Data Mining Hw3 - Maha Alkhairy.

Ex 1 :-

$$A^{-1} = (X^T X + \lambda I)^{-1} \quad ; \quad A = (X^T X + \lambda I)$$

ⓐ

$$\nabla_\omega L(\omega)\Big|_{\omega = \omega^* = (X^T X + \lambda I)^{-1} X^T y} =$$

$$2X^T (X \underbrace{(X^T X + \lambda I)^{-1}}_{A^{-1}} X^T y - y) + 2\lambda \underbrace{(X^T X + \lambda I)^{-1}}_{A^{-1}} X^T y$$

$$= 2X^T (X A^{-1} X^T y - y) + 2\lambda A^{-1} X^T y$$

$$= 2X^T X A^{-1} X^T y - 2X^T y + 2\lambda A^{-1} X^T y$$

$$= (X^T X A^{-1} - I + \lambda A^{-1}) \, 2X^T y$$

$$= ((X^T X + \lambda I) A^{-1} - I) \, 2X^T y$$

$$= (A A^{-1} - I) \, 2X^T y$$

$$= (I - I) \, 2X^T y = 0$$

Rules:- $\vec{\nabla}_\theta \theta^T A \theta = 2A\theta$ $\quad \vec{\nabla}_\theta \theta^T x = x$

(b)

$$P(w|y) = \frac{P(y,w)}{P(y)} = \frac{P(y|w)P(w)}{P(y)}$$

$$\ln(P(w|y)) = \ln(P(y|w)) + \ln(P(w)) + \underbrace{\ln(P(y))}_{\text{constant}}$$

$$= \ln\left(\frac{1}{\sqrt{(2\pi)^D |\Sigma|}} e^{-\frac{1}{2}(y-Xw)^T(\sigma^2 I)^{-1}(y-Xw)}\right) +$$

$$\ln\left(\frac{1}{\sqrt{(2\pi)^D |\Sigma|}} e^{-\frac{1}{2}(w)^T(s^2 I)^{-1}(w)}\right) + \text{Constant}$$

$$= \text{Constant} + -\frac{1}{2}(y-Xw)^T(\sigma^2 I)^{-1}(y-Xw)$$

$$-\frac{1}{2} w^T (s^2 I)^{-1} w$$

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

$$-\frac{1}{2} w^T \left(\frac{1}{s^2} I\right) w$$

$$= -\frac{1}{2}\left(\frac{1}{\sigma^2}\right)(y^T - w^T X^T)(y - Xw) - \frac{1}{2}\frac{1}{s^2} w^T w$$

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

$$= -\frac{1}{2}\left(\frac{1}{\sigma^2}\right)(y^T y - 2 w^T X^T y + (w^T \underbrace{X^T X}_{A} w)) - \frac{1}{2}\left(\frac{1}{s^2}\right) w^T I w$$

$$\frac{d\ln(P(w|y))}{d(w)} = -\frac{1}{\cancel{2}}\left(\frac{1}{\sigma^2}\right)(0 - \cancel{2} X^T y + \cancel{2} X^T X w) - \frac{1}{\cancel{2}}\left(\frac{1}{s^2}\right)(\cancel{2} w)$$

$$\lambda = \frac{\sigma^2}{s^2}$$

$$\frac{d\ln(p(w|y))}{dw} = (-X^T y + X^T X w) + \frac{\sigma^2}{s^2} w) = 0$$

$$\Rightarrow X^T X w + \frac{\sigma^2}{s^2} w = X^T y$$

$$\Rightarrow (X^T X + \lambda I) w = X^T y$$

$$\Rightarrow \boxed{w = (X^T X + \lambda I)^{-1} X^T y}$$

(c) $w \sim \mathcal{N}(m_0, \Sigma_0)$ $\quad p(y|X, m_0, S_0) = \int p(y|w, X)\, p(w|m_0, S_0)\, dw$.

$$y|x, m_0, \Sigma_0 \sim \mathcal{N}(\mu, \Sigma)$$

$$y = f + \varepsilon \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2 I) \qquad f = Xw$$

$$E(y) = E(f) + E(\varepsilon) = X E(w) + 0 = X m_0$$

$$Cov(y) = Cov(f) + Cov(\varepsilon) = X\, cov(w)\, X^T + \sigma^2 I$$

$$Cov(y) = X s_0 X^T + \sigma^2 I$$

$$\boxed{y|x, m_0, \Sigma_0 \sim \mathcal{N}(X m_0,\ X s_0 X^T + \sigma^2 I)}$$

(d) $y_*|X_*, m_0, \Sigma_0 \sim \mathcal{N}(X_* m_0,\ X_* S_0 X_*^T + \sigma^2 I)$

(d) $p(y_*|y, X, x_*) = \int_{dw} p(y_*|w, x_*)\, p(w|y, X)$

$y_*|y, X, x_* \sim \mathcal{N}(f_*, \sigma_*)$.

$\alpha|\beta \sim \mathcal{N}(a + CB^{-1}(\beta - b),\ A - CB^{-1}C^T$

$$\begin{array}{cc} 1\times 1 & 1\times N \\ \begin{bmatrix} [\ ] & [\quad] \\ [\ ] & [\quad] \end{bmatrix} \\ N\times 1 & N\times N \end{array}$$

$\alpha = y_*$ $\beta = y$ $a = E(y_*)$ $b = E(y)$

$A = Cov(y_*)$ $B = Cov(y)$

$C = Corr(y_*, y)$

$y_* = f(x_*) + \varepsilon$

$y_* = x_* w + \varepsilon$

$y = f(x) + \varepsilon$

$= Xw + \varepsilon$

$a = X_*^T m_0 = E(y_*)$

$b = X m_0 = E(y)$

$A = x_*^T S_0 x + \sigma^2 I$

$B = X S_0 X^T + \sigma^2 I$

$C^T = X S_0 X_*$

$C = X_*^T S_0 X^T$

$x' = \begin{bmatrix} -X_*^T- \\ -X- \end{bmatrix}$ $\Sigma' = X' S_0 X'^T$

$\begin{bmatrix} y_* \\ y \end{bmatrix} \sim N\left(\begin{bmatrix} a \\ b \end{bmatrix}, \begin{bmatrix} A & C \\ C^T & B \end{bmatrix}\right)$

$\Sigma' = \begin{bmatrix} -X_*^T- \\ -X- \end{bmatrix} S_0 \begin{bmatrix} | & | \\ x_* & X \\ | & | \end{bmatrix}^T$

$$= \begin{array}{cc} 1\times 1 & 1\times N \\ \begin{bmatrix} A & C \\ C^T & B \end{bmatrix} \\ N\times 1 & N\times N \end{array} \Bigg\}(N+1)\times(N+1)$$

$\Rightarrow C = x_*^T S_0 X^T$

$I = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$

$y : \underset{N}{\begin{bmatrix} \\ \\ \end{bmatrix}}$

$$y_* | y \sim \mathcal{N}\left(a + CB^{-1}(\beta - b) \; ; \; A - CB^{-1}C^T \right)$$

$X : \overbrace{\begin{bmatrix} - x_*^T - \\ \\ - \end{bmatrix}}^{d}$ N

$$\Rightarrow y_* | y \sim \mathcal{N}\Big(\underbrace{\underbrace{x_*^T m_0}_{1\times1} + \underbrace{\underbrace{(x_*^T S_0 X^T)}_{1\times N} \underbrace{(X S_0 X^T + \sigma^2 I)^{-1}}_{N\times N} \underbrace{(\overbrace{y}^{N\times1} - \underbrace{X}_{N\times d} \underbrace{m_0}_{d\times1})}_{N\times1}}_{1\times1}} \; ,$$

$S_0 = d\times d$

$m_0 : d\times1$

$$\underbrace{\underset{1\times d}{x_*^T}\underset{d\times d}{S_0}\underset{d\times1}{x_*}}_{1\times1} - \underbrace{(x_*^T S_0 X^T)}_{1\times N} \underbrace{(X S_0 X^T + \sigma^2 I)}_{N\times N} \underbrace{(X S_0 x_*)}_{N\times1} \Big)$$

$$\boxed{\begin{aligned} y_* | y, X, x_* \sim \mathcal{N}\Big(& x_*^T m_0 + (x_*^T S_0 X^T)(X S_0 X^T + \sigma^2 I)^{-1}(y - X m_0), \\ & x_*^T S_0 x_* - (x_*^T S_0 X^T)(X S_0 X^T + \sigma^2 I)(X S_0 x_*) \Big) \end{aligned}}$$

(e) if $m_0 = 0$ $\quad S_0 = s^2 I$:-

$$y_* | y \sim \mathcal{N}\Big((x_*^T s^2 I X^T)(X s^2 I X^T + \sigma^2 I)^{-1}(y),$$
$$x_*^T s^2 I x_* - (x_*^T s^2 I X^T)(X s^2 I X^T + \sigma^2 I)(X s^2 I x_*) \Big)$$

$$\sim \mathcal{N}\Big((x_*^T s^2 X^T)(X s^2 X^T + \sigma^2 I)^{-1}(y),$$
$$x_*^T s^2 x_* - (x_*^T s^2 X^T)(X s^2 X^T + \sigma^2 I)(X s^2 x_*) \Big)$$

$$\boxed{f_* = E(x_*) = (x_*^T s^2 X^T)(X s^2 X^T + \sigma^2 I)^{-1} y}$$