MFDNN

Homework 9

24/05/12

3. $$\left(\frac{\partial \underline{z}}{\partial \underline{x}}\right)_{ij} = \frac{\partial z_i}{\partial x_j}$$

Jacobian matrix

Permute $\underline{z}$ using permutation $\sigma := \Omega \cup \Omega^c$ (in order union, i.e. $\{\Omega_1, \dots, \Omega_{|\Omega|}, \Omega^c_1, \dots, \Omega^c_{n-|\Omega|}\}$) such that

$$\underline{z}' := P_\sigma \underline{z} = \begin{pmatrix} (\underline{z}_\Omega)_1 \\ \vdots \\ (\underline{z}_\Omega)_{|\Omega|} \\ (\underline{z}_{\Omega^c})_1 \\ \vdots \\ (\underline{z}_{\Omega^c})_{n-|\Omega|} \end{pmatrix} \quad \text{by HW8.7}$$

$\underline{z} = P_\sigma^{-1} \underline{z}'$

$\Rightarrow \frac{\partial \underline{z}}{\partial \underline{z}'} = P_\sigma^{-1} = P_{\sigma^{-1}}$

Similarly for $\underline{x}' := P_\sigma \underline{x} \Rightarrow \frac{\partial \underline{x}'}{\partial \underline{x}} = P_\sigma$

Easily, $$\frac{\partial \underline{z}'}{\partial \underline{x}'} = \begin{pmatrix} \underline{I} & \underline{0} \\ * & \operatorname{diag}(e^{\underline{s}_\theta(\underline{x}_\Omega)}) \end{pmatrix}$$

unimportant for det: $\frac{\partial \underline{z}_{\Omega^c}}{\partial \underline{x}_\Omega}$

Then $$\frac{\partial \underline{z}}{\partial \underline{x}} = \frac{\partial \underline{z}}{\partial \underline{z}'} \frac{\partial \underline{z}'}{\partial \underline{x}'} \frac{\partial \underline{x}'}{\partial \underline{x}} = P_{\sigma^{-1}} \frac{\partial \underline{z}'}{\partial \underline{x}'} P_\sigma \quad \text{(as hinted)}$$

$$\left|\frac{\partial \underline{z}}{\partial \underline{x}}\right| = \left|\det\left(P_{\sigma^{-1}} \frac{\partial \underline{z}'}{\partial \underline{x}'} P_\sigma\right)\right| \overset{\text{det properties}}{=} \left|\det(P_{\sigma^{-1}}) \det\left(\frac{\partial \underline{z}'}{\partial \underline{x}'}\right) \det(P_\sigma)\right| \overset{\text{HW8.7}}{=} \left|1 \cdot \exp\left(\underline{1}^T_{n-|\Omega|} \underline{s}_\theta(\underline{x}_\Omega)\right) \cdot 1\right|$$

$$\Rightarrow \log\left|\frac{\partial \underline{z}}{\partial \underline{x}}\right| = \underline{1}^T_{n-|\Omega|} \underline{s}_\theta(\underline{x}_\Omega) \quad \blacksquare$$

4.(a) f and g are PDFs $\Rightarrow f(\underline{x}), g(\underline{x}) \geq 0 \ \forall \underline{x} \in \mathbb{R}^d$ and $\int_{\mathbb{R}^d} f(\underline{x})\, d\underline{x} = \int_{\mathbb{R}^d} g(\underline{x})\, d\underline{x} = 1$

~~This latter property implies that $g(\underline{x})$ is non-zero at least somewhere so $\frac{f(\underline{x})}{g(\underline{x})}$ is defined somewhere and is positive.~~ ~~→ on a set of measure 0~~

~~$D_{KL}(X\|Y) = \int_{\mathbb{R}^d} f(x) \log\left(\frac{f(x)}{g(x)}\right) dx \geq 0$ since f and $\log\frac{f}{g}$ both equal to or greater than 0 over all of $\mathbb{R}^d$~~

~~(equal to 0 iff $f(x) = g(x)$ almost everywhere?)~~

Or use Jensen's inequality again?

Define Random Variable, $Z := \frac{g(\underline{z})}{f(\underline{z})}$ $\qquad$ $-\log$ is convex (HW2.4)

$$D_{KL}(\underline{X}\|\underline{Y}) = \int_{\mathbb{R}^d} f(\underline{x}) \log\left(\frac{f(\underline{x})}{g(\underline{x})}\right) d\underline{x} = \mathbb{E}_f[-\log(Z)] \overset{\text{Jensen's ineq.}}{\geq} -\log(\mathbb{E}_f[Z]) = -\log\left(\int_{\mathbb{R}^d} \frac{g(\underline{x})}{f(\underline{x})} f(\underline{x})\, d\underline{x}\right)$$

$$= -\log\left(\int_{\mathbb{R}^d} g(\underline{x})\, d\underline{x}\right) \overset{g \text{ is a PDF}}{=} -\log(1) = 0$$

$$\Rightarrow D_{KL}(X\|Y) \geq 0 \quad \blacksquare$$

4.(b) Independent $\Rightarrow f(\underline{x}) = f_1(x_1) f_2(x_2) \cdots f_d(x_d)$ i.e. f can be decomposed into separate PDFs

Analogous decomposition for $g(\underline{x})$

$$D_{KL}(X_n\|Y_n) = \int_{\mathbb{R}} f(x_n) \log\left(\frac{f(x_n)}{g(x_n)}\right) dx_n$$

$$D_{KL}(\underline{X}\|\underline{Y}) = \int_{\mathbb{R}^d} f(\underline{x}) \log\left(\frac{f(\underline{x})}{g(\underline{x})}\right) d\underline{x} = \underbrace{\int_{\mathbb{R}} \cdots \int_{\mathbb{R}}}_{d \text{ times}} f(\underline{x}) \log\left(\frac{f(\underline{x})}{g(\underline{x})}\right) dx_1 \cdots dx_d$$ by Fubini's theorem (PDF Lebesgue-integrable)

$$= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_1(x_1) \cdots f_d(x_d) \log\left(\frac{f_1(x_1) \cdots f_d(x_d)}{g_1(x_1) \cdots g_d(x_d)}\right) dx_1 \cdots dx_d$$

$$= \sum_{n=1}^{d} \left(\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_1(x_1) \cdots f_d(x_d) \log\left(\frac{f_n(x_n)}{g_n(x_n)}\right) dx_1 \cdots dx_d \right)$$

$$= \sum_{n=1}^{d} \left(\int_{\mathbb{R}} f_1(x_1) \cdots \int_{\mathbb{R}} f_{n-1}(x_{n-1}) \int_{\mathbb{R}} f_n(x_n) \log\left(\frac{f_n(x_n)}{g_n(x_n)}\right) \int_{\mathbb{R}} f_{n+1}(x_{n+1}) \cdots \int_{\mathbb{R}} f_d(x_d)\, dx_d \cdots dx_{n+1}\, dx_n\, dx_{n-1} \cdots dx_1 \right)$$

$$= \sum_{n=1}^{d} \left(\int_{\mathbb{R}} f_n(x_n) \log\left(\frac{f_n(x_n)}{g_n(x_n)}\right) dx_n \right)$$ since all other integrals evaluate to 1 since f_i is a PDF

$$= \sum_{n=1}^{d} D_{KL}(X_n\|Y_n) = D_{KL}(X_1\|Y_1) + \ldots + D_{KL}(X_d\|Y_d) \quad \blacksquare$$

5. $\underline{N}_i := \mathcal{N}(\mu_i, \Sigma_i)$

Positive definite: $\underline{x}^T \Sigma_i \underline{x} > 0 \ \forall \underline{x} \in \mathbb{R}^d \setminus \{\underline{0}\}$

⇒ invertible with inverse also pos-def

⇒ symmetric

PDFs:

$$n_i(\underline{x}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} \exp\left(-\frac{1}{2}(\underline{x}-\mu_i)^T \Sigma_i^{-1} (\underline{x}-\mu_i)\right)$$

$$D_{KL}(\underline{N}_0 \| \underline{N}_1) = \int_{\mathbb{R}^d} n_0(\underline{x}) \log\left(\frac{n_0(\underline{x})}{n_1(\underline{x})}\right) d\underline{x}$$

$$= \frac{1}{\sqrt{(2\pi)^d |\Sigma_0|}} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}(\underline{x}-\mu_0)^T \Sigma_0^{-1} (\underline{x}-\mu_0)\right)\Bigg(\log\left(\frac{\sqrt{|\Sigma_1|}}{\sqrt{|\Sigma_0|}}\right) - \frac{1}{2}(\underline{x}-\mu_0)^T \Sigma_0^{-1}(\underline{x}-\mu_0) + \frac{1}{2}(\underline{x}-\mu_1)^T \Sigma_1^{-1}(\underline{x}-\mu_1)\Bigg) d\underline{x}$$

$$= \frac{1}{2}\log\left(\frac{\det\Sigma_1}{\det\Sigma_0}\right) + \frac{1}{(2\pi)^{d/2}|\Sigma_0|^{1/2}} \exp\left(-\frac{1}{2}(\underline{x}-\mu_0)^T \Sigma_0^{-1}(\underline{x}-\mu_0)\right)$$

$$+ \frac{1}{\sqrt{(2\pi)^d|\Sigma_0|}} \int_{\mathbb{R}^d} \frac{1}{2}(\underline{x}-\mu_1)^T \Sigma_1^{-1}(\underline{x}-\mu_1) \exp\left(-\frac{1}{2}(\underline{x}-\mu_0)^T \Sigma_0^{-1}(\underline{x}-\mu_0)\right) d\underline{x} \ \ldots ?$$

Via expectations:

i.e. $\underline{x}$ is a R.V. $\underline{x} \sim \mathcal{N}(\mu_0, \Sigma_0)$ i.e. like N_0

$$D_{KL}(\underline{N}_0 \| \underline{N}_1) = \mathbb{E}_{\underline{N}_0}\left[\log\left(\frac{n_0(\underline{x})}{n_1(\underline{x})}\right)\right] \quad \left(= \int_{\mathbb{R}^d} n_0(\underline{x}) \log\left(\frac{n_0(\underline{x})}{n_1(\underline{x})}\right) d\underline{x}\right)$$

via PDFs

$$= \mathbb{E}_{\underline{N}_0}\left[\frac{1}{2}\log\frac{\det\Sigma_1}{\det\Sigma_0} - \frac{1}{2}(\underline{x}-\mu_0)^T\Sigma_0^{-1}(\underline{x}-\mu_0) + \frac{1}{2}(\underline{x}-\mu_1)\Sigma_1^{-1}(\underline{x}-\mu_1)\right]$$

via linearity

$$= \frac{1}{2}\left(\log\frac{\det\Sigma_1}{\det\Sigma_0} + \mathbb{E}_{\underline{N}_0}\left[(\underline{x}-\mu_1)^T\Sigma_1^{-1}(\underline{x}-\mu_1)\right] - \mathbb{E}_{\underline{N}_0}\left[(\underline{x}-\mu_0)^T\Sigma_0^{-1}(\underline{x}-\mu_0)\right]\right)$$

rewrite

Quadratic form $\mathbb{E}[\underline{x}^T\Lambda\underline{x}] = tr[\Lambda\Sigma] + \mu^T\Lambda\mu$

$$= \frac{1}{2}\left(\log\frac{\det\Sigma_1}{\det\Sigma_0} + \mathbb{E}_{\underline{N}_0}\left[(\underline{x}-\mu_0+(\mu_0-\mu_1))^T\Sigma_1^{-1}(\underline{x}-\mu_0+(\mu_0-\mu_1))\right] - tr(\underbrace{\Sigma_0^{-1}\Sigma_0}_{I_d}) - 0\right)$$

mean of $\underline{x}-\mu_0$ is $\underline{0}$

Σ_1^{-1} pos-def (⇒ symmetric)

$(\underline{a}^T X \underline{b})^T = \underline{b}^T X^T \underline{a}$

$\underline{a}^T X \underline{b} = \underline{b}^T X \underline{a}$

since mean of $\underline{x}$ is μ_0. (covariance matrix unchanged)

$$= \frac{1}{2}\left(\log\frac{\det\Sigma_1}{\det\Sigma_0} - d + \mathbb{E}_{\underline{N}_0}\left[(\underline{x}-\mu_0)\Sigma_1^{-1}(\underline{x}-\mu_0)\right] + 2(\mu_0-\mu_1)\Sigma_1^{-1}\mathbb{E}_{\underline{N}_0}[\underline{x}-\mu_0] + (\mu_0-\mu_1)^T\Sigma_1^{-1}(\mu_0-\mu_1)\right)$$

$$= \frac{1}{2}\left(\log\frac{\det\Sigma_1}{\det\Sigma_0} - d + tr(\Sigma_1^{-1}\Sigma_0) + 0 + 0 + (\mu_0-\mu_1)^T\Sigma_1^{-1}(\mu_0-\mu_1)\right)$$

$\underline{a}^T X \underline{b} = (-1)\cdot\underline{a}^T X \underline{b}\cdot(-1) = (-\underline{a})^T X (-\underline{b})$

simple sign change

$$= \frac{1}{2}\left(tr(\Sigma_1^{-1}\Sigma_0) + (\mu_1-\mu_0)^T\Sigma_1^{-1}(\mu_1-\mu_0) - d + \log\frac{\det\Sigma_1}{\det\Sigma_0}\right) \ \blacksquare$$

6. $$f(\theta) = g(\theta,\phi) + h(\theta,\phi) \quad \forall \phi \in \Phi$$

$\forall \theta \in \Theta \ \exists \phi \in \Phi$ s.t. $h(\theta,\phi) = 0$ so,

$$= g(\theta, \phi^*(\theta)) \qquad \text{with } \phi^*(\theta) := \underset{\phi \in \Phi}{\operatorname{argmin}}\, h(\theta,\phi)$$

$$\Rightarrow \max_{\theta} f(\theta) = \max_{\theta} g(\theta, \phi^*(\theta))$$

?

$$= \sup_{\theta,\phi} g(\theta,\phi) = \sup_{\theta}\left(\sup_{\phi} g(\theta,\phi)\right) \ ?$$

$$\Rightarrow \underset{\theta}{\operatorname{argmax}}\, f = \{\theta \mid (\theta,\phi) \in \underset{\theta,\phi}{\operatorname{argmax}}\, g\}$$