EE 475 HW5

Name : Zhisheng Lin
student #: 3226299

Problem 1:

from likelihood: $P(\bar{y}|\tilde{X},\tilde{w}) = N_{\bar{y}}(\tilde{X}^T\tilde{w}, \sigma^2 I)$

$P(\tilde{w}) = N_{\tilde{w}}(0, \sigma_p^2 I)$

$$P(\tilde{w}|\tilde{X},\bar{y}) \propto P(\bar{y}|\tilde{X},\tilde{w}) \cdot P(\tilde{w}) = N_{\bar{y}}(\tilde{X}^T\tilde{w}, \sigma^2 I) \cdot N_{\tilde{w}}(0, \sigma_p^2 I)$$

$$\propto \exp\{-\tfrac{1}{2}(\bar{y}-\tilde{X}^T\tilde{w})^T\sigma^{-2}I(\bar{y}-\tilde{X}^T\tilde{w})\} \cdot \exp\{-\tfrac{1}{2}\tilde{w}^T\sigma_p^{-2}I\tilde{w}\}$$

$$= \exp\{-\tfrac{1}{2\sigma^2}(\bar{y}^T-\tilde{w}^T\tilde{X})(\bar{y}-\tilde{X}^T\tilde{w}) - \tfrac{1}{2}\tilde{w}^T\sigma_p^{-2}I\tilde{w}\}$$

$$= \exp\{-\tfrac{1}{2\sigma^2}(\bar{y}^T\bar{y} - \bar{y}^T\tilde{X}^T\tilde{w} - \tilde{w}^T\tilde{X}\bar{y} + \tilde{w}^T\tilde{X}\tilde{X}^T\tilde{w}) - \tfrac{1}{2}\tilde{w}^T\sigma_p^{-2}I\tilde{w}\}$$

$$= \exp\{-\tfrac{1}{2\sigma^2}(\bar{y}^T\bar{y} - 2\bar{y}^T\tilde{X}^T\tilde{w} + \tilde{w}^T\tilde{X}\tilde{X}^T\tilde{w}) - \tfrac{1}{2}\tilde{w}^T\sigma_p^{-2}I\tilde{w}\} \quad \text{---- (a)}$$

Assume $P(\tilde{w}|\tilde{X},\bar{y}) = N(\mu_{\tilde{w}}, \Sigma_{\tilde{w}})$. then its exponential part is $\exp\{-\tfrac{1}{2}(\tilde{w}-\mu_{\tilde{w}})^T\Sigma_{\tilde{w}}^{-1}(\tilde{w}-\mu_{\tilde{w}})\}$

$$\Rightarrow \exp\{-\tfrac{1}{2}(\underbrace{\tilde{w}^T\Sigma_{\tilde{w}}^{-1}\tilde{w}}_{\text{quadratic term}} - \underbrace{2\mu_{\tilde{w}}^T\Sigma_{\tilde{w}}^{-1}\tilde{w}}_{\text{Linear term}} + \mu^T\Sigma^{-1}\mu)\}$$

So, we can find from the exponential part of the standard Gussian distribution that

① we can get $\Sigma_{\tilde{w}}$ from quadratic term

② we can get $\mu_{\tilde{w}}$ from linear term

Then, from the formula (a), we can get quadratic term:

$$-\tfrac{1}{2}(\tilde{w}^T(\sigma^{-2}\tilde{X}\tilde{X}^T + \sigma_p^{-2}I)\tilde{w}) \Rightarrow \Sigma_w = (\sigma^{-2}\tilde{X}\tilde{X}^T + \sigma_p^{-2}I)^{-1} = A^{-1}$$

linear term:

$$\sigma^{-2}\bar{y}^T\tilde{X}^T = \mu_{\tilde{w}}^T\Sigma_{\tilde{w}}^{-1} = \mu_{\tilde{w}}^T \cdot A \xrightarrow{\text{transpose}} A\mu_{\tilde{w}} = \sigma^{-2}\tilde{X}\bar{y} \Rightarrow \mu_{\tilde{w}} = \sigma^{-2}A^{-1}\tilde{X}\bar{y}$$

$$\boxed{\text{Finally, We show that } Pr(\tilde{w}|\tilde{X},\bar{y}) = Norm_{\tilde{w}}[\sigma^{-2}A^{-1}\tilde{X}\bar{y}, A^{-1}] \quad \text{where } A = \sigma^{-2}\tilde{X}\tilde{X}^T + \sigma_p^{-2}I}$$

Problem 2:

$$Pr(y^* | \bar{X}^*, \tilde{X}, \bar{y}) = \int P(y^* | \bar{x}^*, \tilde{w}) \cdot P(\tilde{w} | \tilde{X}, \bar{y}) \, dw$$

$$\because P(y^* | \bar{X}^*, \tilde{w}) = N_{y^*} (\bar{X}^{*T} \cdot \tilde{W}, \sigma^2) \quad \text{and}$$

$$P(\tilde{w} | \tilde{X}, \bar{y}) = N_{\tilde{w}} (\sigma^{-2} A^{-1} \tilde{X} \bar{y}, A^{-1}), \text{where } A = \sigma^{-2} \tilde{X} \tilde{X}^T + \sigma_p^{-2} I$$

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

where $A = \sigma^{-2} \tilde{X} \tilde{X}^T + \sigma_p^{-2} I$

Problem 3:

$$L = \sum_{i=1}^{P} y_i \log(\text{sig}(a_i)) + \sum_{i=1}^{P} (1 - y_i) \log(1 - \text{sig}(a_i))$$

$$= \sum_{i=1}^{P} \left[y_i \log(\text{sig}(a_i)) - y_i \log(1 - \text{sig}(a_i)) \right] + \sum_{i=1}^{P} \log(1 - \text{sig}(a_i))$$

$$= \sum_{i=1}^{P} y_i \cdot \log \frac{\text{sig}(a_i)}{1 - \text{sig}(a_i)} + \sum_{i=1}^{P} \log(1 - \text{sig}(a_i))$$

$$= \sum_{i=1}^{P} y_i \cdot \log e^{X_i^T \tilde{w}} + \sum_{i=1}^{P} \left[\log e^{-X_i^T \tilde{w}} - \log(1 + e^{-X_i^T \tilde{w}}) \right]$$

$$= \sum_{i=1}^{P} y_i \cdot X_i^T \tilde{w} - \sum_{i=1}^{P} X_i^T \tilde{w} - \sum_{i=1}^{P} \log(1 + e^{-X_i^T \tilde{w}})$$

Then:

$$\nabla_{\tilde{w}} L = \sum_{i=1}^{P} y_i \bar{X}_i - \sum_{i=1}^{P} \bar{X}_i + \sum_{i=1}^{P} \frac{e^{-X_i^T \tilde{w}}}{1 + e^{-X_i^T \tilde{w}}} \cdot \bar{X}_i$$

$$= \sum_{i=1}^{P} y_i \bar{X}_i - \sum_{i=1}^{P} \left(1 - \frac{e^{-X_i^T \tilde{w}}}{1 + e^{-X_i^T \tilde{w}}} \right) \bar{X}_i$$

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

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

So, it's been proved.

Problem 4:

$$\because \nabla_{\tilde{w}} L = -\sum_{i=1}^{p} (\mathrm{sig}(a_i) - y_i)\, \bar{x}_i$$

$$\therefore \nabla_{\tilde{w}}^2 L = -\sum_{i=1}^{p} \frac{e^{-x_i^T \tilde{w}}}{(1+e^{-x_i^T \tilde{w}})^2} \cdot \bar{x}_i \cdot \bar{x}_i^T$$

$$= -\sum_{i=1}^{p} \frac{1}{1+e^{-x_i^T \tilde{w}}} \cdot \left(1 - \frac{1}{1+e^{-x_i^T \tilde{w}}}\right) \cdot \bar{x}_i\, \bar{x}_i^T$$

$$= -\sum^{p} (\mathrm{sig}(a_i))(1 - \mathrm{sig}(a_i))\, \bar{x}_i \bar{x}_i^T$$

So, it's been proved.