MathDNN Homework 8

Department of Computer Science and Engineering
2021-16988 Jaewan Park

Problem 1

Since $\mathcal{T}$ is a 2×2 average pool operator with stride 2, A will be given as

$$A = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & 0 & \cdots & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} & 0 & \cdots & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & 0 & \cdots & 0 & 0 & \cdots \\ \vdots & & & & \ddots & \ddots & & & & & & & \ddots & \ddots & & & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & \cdots & 0 & \frac{1}{4} & \frac{1}{4} & \cdots \\ & & & & & & & & & & & & & & & & \\ & & & & & & & & & & & & & & & & \ddots \end{bmatrix}.$$

and

$$[\mathcal{T}(X)]_{i,j} = \frac{1}{4}\Big([X]_{2i-1,2j-1} + [X]_{2i-1,2j} + [X]_{2i,2j-1} + [X]_{2i,2j}\Big).$$

So by the definition of $\mathcal{T}^\top$, we can calculate

$$\begin{aligned} \sum_{i=1}^{m/2}\sum_{j=1}^{n/2}[Y]_{i,j}[\mathcal{T}(X)]_{i,j} &= \sum_{i=1}^{m/2}\sum_{j=1}^{n/2}\frac{1}{4}[Y]_{i,j}\Big([X]_{2i-1,2j-1} + [X]_{2i-1,2j} + [X]_{2i,2j-1} + [X]_{2i,2j}\Big) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{1}{4}[Y]_{\lceil i/2\rceil,\lceil j/2\rceil}[X]_{i,j} = \sum_{i=1}^{m}\sum_{j=1}^{n}\left[\mathcal{T}^\top(Y)\right]_{i,j}[X]_{i,j}. \end{aligned}$$

Therefore we can compute $\mathcal{T}^\top$ by calculating $\left[\mathcal{T}^\top(Y)\right]_{i,j} = \dfrac{1}{4}[Y]_{\lceil i/2\rceil,\lceil j/2\rceil}$. This is equivalent to $\dfrac{1}{4}$ times the nearest neighbor upsampling.

Problem 2

```
# Using Nearest Neighbor Upsampling
layer = nn.Upsample(scale_factor=r, mode='nearest')

# Using Transpose Convolution
layer = nn.ConvTranspose2d(in_channels, out_channels, kernel_size=r, stride=r)
layer.weight.data = torch.ones(layer.weight.data.shape)
```

The two implementations above are equivalent. Transpose convolution with same kernel size and stride can be understood as nearest neighbor upsampling where all elements of the weight tensor are 1.

Problem 3

(a) Since f is a convex function, we can apply Jensen's inequality to f, which gives

$$\begin{aligned} D_f(X\|Y) &= \int f\left(\frac{p_X(x)}{p_Y(x)}\right)p_Y(x)dx = \mathbf{E}\left[f\left(\frac{p_X(Y)}{p_Y(Y)}\right)\right] \\ &\geq f\left(\mathbf{E}\left[\frac{p_X(Y)}{p_Y(Y)}\right]\right) = f\left(\int \frac{p_X(x)}{p_Y(x)}p_Y(x)dx\right) = f(1) = 0. \end{aligned}$$

(b) If $f(t) = -\log t$,

$$D_f(X\|Y) = \int -\log\left(\frac{p_X(x)}{p_Y(x)}\right)p_Y(x)dx = \int \log\left(\frac{p_Y(x)}{p_X(x)}\right)p_Y(x)dx = D_{\mathrm{KL}}(Y\|X).$$

If $f(t) = t\log t$,

$$D_f(X\|Y) = \int \left(\frac{p_X(x)}{p_Y(x)}\right)\log\left(\frac{p_X(x)}{p_Y(x)}\right)p_Y(x)dx = \int \log\left(\frac{p_Y(x)}{p_X(x)}\right)p_X(x)dx = D_{\mathrm{KL}}(X\|Y).$$

Problem 4

We should show $G(u) \leq x \Leftrightarrow u \leq F(x)$ for all $u \in (0,1)$ and $x \in \mathbb{R}$.

First, if $G(u) \leq x$, we obtain $F(G(u)) \leq F(x)$ since F is nondecreasing. Also, since F is right continuous and $\lim_{x\to-\infty} F(x) = 0$, $G(u)$ exists such that $u \leq F(G(u))$. Therefore $u \leq F(G(u)) \leq F(x)$, so $u \leq F(x)$.

Next, if $u \leq F(x)$, we obtain $G(u) \leq x$ from the definition of $G(u)$.

Therefore we obtain $G(u) \leq x \Leftrightarrow u \leq F(x)$, so

$$\Pr(G(U) \leq t) = \Pr(U \leq F(t)) = F(t).$$

Problem 5

From the relation $Y = \varphi(X) = A^{-1}(X - b)$, we obtain the following.

$$\begin{aligned} p_X(x) &= p_Y\left(A^{-1}(x-b)\right)\left|\det \frac{\partial A^{-1}(x-b)}{\partial x}(x)\right| \\ &= \frac{1}{\sqrt{(2\pi)^n}}e^{-\frac{1}{2}\left\|A^{-1}(x-b)\right\|^2}\left|\det A^{-1}\right| = \frac{1}{\sqrt{(2\pi)^n}}e^{-\frac{1}{2}\left\|A^{-1}(x-b)\right\|^2}\left|\det A\right|^{-1} \\ &= \frac{1}{\sqrt{(2\pi)^n}}e^{-\frac{1}{2}\left(A^{-1}(x-b)\right)^\intercal\left(A^{-1}(x-b)\right)}\frac{1}{\sqrt{\det AA^\intercal}} \\ &= \frac{1}{\sqrt{(2\pi)^n \det AA^\intercal}}e^{-\frac{1}{2}(x-b)^\intercal A^{-1\intercal}A^{-1}(x-b)} = \frac{1}{\sqrt{(2\pi)^n \det AA^\intercal}}e^{-\frac{1}{2}(x-b)^\intercal A^{\intercal -1}A^{-1}(x-b)} \\ &= \frac{1}{\sqrt{(2\pi)^n \det AA^\intercal}}e^{-\frac{1}{2}(x-b)^\intercal (AA^\intercal)^{-1}(x-b)} \end{aligned}$$

Problem 6

All indices in the pseudocode start from 1.

Algorithm 1 Inverse Permutation

```
procedure InversePermutation(σ)
    σ′ = [ ]                                  ▷ Empty List
    while i = 1, 2, ··· , n do
        while j = 1, 2, ··· , n do
            if σ(j) = i then
                σ′(i) = j
                break
            end if
        end while
    end while
    return σ′
end procedure
```

Problem 7

(a) For any $x \in \mathbb{R}^n$,

$$(P_\sigma x)_i = e_{\sigma(i)}^\intercal x = x_{\sigma(i)}.$$

(b) Since the rows of P_σ are standard unit vectors, they are orthonormal and P_σ is an orthogonal matrix. Therefore $P_\sigma P_\sigma^\intercal = P_\sigma^\intercal P_\sigma = I$, and $P_\sigma^\intercal = P_\sigma^{-1}$. Also, for all $i = 1, \cdots, n$, $[P_\sigma]_{i,\sigma(i)} = 1$ and all other elements are 0. Then for $P_\sigma^\intercal$, we can say $[P_\sigma^\intercal]_{\sigma(i),i} = 1$ for all $i = 1, \cdots, n$, which is equivalent to stating $[P_\sigma^\intercal]_{j,\sigma^{-1}(j)} = 1$ for all $i = 1, \cdots, n$. This gives $P_\sigma^\intercal = P_{\sigma^{-1}}$. Therefore $P_\sigma^\intercal = P_\sigma^{-1} = P_{\sigma^{-1}}$

(c) $\det P_\sigma = (-1)^t \det I = (-1)^t$, where t is the number of row changes. Therefore $|\det P_\sigma| = 1$.