Probset #2

3. (c)

Linear regression model : $y = \theta^T x + \epsilon$, $\epsilon \sim \mathcal{N}(0, \sigma^2)$

Gaussian prior $\theta \sim \mathcal{N}(0, \eta^2 I)$

$$X = \Bigg[\overbrace{\;- \; (x^{(i)})^T \; -\;}^{n}\Bigg]\Big\}m \quad \vec{y}\Bigg[y^{(i)}\Bigg]\Big\}m \quad \theta : \Bigg[\theta^{(i)}\Bigg]\Big\}n$$

Come up with a closed form expression for θ_{MAP}

$$\theta_{MAP} = \arg\min_{\theta} \left\{ -\log p(y|\theta, x) + \frac{1}{2}\left(\frac{\theta}{\eta}\right)^2 \right\}$$

$$y^{(i)} = \theta^T x^{(i)} + \epsilon^{(i)}$$

$$y^{(i)} | x^{(i)}, \theta \sim \mathcal{N}(\theta^T x^{(i)}, \sigma^2)$$

$$p(y^{(i)}|x^{(i)}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ \frac{-1}{2\sigma^2}(y^{(i)} - \theta^T x^{(i)})^2 \right\}$$

$$p(\vec{y}|X, \theta) = \prod_{i=1}^{m} p(y^{(i)}|x^{(i)}, \theta)$$

$$= \prod_{i=1}^{m} \left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2\sigma^2}(y^{(i)} - \theta^T x^{(i)})^2 \right\} \right]$$

$$\theta_{MAP} = \arg\min_{\theta} \left\{ -\log p(y|\theta, x) + \frac{1}{2}\left(\frac{\theta}{\eta}\right)^2 \right\}$$

$$= \arg\min_{\theta} \left\{ -\log \prod_{i=1}^{m} \left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2\sigma^2}(y^{(i)} - \theta^T x^{(i)})^2 \right\} \right] + \frac{1}{2}\left(\frac{\theta}{\eta}\right)^2 \right\}$$

$$= \arg\min_{\theta} \left\{ -\sum_{i=1}^{m} \log \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y^{(i)} - \theta^T x^{(i)})^2 \right) + \frac{1}{2}\left(\frac{\theta}{\eta}\right)^2 \right\}$$

$$= \arg\min_{\theta} \left\{ -\sum_{i=1}^{m} \left[\log \frac{1}{\sqrt{2\pi}\sigma} + \left(-\frac{1}{2\sigma^2}(y^{(i)} - \theta^T x^{(i)})^2 \right) \right] + \frac{1}{2}\left(\frac{\theta}{\eta}\right)^2 \right\}$$

$$\theta_{MAP} = \underset{\theta}{\operatorname{argmin}} \left\{ -\sum_{i=1}^{m} \left[\log \frac{1}{\sqrt{2\pi}\sigma} + \left(-\frac{1}{2\sigma^2} \left(y^{(i)} - \theta^T x^{(i)} \right)^2 \right) \right] + \frac{1}{2} \left(\frac{\theta}{\eta} \right)^2 \right\}$$

$$= \underset{\theta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{m} \left(\frac{1}{2\sigma^2} \left(y^{(i)} - \theta^T x^{(i)} \right)^2 \right) + \frac{1}{2} \left(\frac{\theta}{\eta} \right)^2 \right\}$$

$$= \underset{\theta}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma^2} \overbrace{(\underbrace{\vec{y}}_{(m,1)} - \underbrace{X\theta}_{(m,1)})^T (\vec{y} - X\theta)}^{(1,1)} + \frac{1}{2\eta^2} \theta^T \theta \right\}$$

$$= \underset{\theta}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma^2} (\vec{y}^T - \theta^T X^T)(\vec{y} - X\theta) + \frac{1}{2\eta^2} \theta^T \theta \right\}$$

$$= \underset{\theta}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma^2} \left[y^T y - \underbrace{y^T X\theta - \theta^T X^T \vec{y}}_{=} + \theta^T X^T X \theta \right] + \frac{1}{2\eta^2} \theta^T \theta \right\}$$

$y : (m,1)$, $X : (m,n)$ $\theta : (n,1)$

$y^T X \theta : (1,m)(m,n)(n,1) \in \mathbb{R}$

$\theta^T X^T y : (1,n)(n,m)(m,1) \in \mathbb{R}$

$=$

$$l(\theta) = \frac{1}{2\sigma^2} y^T y - \frac{1}{\sigma^2} y^T X \theta + \frac{1}{2\sigma^2} \theta^T X^T X \theta + \frac{1}{2\eta^2} \theta^T \theta$$

$$\frac{\partial l(\theta)}{\partial \theta} = \nabla_\theta l(\theta) = -\frac{1}{\sigma^2} X^T y + \frac{1}{\sigma^2} X^T X \theta + \frac{1}{\eta^2} \theta = 0$$

$$\frac{1}{\sigma^2} X^T X \theta + \frac{1}{\eta^2} \theta = \frac{1}{\sigma^2} X^T y$$

$$\left(\frac{1}{\sigma^2} X^T X + \frac{1}{\eta^2} \right) \theta = \frac{1}{\sigma^2} X^T y$$

$$\left(X^T X + \left(\frac{\sigma}{\eta} \right)^2 \right) \theta = X^T y$$

$$\theta = \left(X^T X + \left(\frac{\sigma}{\eta} \right)^2 I \right)^{-1} X^T y$$

$$\therefore \theta_{MAP} = \underset{\theta}{\operatorname{argmin}}\, l(\theta)$$

$$= \left(X^T X + \frac{\sigma^2}{\eta^2} I \right)^{-1} X^T y$$