PML Theory

Sam Laing 6283670
Albert Catalan 6443478

a) We have

$$p(y|X) = \int_{\forall f} p(y|f, X)\, p(f|X)\, df$$

we have $p(f) \sim N(\mu, k) \Rightarrow p(f|X) = N(f; \mu_x, K_{xx})$

$$\Rightarrow p(y|X) = \int_{\forall f} N(y; f(X), \Lambda)\, N(f; \mu_x, K_{xx})\, df$$

$$= \int_{\forall f} N(y; \mu_x, \Lambda + K_{xx})\, N(f; (\Lambda^{-1} + K_{xx})^{-1}(\Lambda^{-1} f(x) + K_{xx}^{-1}\mu_x), (\Lambda^{-1} + K_{xx}^{-1})^{-1})\, df$$

By the Result proved in last week's hw.

$$= N(y; \mu_x, \Lambda + K_{xx}) \int_{\forall f} N(f; (\Lambda^{-1} + K_{xx}^{-1})^{-1}(\Lambda^{-1} f(x) + K_{xx}^{-1}\mu_x), (\Lambda^{-1} + K_{xx}^{-1})^{-1})\, df$$

$$= N(y; \mu_x, \Lambda + K_{xx}) \cdot 1$$

(since we integrate over a distribution)

$\boxed{//}$

b) i) $y \mid x \sim N(\mu_x, \Lambda + K_{xx})$

$$\mathbb{E}_{y|x}\left[\log(p(y|x))\right]$$

$$= -\frac{1}{2}\Big\{ \mathbb{E}_{y|x}\left[(y-\mu_x)^T (K_{xx}+\Lambda)^{-1}(y-\mu_x)\right]$$

doesn't depend on y

$$+ \mathbb{E}_{y|x}\left[\log\det(K_{xx}+\Lambda)\right] + n\log 2\pi \Big\}$$

$$= -\frac{1}{2}\operatorname{tr}\left[(K_{xx}+\Lambda)^{-1}(\cancel{A}K_{xx}+\Lambda)\right]$$

By suggested identity

$$-\frac{1}{2}(\cancel{y}^{\mu_x}-\mu_x)^T (K_{xx}+\Lambda)^{-1}(\cancel{y}^{\mu_x}-\mu_x)$$

$$-\frac{1}{2}\log\det(K_{xx}+\Lambda) - n\log 2\pi$$

$$= -\frac{1}{2}\operatorname{tr}(\mathbb{1}) + 0 - \frac{1}{2}\log\det(K_{xx}+\Lambda)$$

$$-\frac{n}{2}\log 2\pi$$

$$= -\frac{n}{2}(1+\log 2\pi) - \frac{1}{2}\log\det(K_{xx}+\Lambda)$$

(b) (iii)

$$\mathcal{K}_{xx} = \mathcal{K}_{xx}(\theta)$$

$$\log p(y|X) = -\frac{1}{2}\Big[(y-\mu_x)^T(\mathcal{K}_{xx}+\Lambda^{-1})^{-1}(y-\mu_x) + \log\det(\mathcal{K}_{xx}+\Lambda^{-1}) + n\log 2\pi\Big]$$

Lets first prove the result & then prove the required identities:

$$\partial_\theta \log p(y|X) = \frac{1}{p(y|X)}\,\partial_\theta p(y|X)$$

we note that

$$\partial_\theta\left[(y-\mu_x)^T(\mathcal{K}_{xx}+\Lambda)^{-1}(y-\mu_x)\right]$$

$$= \left(\partial_\theta (y-\mu_x)^T\right)(\mathcal{K}_{xx}+\Lambda)^{-1}(y-\mu_x) + (y-\mu_x)^T\partial_\theta(\mathcal{K}_{xx}+\Lambda)^{-1}(y-\mu_x)$$

Chain rule

$$+ (y-\mu_x)(\mathcal{K}_{xx}+\Lambda)^{-1}\partial_\theta(y-\mu_x)$$

$$= (y-\mu_x)^T\,\partial_\theta(\mathcal{K}_{xx}+\Lambda)^{-1}(y-\mu_x)$$

by result we will prove

$$= -(y-\mu_x)^T\left[(\mathcal{K}_{xx}+\Lambda)^{-1}\frac{\partial \mathcal{K}_{xx}}{\partial\theta}(\mathcal{K}_{xx}+\Lambda)^{-1}\right](y-\mu_x)$$

~~and~~

$$\cancel{A)\ \partial_\theta \log p(y|X) = (y-\mu_x)^T(\mathcal{K}_{xx}+\Lambda^{-1})\frac{\partial \mathcal{K}_{xx}}{\partial\theta}(\mathcal{K}_{xx}+\Lambda^{-1})^{-1}(y-\mu_x)}$$

using 2nd result

$$\Rightarrow \partial_\theta \log p(y|x) \overset{\circ}{=} \frac{1}{p(y|x)} \left[-(y-\mu_x)^T (K_{xx}+\Lambda)^{-1} \frac{\partial K_{xx}}{\partial \theta} (K_{xx}+\Lambda)^{-1} (y-\mu_x) + tr\left(K_{xx}^{-1} \frac{\partial K_{xx}}{\partial \theta}\right) \right]$$

We now prove the two statements we made use of above:

- let $K \equiv K_{xx}(\theta)$ have inverse $H \equiv H_{xx}(\theta)$

Then $KH = \mathbb{1}$. differentiating both sides (and noting that the product rule applies for matrices) gives:

$$(\partial_\theta K) H + K(\partial_\theta H) = 0$$

$$\Rightarrow K(\partial_\theta H) = -(\partial_\theta K) H$$

multiplying both sides from the left by $K^{-1} = H$ then gives:

$$\partial_\theta (K_{xx}(\theta))^{-1} = -K_{xx}^{-1} \frac{\partial K_{xx}}{\partial \theta} K_{xx}^{-1} \quad \boxed{\text{Th}}$$

(this result holds $\forall$ matrices of full rank)

- We observe that since $K_{xx}(\theta)$ is psd., $\exists$ $U(\theta)$ orthogonal, s.t. $D(\theta)$ positive diagonal such that:

$$K_{xx}(\theta) = U(\theta) D(\theta) U(\theta)^T$$

So we have $D(\theta) = diag(\lambda_j(\theta))_{j=1}^n$ for continuous $(\lambda_j : \mathbb{R} \to \mathbb{R})_{j=1}^n$

then $\det(K_{xx}(\theta)) = \prod_{j=1}^{n} \lambda_j(\theta)$

$$\Rightarrow \partial_\theta \log(\det(K_{xx}(\theta))) = \partial_\theta \log \prod_{j=1}^{n} \lambda_j(\theta)$$

$$= \partial_\theta \sum_{j=1}^{n} \log \lambda_j(\theta)$$

$$(*) \quad = \sum_{j=1}^{n} \frac{1}{\lambda_j(\theta)} \lambda_j'(\theta) \quad \text{(product rule)}$$

we observe that $K_{xx}^{-1} = U D(\theta) U^T$
and $D^{-1}(\theta) = \mathrm{diag}(1/\lambda_j(\theta))_{j=1}^{n}$

also $\frac{\partial K_{xx}}{\partial \theta} = U \,\mathrm{diag}(\lambda_j'(\theta))_j \, U^T$

we therefore recognize the expression $(*)$ as

$$\mathrm{tr}\left(U \,\mathrm{diag}(1/\lambda_j(\theta))_{j=1}^{n} U^T \cdot U \,\mathrm{diag}(\lambda_j'(\theta))_{j=1}^{n} U \right)$$

(since $U^T = U^{-1}$)

$$= \mathrm{tr}\left(K_{xx}^{-1} \frac{\partial K_{xx}}{\partial \theta} \right)$$

$\blacksquare$