Lecture 12 – Bayesian Linear Regression

Goal: Find BLR posterior

$$p(\theta) = \mathcal{N}(\theta;\, 0,\, I_M)$$
$$p(y|\theta) = \mathcal{N}(y;\, \phi(x)\theta,\, \sigma^2 I_N)$$

Method 1: Direct algebra

Similar procedure to before...

$$\log p(\theta|y) = \log p(y|\theta) + \log p(\theta)$$
$$= c - \frac{1}{2\sigma^2}(y-\phi(x)\theta)^T(y-\phi(x)\theta) - \frac{1}{2}\theta^T\theta$$

Let's try to equate coefficients with Gaussian (since quadratic).

$$\log p(\theta|y) = \log \mathcal{N}(\theta;\, \mu_p,\, \Sigma_p)$$
$$= c - \frac{1}{2}(\theta-\mu_p)^T\Sigma_p^{-1}(\theta-\mu_p)$$
$$= c' - \frac{1}{2}\theta^T\Sigma_p^{-1}\theta + \mu_p^T\Sigma_p^{-1}\theta$$
$$= -\frac{1}{2}\theta^T\left[\frac{1}{\sigma^2}\phi(x)^T\phi(x) + I_M\right]\theta + \frac{1}{\sigma^2}y^T\phi(x)\theta$$

$$\Rightarrow \Sigma_p = \left[\frac{1}{\sigma^2}\phi(x)\phi(x)^T + I_M\right]^{-1}$$
$$\mu_p^T\Sigma_p^{-1} = \frac{1}{\sigma^2}y^T\phi(x)$$
$$\Rightarrow \mu_p = \frac{1}{\sigma^2}\Sigma_p\phi(x)^T y = \left[\frac{1}{\sigma^2}\phi(x)\phi(x)^T + I_M\right]^{-1}\frac{1}{\sigma^2}\phi(x)^T y \quad \square$$

Method 2: Joint Gaussian

$$p(\theta, y) = \mathcal{N}\left(\begin{bmatrix}\theta\\ y\end{bmatrix}; \begin{bmatrix}\mathbb{E}[\theta]\\ \mathbb{E}[y]\end{bmatrix}, \begin{bmatrix}\mathbb{V}[\theta] & \mathbb{C}[\theta, y]\\ \mathbb{C}[y, \theta] & \mathbb{V}[y]\end{bmatrix}\right)$$

$$\begin{aligned} \theta &= \theta \\ y &= \phi(x)\,\theta + \epsilon \end{aligned}$$

$$p(\theta, \epsilon) = \mathcal{N}\left(\begin{bmatrix}\theta\\ \epsilon\end{bmatrix}; 0, \begin{bmatrix}I_M & 0\\ 0 & \sigma^2 I_N\end{bmatrix}\right)$$

↳ Since $\theta \perp\!\!\!\perp \epsilon$ (i.e. independent)

All the following expectations are under $p(\theta, \epsilon)$

$$\mathbb{E}[\theta] = 0$$

$$\mathbb{E}[y] = \cancelto{0}{\mathbb{E}[\phi(x)\theta]} + \cancelto{0}{\mathbb{E}[\epsilon]} = 0$$

$$\mathbb{V}[\theta] = I_M$$

$$\begin{aligned} \mathbb{V}[y] &= \mathbb{V}[\phi(x)\theta] + \mathbb{V}[\epsilon] && \text{- using } \theta \perp\!\!\!\perp \epsilon \text{ + LOTUS} \\ &= \phi(x)\phi(x)^T + \sigma^2 I_N && \text{- } \mathbb{V}[Ax] = A\mathbb{V}[x]A^T \\ \mathbb{C}[\theta, y] &= \mathbb{E}[\theta y^T] - \cancelto{0}{\mathbb{E}[\theta]}\,\cancelto{0}{\mathbb{E}[y]^T} && \text{- identity} \\ &= \mathbb{E}[\theta(\phi(x)\theta + \epsilon)^T] && \text{- LOTUS} \\ &= \mathbb{E}[\theta\theta^T\phi(x)^T] + \cancelto{0}{\mathbb{E}[\theta]}\,\cancelto{0}{\mathbb{E}[\epsilon]^T} && \text{- independence} \\ &= \phi(x)^T \end{aligned}$$

$$\Longrightarrow p(\theta \mid y) = \mathcal{N}\left(\theta; \phi(x)^T[\phi(x)\phi(x)^T + \sigma^2 I_N]^{-1} y,\ I_M - \phi(x)^T[\phi(x)\phi(x)^T + \sigma^2 I_N]^{-1}\phi(x)\right) \quad \square$$

New Goal: Find predictive distribution $p(y^*|y)$

$$y^* = \phi(x^*)\theta + \epsilon^* \qquad y^*, \epsilon^* \in \mathbb{R}^{N^*}$$
$$y = \phi(x)\theta + \epsilon \qquad y, \epsilon \in \mathbb{R}^{N}$$

$$\theta \sim N(0, I_M)$$
$$\epsilon \sim N(0, \sigma^2 I_N)$$
$$\epsilon^* \sim N(0, \sigma^2 I_{N^*})$$

$\Rightarrow$ Linear Gaussian model

$\Rightarrow$ All joints are Gaussian

$\Rightarrow$ All conditionals are Gaussian

$$p(y^*|y) = N(y^*; \mu^*, \Sigma^*)$$

$$
\begin{aligned}
\mu^* &= \mathbb{E}_{y^*}[y^*|y] \\
&= \mathbb{E}_{\theta,\epsilon^*}[\phi(x^*)\theta|y] + \cancel{\mathbb{E}[\epsilon^*|y]}^{\,0} && \text{- LOTUS} \\
&= \phi(x^*)\mu_p \quad \text{From earlier posterior!} \\
&= \underline{\underline{\phi(x^*)\left[\tfrac{1}{\sigma^2}\phi(x)\phi(x)^T + I_M\right]^{-1}\tfrac{1}{\sigma^2}\phi(x)^T y}}
\end{aligned}
$$

$$
\begin{aligned}
\Sigma^* &= \mathbb{V}_{y^*}[y^*|y] \\
&= \mathbb{V}[\phi(x^*)\theta|y] + \mathbb{V}[\epsilon^*|y] && \text{- } \theta \perp\!\!\!\perp \epsilon^* | y \text{ + LOTUS} \\
&= \phi(x^*)\Sigma_p\phi(x^*)^T + \sigma^2 I_{N^*}
\end{aligned}
$$

From earlier posterior! Note: Not the prior b/c we need $\mathbb{V}[\phi(x^*)\theta|\underline{y}]$!

$$= \underline{\underline{\phi(x^*)\left[\tfrac{1}{\sigma^2}\phi(x)\phi(x)^T + I_M\right]^{-1}\phi(x^*) + \sigma^2 I_{N^*}}}$$