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

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

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

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

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

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

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

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

Problem 4

Since $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$, for all $t \in (0,1)$, the intermediate value theorem gives there exists at least one t such that $F(t) = u$. Let $\min\{t \in \mathbb{R} \mid F(t) = y\} = t_u$, then t_u always exists. Since $t_u \in \{x \in \mathbb{R} \mid u \leq F(x)\}$, from the definition of G, we can say

$$G(u) \leq t_u.$$

Also, since F is increasing and $F(t_u) = u \leq F(G(u))$, we obtain

$$t_u \leq G(u).$$

Totally, $G(u) = t_u$, so $F(G(u)) = F(t_u) = u$. Therefore

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

Problem 5

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

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

Problem 6

All indices is the pseudocode start from 1.

Algorithm 1 Inverse Permutation

```
σ' = [ ]                                         ▷ Empty List
procedure InversePermutation(σ)                  ▷ Sort A[p, ··· , r]
    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
```