MFDNN HW8

MinGyu Shin

May 1, 2024

Problem 1 : *Transpose of downsampling.*

Since average operation is linear, the downsampling operator is also linear. Therefore, $\mathcal{T}$ has a matrix representation $A \in \mathbb{R}^{(mn/4)\times(mn)}$ such that

$$\mathcal{T}(X) = (A(X.\text{reshape}(mn))).\text{reshape}(m/2, n/2).$$

Let $a_{ij} \in \mathbb{R}^{m\times n}$ such that only components $(2i-1 : 2i, 2j-1 : 2j)$ where the average operation is applied are $1/4$, or 0, otherwise. Then

$$A = \begin{bmatrix} {a'_{11}}^\intercal \\ {a'_{12}}^\intercal \\ \vdots \\ {a'_{1n/2}}^\intercal \\ {a'_{21}}^\intercal \\ \vdots \\ {a'_{m/2n/2}}^\intercal \end{bmatrix}$$

where $a' = a.\text{reshape}(mn)$.

As a' indicates that how the each component of X affect to a single $\mathcal{T}(X)$, the column of A indicates that how the single component of X affects to each component of $\mathcal{T}(X)$. Since there is no overlap for downsampling, the only single component of the column is nonzero, especially $1/4$. Moreover, there are exactly same four column. Let $Y \in \mathbb{R}^{(m/2)\times(n/2)}$. For $A^\intercal(Y.\text{reshape}(mn/4))$, ij component of Y affects to four components of result, whose indices are $(2i-1 : 2i, 2j-1 : 2j)$. Thus the result is an instance of $1/4$ the nearest neighbor upsampling.

Problem 2 : *Nearest neighbor upsampling.*

The code below represents the equivalent layer to nearest neighbor upsampling with scale factor r.

```
layer2 = nn.ConvTranspose2d(in_channels=1, out_channels=1, kernel_size=r, stride=r, bias=False)
layer2.weight.data = torch.ones_like(layer2.weight.data)
```

Figure 1: Nearest neighbor upsampling by ConvTranspose2d.

Problem 3 : *f-divergence.*

(a)

$$D_f(X\|Y) = \mathop{\mathbb{E}}_{Y}[f(\frac{p_X}{p_Y})] \geq f(\mathop{\mathbb{E}}_{Y}[\frac{p_X}{p_Y}]) = f\left(\int p_X(t)dt\right) = f(1) = 0,$$

where we used the Jensen's inequality.

(b) Note that the KL divergence $D_{KL}(X\|Y) = \int \log\left(\frac{p_X(t)}{p_Y(t)}\right) p_X(t)dt$.

$$D_f(X\|Y) = \int -\log(\frac{p_X(t)}{p_Y(t)})p_Y(t)dt = \int \log(\frac{p_Y(t)}{p_X(t)})p_Y(t)dt = D_{KL}(Y\|X),$$

with $f = \log t$,

$$D_f(X\|Y) = \int \frac{p_X(t)}{p_Y(t)} \log(\frac{p_X(t)}{p_Y(t)})p_Y(t)dt = \int \log(\frac{p_Y(t)}{p_X(t)})p_X(t)dt = D_{KL}(X\|Y),$$

with $f = t\log t$.

Problem 4 : *Generalized inverse tranform sampling.*

First, note that for $u \in (0,1), \exists x \in \mathbb{R}$ such that $x = G(u)$ since F is nondecreasing and right-continuous. Additionally,

Claim : $u \leq F(x) \Leftrightarrow G(u) \leq x$ for $u \in (0,1), x \in \mathbb{R}$

Proof : $(\Rightarrow)$ $G(u) = \inf\{t \in \mathbb{R} | u \leq F(t)\} \leq x$.

$(\Leftarrow)$ $u \leq F(G(u)) \leq F(x)$ since F is nondecresing.

Thus, $P(G(U) \leq t) = P(U \leq F(t)) = F(t)$.

Problem 5 : *Change of variables formula for Gaussians.*

Let $\varphi(P) = A^{-1}(P-b)$. Then for given X, $\varphi(X) = A^{-1}(AY+b-b) = Y$. Since A is invertible, φ is one-to-one. Moreover, φ is differentiable with derivatives $\frac{\partial\varphi}{\partial X} = A^{-1}$. Note that

$$(A^{-1})^\intercal A^{-1} = (A^\intercal)^{-1}A^{-1} = (AA^\intercal)^{-1} = \Sigma^{-1},$$

$$\det\Sigma = \det(AA^\intercal) = \det A \det A^\intercal = (\det A)^2 \Leftrightarrow \det A = \sqrt{\det\Sigma}.$$

Thus, X is a continuous random vector with density

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

Problem 6 : *Inverse permutation.*

The pseudocode below describes the algorithm for computing σ^{-1} given σ.

1. Construct a $n \times n$ matrix Σ whose i-th column is $e_{\sigma(i)}$.

2. Compute $\Sigma[1\ 2\ \ldots n]^\intercal$.

In 1, $e_i \in \mathbb{R}^n$ is standard unit vector.

Problem 7 : *Permutation matrix.*

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

(b) Since $(P_\sigma P_\sigma^\intercal)_{ij} = e_{\sigma(i)}^\intercal e_{\sigma(j)} = \begin{cases} 1 & i = j \\ 0 & o.w. \end{cases}$, $P_\sigma P_\sigma^\intercal = I_n$, that is, $P_\sigma^\intercal = P_\sigma^{-1}$.

Since $\sigma^{-1}(\sigma(i)) = i$, for $x \in \mathbb{R}^n, P_{\sigma^{-1}} P_\sigma x = x$, that is, $P_\sigma^{-1} = P_{\sigma^{-1}}$

(c) $1 = |\det I_n| = |\det P_\sigma^\intercal P_\sigma| = |\det P_\sigma^\intercal||\det P_\sigma| = |\det P_\sigma|^2$

$\Rightarrow |\det P_\sigma| = 1.$

More generally, for any orthogonal matrix $P, |\det P| = 1$.