PS 3.10 Bayesian Regression w/ Gaussian Prior

$\Phi = [\phi(x_1) \dots \phi(x_n)]$ ⇐ inputs (features)

$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ ⇐ observations

function: $\Phi^T\theta$ $\quad \theta \in \mathbb{R}^D$

$\boxed{y = \Phi^T\theta + \epsilon}$, $\epsilon \sim N(0, \Sigma)$ m.v. Gaussian

Prior: $p(\theta) = N(\theta | 0, \Gamma)$

a) Given dataset $D = \{(x_i, y_i)\}_{i=1}^N$

Find $p(\theta | D)$.

$$p(\theta | D) = \frac{p(D|\theta)\, p(\theta)}{p(D)}$$

$$p(D|\theta) = p(y|x,\theta) = N(y | \Phi^T\theta, \Sigma)$$

$$p(\theta|D) = \frac{N(y|\Phi^T\theta, \Sigma)\, N(\theta|0,\Gamma)}{\underbrace{p(D)}_{\text{no } \theta}} : e^{-\frac{1}{2}\|\Phi^T\theta - y\|^2_\Sigma}\, e^{-\frac{1}{2}\|\theta\|^2_\Gamma}$$

Since θ is in the exponent of numerator, look at log numerator as a function of θ:

$$\log N(y | \Phi^T\theta, \Sigma) + \log N(\theta | 0, \Gamma) =$$

$$= \underbrace{-\frac{1}{2}\|y - \Phi^T\theta\|^2_\Sigma}_{} \underbrace{- \frac{1}{2}\log|\Sigma|}_{\times} \underbrace{- \frac{n}{2}\log 2\pi}_{\times} \underbrace{- \frac{1}{2}\|\theta\|^2_\Gamma}_{} \underbrace{- \frac{1}{2}\log|\Gamma| - \frac{D}{2}\log 2\pi}_{\times}$$

$$= -\frac{1}{2}(\underbrace{y^T\Sigma^{-1}y}_{\times} - 2y^T\Sigma^{-1}\Phi^T\theta + \underbrace{\theta^T\Phi\Sigma^{-1}\Phi^T\theta}_{①}) - \frac{1}{2}\underbrace{\theta^T\Gamma^{-1}\theta}_{①} + \text{const}$$

$$= -\frac{1}{2}\left(\theta^T\underbrace{(\underbrace{\Phi\Sigma^{-1}\Phi^T}_{①} + \Gamma^{-1})}_{A}\theta - 2\underbrace{y^T\Sigma^{-1}\Phi^T}_{b^T}\theta\right) + \text{const}$$

$$= -\frac{1}{2}(\theta^TA\theta - 2b^T\theta) + \text{const}$$

completing the square (PS1-10) (A is sym)

$f(x) = x^TAx - 2x^Tb + c$
$= (x-d)^TA(x-d) + e$
$\begin{cases} d = A^{-1}b \\ e = c - b^TA^{-1}b \end{cases}$

$$= -\frac{1}{2}(\theta - \hat{\mu})^T A (\theta - \hat{\mu}) + \text{const}$$

$\hat{\mu} = A^{-1}b$

$$\boxed{\hat{\mu} = (\Phi\Sigma^{-1}\Phi^T + \Gamma^{-1})^{-1}\Phi\Sigma^{-1}y}$$

$$\boxed{\hat{\Sigma} = A^{-1} = (\Phi\Sigma^{-1}\Phi^T + \Gamma^{-1})^{-1}}$$

$$= -\frac{1}{2}\|\theta - \hat{\mu}\|^2_{\hat{\Sigma}} + \text{const}$$

$$p(\theta|D) \propto e^{-\frac{1}{2}\|\theta - \hat{\mu}\|^2_{\hat{\Sigma}}} \cdot \text{const} \; \boxed{= N(\theta | \hat{\mu}, \hat{\Sigma})}$$

CS5487 Lecture Notes (2022B)
Prof. Antoni B. Chan
Dept of Computer Science
City University of Hong Kong

c) MAP estimate w/ $\Gamma = \alpha I$, $\Sigma = \sigma^2 I$

$$\hat{\theta}_{MAP} = \arg\max_{\theta} \; p(\theta | D)$$

$$= \arg\max_{\theta} \; N(\theta | \hat{\mu}, \hat{\Sigma})$$

$$= \hat{\mu}$$

$$= (\Phi \Sigma^{-1} \Phi^T + \Gamma^{-1})^{-1} \Phi \Sigma^{-1} y \quad \Leftarrow \quad (b)$$

$$= \Big(\Phi \tfrac{1}{\sigma^2} I \Phi^T + \tfrac{1}{\alpha} I\Big)^{-1} \Phi \Big[\tfrac{1}{\sigma^2}\Big] I \, y \quad \cdot \sigma^2$$

$$= \Big(\Phi \Phi^T + \underbrace{\tfrac{\sigma^2}{\alpha}}_{\lambda} I\Big)^{-1} \Phi y$$

$$\boxed{\hat{\theta}_{MAP} = (\Phi \Phi^T + \lambda I)^{-1} \Phi y}$$

(soln of regularized LS problem)

e) Given a novel x_* , find distribution of $f_* = \phi(x_*)^T \theta$

r.v.

i.e. find

$$p(f_* | x_*, D)$$

$$p(\theta | D) = N(\theta | \hat{\mu}, \hat{\Sigma})$$

PS 1-1

$y = Ax + b$

$E[y] = A E[x] + b$

$\mathrm{cov}(y) = A \, \mathrm{cov}(x) A^T$

$$\underbrace{f_* = \phi_*^T \theta}_{\text{linear xform of m.v. Gaussian}} \sim p(\theta | D) = N(\theta | \hat{\mu}, \hat{\Sigma})$$

Thus,

$$p(f_* | x_*, D) = N\Big(f_* \,\Big|\, \underbrace{\phi_*^T \hat{\mu}}, \underbrace{\phi_*^T \hat{\Sigma} \phi_*}\Big)$$

uncertainty of the predicted function value.

[Gaussian Process regression
Bayesian Linear regression