MathDNN HW 8

2018-13260 차재현

1 Problem 1

T : Linear?

Regard a domain of $\mathcal{T}$ as $(\mathbb{R}^4)^{m/2 \times n/2}$, where each $\mathbb{R}^4$ is 2×2 submatrix of domain which shrinks to $\mathbb{R}$ by $\mathcal{T}$. In this case, we can express $\mathcal{T}$ as

$$[\mathcal{T}] = \begin{pmatrix} I'_{1\times 4} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & I'_{1\times 4} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & I'_{1\times 4} \end{pmatrix},$$

where $I'_{1\times 4} = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$ and $\mathbf{0} = (0 \ 0 \ 0 \ 0)$. Since transpose is unique and it is independent of choice of basis (because so does dual map), we have matrix expression

$$[\mathcal{T}]^\top = \begin{pmatrix} {I'_{1\times 4}}^\top & \mathbf{0}^\top & \cdots & \mathbf{0}^\top \\ \mathbf{0}^\top & {I'_{1\times 4}}^\top & \cdots & \mathbf{0}^\top \\ \vdots & \vdots & & \vdots \\ \mathbf{0}^\top & \mathbf{0}^\top & \cdots & {I'_{1\times 4}}^\top \end{pmatrix}$$

under same basis. This means following. If a down-sampled image Y is given, its scale-up image X given by $\mathcal{T}^\top$ can be written as (using zero-based index)

$$X_{2i,2j} = X_{2i+1,2j} = X_{2i,2j+1} = X_{2i+1,2j+1} = \frac{Y_{i,j}}{4}.$$

Note that, also, this is a nearest-neighbor by Cauchy-Schwartz inequality

$$(a^2 + b^2 + c^2 + d^2)(4) \geq (a + b + c + d)^2 = 1,$$

where equation holds when $a = b = c = d = 1/4$.

2 Problem 3

(a) The measure $d\mu = p_Y(x)dx$ is a probability measure, so Jensen's inequality concludes

$$D_f(X||Y) = \int f\left(\frac{p_X(x)}{p_Y(x)}\right) p_Y(x)dx \geq f\left(\int \frac{p_X(x)}{p_Y(x)} p_Y(x)dx\right) \geq f(1) = 0.$$

(b) First of all, since both function is defined only on $t > 0$

$$(-\log t)'' = \frac{1}{t^2} > 0, \qquad (t \log t)'' = \frac{1}{t} > 0,$$

so both are convex. Also, $f(1) = 0$ for both cases.

Take $f(t) = -\log t$. In this case,

$$D_f(X||Y) = \int p_Y(x) \log p_Y(x) - p_Y(x) \log p_X(x) dx, \tag{1}$$

which is a generalization of

$$D_{\mathrm{KL}}(Y||X) = \sum_x p_Y(x) \log p_Y(x) - p_Y(x) \log p_X(x). \tag{2}$$

Explicitly, (2) is just (1) with discrete measure.

Next, take $f(t) = t \log t$. In this case,

$$D_f(X||Y) = \int p_X(x) \log \frac{p_X(x)}{p_Y(x)} = \int p_X(x) \log p_X(x) - p_X(x) \log p_Y(x) dx,$$

which is a generalization of $D_{\mathrm{KL}}(X||Y)$, by the same argument.

3 Problem 4

Being CDF, F is nondecreasing.

By the definition of G

$$\{u \in [0,1] \mid G(u) \leq t\} = \{u \in [0,1] \mid \inf_x \{u \leq F(x)\} \leq t\}.$$

Since F is nondecreasing, $x_1 \leq x_2$ is equivalent to $F(x_1) \leq F(x_2)$. Thus we can restrict inside of inf in right side to $u = F(x)$. Then

$$\{u \in [0,1] \mid G(u) \leq t\} = \{u \in [0,1] \mid \inf_x \{u = F(x)\} \leq t\}.$$

Let $x_0 = \inf_x \{u = F(x)\}$. Such x_0 exists because $\lim_{x \to -\infty} = 0$ and well-ordered property of $\mathbb{R}$. If x_0 is the unique inverse image of F at u, trivially $F(x_0) = u$.

Suppose otherwise. Since F is nondecreasing, $F(x) \geq F(x_0)$ for all $x \geq x_0$. Then there is $x_1 > x_0$ satisfying $F(x_1) = u$. Thus $F(x_0) \leq u$. By the definition of x_0, no $x \in (x_0, x_1)$ has its function value $F(x)$ strictly less than u. Then, right continuity of F gives

$$F(x_0) = \lim_{x \to x_0+} F(x) = u.$$

Thus we get

$$\{u \in [0,1] \mid G(u) \leq t\} = \{u \in [0,1] \mid u = F(x), x \leq t\}.$$

Applying nondecreasing property once again, this reduces to

$$\{u \in [0,1] \mid G(u) \leq t\} = \{u \in [0,1] \mid u \leq F(t)\}.$$

Taking union over all $u \in U$, therefore, we conclude

$$\mathbb{P}(G(U) \leq t) = \mathbb{P}(U \leq F(t)) = F(t).$$

4 Problem 5

Since A is invertible, we have $Y = A^{-1}(X - b)$. Take $\phi(x) = A^{-1}(x - b)$. For $y = \phi(x)$, we have

$$p_X(x) = p_Y(y) \left| \det \frac{\partial \phi}{\partial x}(x) \right| = \frac{1}{\sqrt{(2\pi)^n}} e^{-\frac{\|y\|^2}{2}} \left| \det \frac{\partial \phi}{\partial x}(x) \right| \tag{3}$$

We make some modifications on previous expressions. First, since ϕ is a combination of transpose and linear, its derivative is equal to the derivative of linear part. That is,

$$\left| \det \frac{\partial \phi}{\partial x}(x) \right| = |\det A^{-1}| = |(\det A)|^{-1}.$$

Also, since $\det A = \det A^\top$, we can rewrite as

$$\left| \det \frac{\partial \phi}{\partial x}(x) \right| = \sqrt{\det A \det A^\top}^{-1} = \sqrt{\det \Sigma}^{-1}.$$

Next, definition of norm in vector space gives

$$\begin{aligned} \|y\|^2 = \langle y, y \rangle = y^\top y &= (A^{-1}(x - b))^\top A^{-1}(x - b) \\ &= (x - b)^\top (AA^\top)^{-1}(x - b) \end{aligned}$$

$$= (x-b)^\top \Sigma^{-1}(x-b)$$

Substituting both results to (3) gives

$$p_X(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-b)^\top \Sigma^{-1}(x-b)}.$$

5 Problem 6

For me, it is more convenient to use python code instead of pseudocode. I append the answer.

6 Problem 7

(a) Given x, the i-th element of $P_\sigma x$ is $e_{\sigma(i)}^\top x = x_{\sigma(i)}$.

(b) Since each $e_{\sigma(i)}$ is a unit vector, inner product of rows $\left\langle e_{\sigma(i)}, e_{\sigma(j)} \right\rangle = \delta_{ij}$, where δ_{ij} is usual dirac symbol. Thus rows of P_σ consists orthonormal system, so P_σ is unitary. Thus we get $P_\sigma^\top = (P_\sigma)^{-1}$. Next, take $x = e_j$ in (a) over $j = 1, \cdots, n$, we get $P_\sigma x_j = x_{\sigma(j)}$. This is equivalent to $(P_\sigma)^{-1} x_{\sigma(j)} = x_j$, and since σ is one-to-one, $(P_\sigma)^{-1} x_j = x_{\sigma^{-1}(j)}$. We conclude $(P_\sigma)^{-1} = P_{\sigma^{-1}}$.

(c) From (b) and the identity $\det A = \det A^\top$, we have

$$1 = \det I = \det(P_\sigma {P_\sigma}^{-1}) = \det(P_\sigma P_\sigma^\top) = (\det P_\sigma)^2.$$