1}{B}\sum_{i=1}^{B}\left(\frac{X_i\sin(X_i)(X_i-\mu)}{e^{2\tau}},\ X_i\sin(X_i)\left(\frac{(X_i-\mu)^2}{e^{2\tau+1}} - 1\right)\right)$$

with $X_i\sim\mathcal{N}(\mu,e^{2\tau})$

$$(b)\ \nabla_{\mu,\tau}\mathbb{E}_{X\sim\mathcal{N}(\mu,e^{2\tau})}\left[X\sin(X)\right] = \mathbb{E}_{Y\sim\mathcal{N}(0,1)}\left[\nabla_{\mu,\tau}\left((\mu+e^{\tau}Y)\sin(\mu+e^{\tau}Y)\right)\right]$$

$$= \mathbb{E}_{Y\sim\mathcal{N}(0,1)}\left[\left(\sin(\mu+e^{\tau}Y)+(\mu+e^{\tau}Y)\cos(\mu+e^{\tau}Y),\ e^{\tau}Y\sin(\mu+e^{\tau}Y)+(\mu+e^{\tau}Y)e^{\tau}Y\cos(\mu+e^{\tau}Y)\right)\right]$$

$$\approx \frac{1}{B}\sum_{i=1}^{B}\left(\sin(\mu+e^{\tau}Y_i)+(\mu+e^{\tau}Y_i)\cos(\mu+e^{\tau}Y_i),\ e^{\tau}Y_i\sin(\mu+e^{\tau}Y_i)+(\mu+e^{\tau}Y_i)e^{\tau}Y_i\cos(\mu+e^{\tau}Y_i)\right)$$

with $Y_i\sim\mathcal{N}(0,1)$

$$\nabla_{\mu,\tau}\left(\frac{1}{2}(\mu-1)^2 + e^{\tau} - \log e^{\tau}\right) = \nabla_{\mu,\tau}\left(\frac{1}{2}(\mu-1)^2 + e^{\tau} - \tau\right)$$

$$= \left(\mu-1,\ e^{\tau}-1\right)$$