2020-14281 이시웅

#1. $\sum_{i=1}^{m/2}\sum_{j=1}^{n/2} Y_{ij}(T(X))_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n}(T^T(Y))_{ij} X_{ij}$ for all $X \in \mathbb{R}^{m\times n}$, $Y \in \mathbb{R}^{m/2 \times n/2}$

$$= \sum_{i=1}^{m/2}\sum_{j=1}^{n/2} Y_{ij}\left(\sum_{k=2i-1}^{2i}\sum_{l=2j-1}^{2j} \frac{1}{4} X_{kl}\right) = \sum_{i=1}^{m/2}\sum_{j=1}^{n/2}\sum_{k=2i-1}^{2i}\sum_{l=2j-1}^{2j} \frac{1}{4} Y_{ij} X_{kl}$$

$$= \sum_{k=1}^{m}\sum_{l=1}^{n}\sum_{i=1}^{m/2}\sum_{j=1}^{n/2} \frac{1}{4} Y_{ij} X_{kl} \mathbb{1}_{\{\lceil k/2 \rceil = i\}} \mathbb{1}_{\{\lceil l/2 \rceil = j\}}$$

$$= \sum_{k=1}^{m}\sum_{l=1}^{n} \frac{1}{4} Y_{\lceil k/2 \rceil \lceil l/2 \rceil} X_{kl}$$

$\therefore$ Thus $(T^T(Y))_{ij} = \frac{1}{4} Y_{\lceil i/2 \rceil \lceil j/2 \rceil}$

i.e. T^T is $\frac{1}{4}$ times the nearest neighbor upsampling (scale = 2)

#2. layer = nn.Upsample (scale_factor = r, mode = 'nearest')

$\Leftrightarrow$

```
                               Cin Cout
                               //  //
layer = nn.ConvTranspose2d ( c, c, kernel_size=r, stride=r, bias=False)
layer.weight.data = torch.zeros (c, c, r, r)
for i in range(c):
    layer.weight.data[i, i, :, :] = 1
```

#3 f-divergence

By Jensen's inequality, since f is convex.

(a) $D_f(X\|Y) = \int f\left(\frac{P_X(x)}{P_Y(x)}\right) P_Y(x)\,dx = \mathbb{E}_Y\left[f\left(\frac{P_X}{P_Y}\right)\right] \geq f\left(\mathbb{E}_Y\left[\frac{P_X}{P_Y}\right]\right) = f(1) = 0$.

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

$$D_f(X\|Y) = \int (\log P_Y(x) - \log P_X(x)) P_Y(x)\,dx = \int P_Y(x)\log P_Y(x)\,dx - \int P_Y(x)\log P_X(x)\,dx = D_{KL}(Y\|X)$$

when $f = t\log t$

$$D_f(X\|Y) = \int \left(\frac{P_X(x)}{P_Y(x)}\log P_X(x) - \frac{P_X(x)}{P_Y(x)}\log P_Y(x)\right) P_Y(x)\,dx = \int P_X(x)\log P_X(x)\,dx - \int P_X(x)\log P_Y(x)\,dx = D_{KL}(X\|Y).$$

#4. $F: \mathbb{R} \to [0,1]$, a cdf, $G: (0,1) \to \mathbb{R}$

$U \sim \text{Uniform}([0,1])$ $\qquad G(u) = \inf\{x \in \mathbb{R} \mid u \leq F(x)\}$

Show $G(U)$ is a r.v with cdf F

If $G(u) \leq t$, $F(G(u)) \leq F(t)$ and $\lim_{h \to 0^+} F(G(u)+h) \leq F(t)$

Then $F(G(u)+h) \geq u$ (h: positive) so $F(t) \geq \lim_{h\to 0^+} u = u$.

If $F(t) \geq u$, by definition (of infimum) $t \geq G(u)$

Thus we have $G(u) \leq t \iff u \leq F(t)$.

Thus, $P(G(U) \leq t) = P(U \leq F(t)) = F(t)$. Thus $G(U)$ is r.v with CDF F.

#5. $Y = A^{-1}(X-b) = \phi(X)$, $\phi: \mathbb{R}^n \to \mathbb{R}^n$, one to one differentiable.

Then
$$P_X(x) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\|y\|^2} \cdot \left|\det \frac{\partial \phi}{\partial x}(x)\right|$$

$$= \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\|A^{-1}(x-b)\|^2} \left|\det A^{-1}\right| = \frac{1}{(2\pi)^{n/2}\det A} e^{-\frac{1}{2}(A^{-1}(x-b))^T A^{-1}(x-b)}$$

$$= \frac{1}{\sqrt{(2\pi)^n(\det A)^2}} e^{-\frac{1}{2}(x-b)^T (A^{-1})^T A^{-1}(x-b)} = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-b)^T (AA^T)^{-1}(x-b)}$$

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

#6 $\sigma^{-1}(\sigma(i)) = i$ for $i = 1, \dots, n$. Algorithm for computing σ^{-1} given σ?

```
Vector<int> σ⁻¹(n);
for (int i=1 ; i <=n ; i++)
    σ⁻¹[σ[i]] = i ;
```

하면 끝.

#7. $P_\sigma = \begin{bmatrix} e_{\sigma(1)}^T \\ \vdots \\ e_{\sigma(n)}^T \end{bmatrix} \in \mathbb{R}^{n \times n}$

(a) $(P_\sigma x)_i = e_{\sigma(i)}^T x = (x)_{\sigma(i)}$ // $\quad (x \in \mathbb{R}^n)$

(b) Note that $e_{\sigma(i)}^T$ and $e_{\sigma(j)}^T$ $\underset{i \neq j}{\wedge}$ are orthonormal. $\left(e_{\sigma(i)} \cdot e_{\sigma(j)} = 0, \ |e_{\sigma(i)}| = |e_{\sigma(j)}| = 1 \right)$

Thus, P_σ is an orthogonal matrix. Thus, $P_\sigma P_\sigma^T = P_\sigma^T P_\sigma = I$

Thus, $(P_\sigma)^{-1} = P_\sigma^T$

Also, $(P_{\sigma^{-1}})_{ij} = (e_{\sigma^{-1}(i)})_j = \begin{cases} 1 & \text{if } \sigma^{-1}(i) = j \\ 0 & \text{else} \end{cases}$

$(P_\sigma^T)_{ij} = (P_\sigma)_{ji} = (e_{\sigma(j)})_i = \begin{cases} 1 & \text{if } \sigma(j) = i \iff j = \sigma^{-1}(i) \\ 0 & \text{else.} \end{cases}$

Thus $P_\sigma^T = P_{\sigma^{-1}}$ holds.

(c) Since $\det(P_\sigma P_\sigma^T) = (\det(P_\sigma))^2 = 1$, $|\det P_\sigma| = 1$