EECS495

HW4

Liu Cao

Problem 1

Problem 1

$\because \quad Pr(\bar{y} \mid \tilde{X}, \tilde{w}) = Norm_{\bar{y}}[\tilde{X}^T\tilde{w}, s^2 I]$

We model the prior $Pr(\tilde{w})$ as normal with zero mean and spherical covariance.

$$Pr(\tilde{w}) = Norm_{\tilde{w}}[\bar{0}, \sigma_p^2 I]$$

Using Baye's theorem, the posterior distribution is:

$$Pr(\tilde{w} \mid \tilde{X}, \bar{y}) = \frac{Pr(\bar{y} \mid \tilde{X}, \tilde{w}) \cdot Pr(\tilde{w})}{Pr(\bar{y} \mid \tilde{X})}$$

We can simplify the Bayesian probability, we don't need to care about the terms unrelated to $\tilde{w}$.

$$\therefore Pr(\tilde{w} \mid \tilde{X}, \bar{y}) \propto Pr(\bar{y} \mid \tilde{X}, \tilde{w}) \cdot Pr(\tilde{w})$$

$$= \exp\left(\frac{-1}{2\delta^2}(\bar{y} - \tilde{X}^T\tilde{w})^T(\bar{y} - \tilde{X}^T\tilde{w})\right)\exp\left(-\frac{1}{2}\tilde{w}^T\Sigma_p^{-1}\tilde{w}\right)$$

$$= C\exp\left(\frac{1}{2\delta^2}\tilde{w}^T\tilde{X}\bar{y} + \frac{1}{2\delta^2}\bar{y}^T\tilde{X}^T\tilde{w} - \frac{1}{2\delta^2}\tilde{w}^T\tilde{X}\tilde{X}^T\tilde{w} - \frac{1}{2}\tilde{w}^T\Sigma_p^{-1}\tilde{w}\right)$$

$$\propto \exp\left(\frac{1}{2\delta^2}\tilde{w}^T\tilde{X}\bar{y} + \frac{1}{2\delta^2}\bar{y}^T\tilde{X}^T\tilde{w} - \frac{1}{2\delta^2}\tilde{w}^T\tilde{X}\tilde{X}^T\tilde{w} - \frac{1}{2}\tilde{w}^T\Sigma_p^{-1}\tilde{w}\right)$$

$\because \tilde{w}^T\tilde{X}\bar{y} = \bar{y}^T\tilde{X}^T\tilde{w}$ (a scalar) $\Rightarrow$

$$= \exp\left(\frac{1}{\delta^2}\bar{y}^T\tilde{X}^T\tilde{w} - \frac{1}{2\delta^2}\tilde{w}^T\tilde{X}\tilde{X}^T\tilde{w} - \frac{1}{2}\tilde{w}^T\Sigma_p^{-1}\tilde{w}\right)$$

$$= \exp\left(\frac{1}{\delta^2}\bar{y}^T\tilde{X}^T\tilde{w} - \frac{1}{2}\tilde{w}^T\left(\frac{1}{\delta^2}\tilde{X}\tilde{X}^T + \Sigma_p^{-1}\right)\tilde{w}\right)$$

We need ~~to~~ write the function in a quadratic form: $-\frac{1}{2}(\tilde{w}-\bar{w})^T A(\tilde{w}-\bar{w})$

By observation, we can obtain $A = \frac{1}{\delta^2}\tilde{X}\tilde{X}^T + \Sigma_p^{-1}$

$$\frac{1}{\delta^2}\bar{y}^T\tilde{X}^T\tilde{w} = -\frac{1}{2}(\tilde{w}-\bar{w})^T A(\tilde{w}-\bar{w}) - \left(-\frac{1}{2}\tilde{w}^T A\tilde{w}\right) - \left(-\frac{1}{2}\bar{w}^T A\bar{w}\right)$$

$$= \frac{1}{2}(\bar{w}^T A\tilde{w} + \tilde{w}^T A\bar{w}) = \bar{w}^T A\tilde{w} \quad (\bar{w}^T A\tilde{w} \text{ is a scalar})$$

$\therefore \quad \frac{1}{\delta^2}\bar{y}^T\tilde{X}^T = \bar{w}^T A \qquad \therefore \quad \bar{w} = \left(\frac{1}{\delta^2}\bar{y}^T\tilde{X}^T A^{-1}\right)^T = \frac{1}{\delta^2}A^{-1}\tilde{X}\bar{y}$

$\because$ A is symmetric

$$\therefore Pr(\tilde{w}|\tilde{X},\tilde{y}) \propto \exp\left(\frac{1}{\delta^2}\tilde{y}^T\tilde{X}^T\tilde{w} - \frac{1}{2}\tilde{w}^T\left(\frac{1}{\delta^2}\tilde{X}\tilde{X}^T + \Sigma_p^{-1}\right)\tilde{w}\right)$$

$$= \exp\left(-\frac{1}{2}(\tilde{w}-\bar{w})^T A(\tilde{w}-\bar{w})\right)\exp\left(\frac{1}{2}\bar{w}Q\bar{w}\right)$$

$$\propto \exp\left(-\frac{1}{2}(\tilde{w}-\bar{w})^T A(\tilde{w}-\bar{w})\right)$$

$$\therefore Pr[\tilde{w}|\tilde{X},y] = Norm_{\tilde{w}}[\bar{w}, A^{-1}]$$

$$= Norm_{\tilde{w}}[\delta^{-2}A^{-1}\tilde{X}\tilde{y},\ A^{-1}]$$

where $A = \delta^{-2}\tilde{X}\tilde{X}^T + \Sigma_p^{-1}$

$$= \delta^{-2}\tilde{X}\tilde{X}^T + \delta_p^{-2}I.$$

Problem 2

Problem 2.

① We can compute the mean and covariance of $Norm_{y^*}$ respectively.

$$\because \quad Pr(y^*|\tilde{x}^*, \tilde{w}) = Norm_{y^*}[\tilde{x}^{*T}\tilde{w}, \sigma^2] \quad (1)$$

$$Pr(\tilde{w}|\tilde{X}, \tilde{y}) = Norm_{\tilde{w}}[\sigma^{-2}A^{-1}\tilde{X}\tilde{y}, A^{-1}] \quad (2)$$

$\therefore$ For a new data

$y^* = \tilde{x}^{*T} \cdot \tilde{w} + b$ (According to (1)), where b is deviation

And $\tilde{x}^{*T}\tilde{w}$ is independent of b, $b = Norm[0, \sigma^2]$

$\therefore$ first moment: $E[y^*] = E[\tilde{x}^{*T}\tilde{w} + b] = \tilde{x}^{*T}E(\tilde{w}) = \tilde{x}^{*T}\sigma^{-2}A^{-1}\tilde{X}\tilde{y}$

$\because$ second moment: $E[y^{*2}] = E[\tilde{x}^{*T}\tilde{w} \cdot \tilde{w}^T\tilde{x}^*] + E[b \cdot b^T]$

$\because$ ($\tilde{x}^{*T}\tilde{w} = \tilde{w}^T\tilde{x}^*$ Both are scalar) $\quad = \tilde{x}^{*T}E[\tilde{w} \cdot \tilde{w}^T]\tilde{x}^* + \sigma^2$

$$\because \quad E[\tilde{w}\tilde{w}^T] - E[\tilde{w}]E[\tilde{w}^T] = A^{-1} \quad \text{(According to (2))}$$

$$\therefore \quad E[\tilde{w}\tilde{w}^T] = E[\tilde{w}]E[\tilde{w}^T] + A^{-1} \quad \text{(According to (2))}$$

$$= \sigma^{-4}A^{-1}\tilde{X}\tilde{y}\tilde{y}^T\tilde{X}^TA^{-1} + A^{-1} \quad (A \text{ is symmetric})$$

$$\therefore \quad E(y^{*2}) = \sigma^{-4}\tilde{x}^{*T}A^{-1}\tilde{X}\tilde{y}\tilde{y}^T\tilde{X}^TA^{-1}\tilde{x}^{*T} + \tilde{x}^{*T}A^{-1}\tilde{x}^* + \sigma^2$$

$\therefore$ The covariance should be:

$$E[y^{*2}] - E[y^*]^2 = \sigma^{-4}\tilde{x}^{*T}A^{-1}\tilde{X}\tilde{y}\tilde{y}^T\tilde{X}^TA^{-1}\tilde{x}^* + \tilde{x}^{*T}A^{-1}\tilde{x}^* + \sigma^2 - \sigma^{-4}\tilde{x}^{*T}A^{-1}\tilde{X}\tilde{y}y^T\tilde{X}^TA^{-1}\tilde{x}^*$$

$$= \tilde{x}^{*T}A^{-1}\tilde{x}^* + \sigma^2.$$

$$\therefore \quad Pr(y^*|\tilde{x}^*, \tilde{X}, \tilde{y}) = Norm_{y^*}[\sigma^{-2}\tilde{x}^{*T}A^{-1}\tilde{X}\tilde{y}, \tilde{x}^{*T}A^{-1}\tilde{x}^* + \sigma^2].$$

② According to the lecture:

$$\tilde{x}^* \longrightarrow Pr(y^*|\tilde{x}^*, \tilde{X}, \tilde{y})$$

$$= \int Pr(y^* \mid \bar{x}^*, \tilde{w}) \; Pr(\tilde{w} \mid \tilde{x}, \bar{y}) \, d\tilde{w}$$

$$= \int N_{y^*}(\bar{x}^{*T}\tilde{w}, \sigma^2) \, N_{\tilde{w}}\left(\frac{1}{\sigma^2} A^{-1} \tilde{x} \bar{y}, A^{-1}\right) d\tilde{w}$$

$$A = \frac{1}{\sigma^2} \tilde{x}\tilde{x}^T + \frac{1}{\sigma_p^2} I$$

mapping: $\because$

$$p(x) = N(x \mid u, \Lambda^{-1})$$

$$p(y \mid x) = N(y \mid Bx + b, L^{-1})$$

$$p(y) = N(y \mid Bu + b, L^{-1} + B\Lambda^{-1}B^T)$$

$\therefore$ $x = \tilde{w}$,

$$B = \bar{x}^{*T}$$

$$u = \frac{1}{\sigma^2} A^{-1} \tilde{x} \bar{y}$$

$$b = 0$$

$$L^{-1} = \sigma^2$$

$$\Lambda^{-1} = A^{-1}$$

$\therefore$ $Bu + b = \frac{1}{\sigma^2} \bar{x}^{*T} A^{-1} \tilde{x} \bar{y}$

$$L^{-1} B \Lambda^{-1} B^T = x^{*T} B^{-1} \bar{x}^* + \sigma^2$$

$\therefore$ $Pr(y^* \mid \bar{x}^*, \tilde{x}, \bar{y}) = N_{y^*}\left[\sigma^{-2} \bar{x}^{*T} A^{-1} \tilde{x} \bar{y}, \; \bar{x}^{*T} A^{-1} \bar{x}^* + \sigma^2\right]$

Problem 3& Problem 4

Problem 3.

$$\because \nabla_{\tilde{w}} L = \sum_{i=1}^{P} y_i \frac{1}{sig(a_i)} \cdot sig(a_i)(1-sig(a_i))\bar{X}_i$$

$$+\sum_{i=1}^{P}(1-y_i)\cdot\frac{-1}{1-sig(a_i)} sig(a_i)(1-sig(a_i))\bar{X}_i$$

$$= \sum_{i=1}^{P} y_i(1-sig(a_i))\bar{X}_i - \sum_{i=1}^{P}(1-y_i)sig(a_i)\bar{X}_i$$

$$= \sum_{i=1}^{P} y_i \bar{X}_i - \sum_{i=1}^{P} sig(a_i)\bar{X}_i$$

$$= -\sum_{i=1}^{P}(sig(a_i)-y_i)\bar{X}_i$$

Problem 4

According to the resutt of Problem 3.

$$\frac{\partial L}{\partial w_m} = -\sum_{i=1}^{P}\left(sig\left(b+\sum_{q=1}^{N} X_{qi}w_q\right)-y_i\right)X_{mi}$$

$$\therefore \frac{\partial L}{\partial w_m \partial w_n} = -\sum_{i=1}^{P} X_{mi}\cdot sig\left(b+\sum_{q=1}^{N}X_{qi}w_q\right)\left(1-sig\left(b+\sum_{q=1}^{N}X_{qi}w_q\right)\right)X_{ni}$$

$$= -\sum_{i=1}^{P} sig(a_i)(1-sig(a_i))X_{mi}X_{ni}$$

$$= -\sum_{i=1}^{P} sig(a_i)(1-sig(a_i))\bar{X}_i\bar{X}_i^T$$

Problem 5

Problem 5.

$\because Pr(\tilde{w} \mid y_{p+1}, x_{p+1}, \tilde{X}_p, \tilde{y}_p)$

$$= \frac{Pr(\bar{y}_{p+1} \mid \tilde{x}_{p+1}, \tilde{w}) \cdot Pr(\tilde{w} \mid \tilde{X}_p, \tilde{y}_p)}{Pr(\bar{y}_{p+1} \mid \bar{x}_{p+1})}$$

$$= \frac{N_{\bar{y}_{p+1}}(\tilde{x}_{p+1}^T \tilde{w}, \delta_{p+1}^2) \cdot N_{\tilde{w}}(\delta^{-2} A_p^{-1} \tilde{X}_p \bar{y}_p, A_p^{-1})}{N_{y+1}(\bar{x}_{p+1}^T w, \delta_{p+1}^{-2})}$$ $\rightarrow$ this term can be neglected.

$$\propto \exp\left(-\frac{1}{2}(\tilde{w} - \delta^{-2} A_p^{-1} \tilde{X}_{p+1} \bar{y}_{p+1})^T \delta_{p+1}^2 (\tilde{w} - \delta_p^{-2} A_p^{-1} \tilde{x}_{p+1} \bar{y}_{p+1}) + C\right)$$

where C is a constant.

$$\propto \exp\left(\tilde{w}^T(-\frac{1}{2}\delta_{p+1}^2)\tilde{w} + \tilde{w}^T \delta_p^{-2} \delta_{p+1}^2 A_p^{-1} \tilde{x}_{p+1} \bar{y}_{p+1} + C\right)$$

$$\propto \exp(a(\tilde{w}) + C)$$

$$\nabla_{\tilde{w}} a = -\tilde{w}\, \delta_{p+1}^2 + \delta_{p+1}^2 \delta_p^{-2} A_p^{-1} \tilde{x}_{p+1} \bar{y}_{p+1}$$

$$\nabla_{\tilde{w}} a = 0 \Rightarrow \tilde{w} = \delta_p^{-2} A_p^{-1} \tilde{x}_{p+1} \bar{y}_{p+1}$$

which is the mean.

$\therefore$ The resulting posterior distribution is the same by replacing p by $p+1$