1 Problem 1

The posterior distribution is given by:

$$Pr(\tilde{\mathbf{w}}|\tilde{\mathbf{X}}, \bar{\mathbf{y}}) = \frac{Pr(\bar{\mathbf{y}}|\tilde{\mathbf{X}}, \tilde{\mathbf{w}})Pr(\tilde{\mathbf{w}})}{Pr(\bar{\mathbf{y}}|\tilde{\mathbf{X}})}. \tag{1}$$

We focus on the numerator since the denominator is a constant for $\tilde{\mathbf{w}}$. The prior distribution for $\tilde{\mathbf{w}}$ we assume is:

$$\begin{aligned} Pr(\tilde{\mathbf{w}}) &= \text{Norm}_{\tilde{\mathbf{w}}}[0, \sigma_p^2 I] \\ &= C_1 * \exp(-\frac{1}{2}\tilde{\mathbf{w}}^T \sigma_p^{-2} I \tilde{\mathbf{w}}), \end{aligned}$$

where C_1 is a constant of $\tilde{\mathbf{w}}$, and

$$\begin{aligned} Pr(\bar{\mathbf{y}}|\tilde{\mathbf{X}}, \tilde{\mathbf{w}}) &= \text{Norm}_{\bar{\mathbf{y}}}[\tilde{\mathbf{X}}^T \tilde{\mathbf{w}}, \sigma^2 I] \\ &= C_2 * \exp(-\frac{1}{2}(\bar{\mathbf{y}} - \tilde{\mathbf{X}}^T \tilde{\mathbf{w}})^T \sigma^{-2} I (\bar{\mathbf{y}} - \tilde{\mathbf{X}}^T \tilde{\mathbf{w}})), \end{aligned}$$

where C_2 is a constant. Hence:

$$\begin{aligned} Pr(\bar{\mathbf{y}}|\tilde{\mathbf{X}}, \tilde{\mathbf{w}})Pr(\tilde{\mathbf{w}}) &= \text{Constant} * \exp(-\frac{1}{2\sigma^2}(\bar{\mathbf{y}} - \tilde{\mathbf{X}}^T \tilde{\mathbf{w}})^T (\bar{\mathbf{y}} - \tilde{\mathbf{X}}^T \tilde{\mathbf{w}})) \exp(-\frac{1}{2}\tilde{\mathbf{w}}^T \sigma_p^{-2} I \tilde{\mathbf{w}}) \\ &= \text{Constant} * \exp(\frac{1}{2\sigma^2}\tilde{\mathbf{w}}^T \tilde{\mathbf{X}} \bar{\mathbf{y}} + \frac{1}{2\sigma^2}\bar{\mathbf{y}}^T \tilde{\mathbf{X}}^T \tilde{\mathbf{w}} - \frac{1}{2\sigma^2}\tilde{\mathbf{w}}^T \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T \tilde{\mathbf{w}} - \frac{1}{2}\tilde{\mathbf{w}}^T \sigma_p^{-2} \tilde{\mathbf{w}}) \\ &= \text{Constant} * \exp(\frac{1}{\sigma^2}\bar{\mathbf{y}}^T \tilde{\mathbf{X}}^T \tilde{\mathbf{w}} - \frac{1}{2\sigma^2}\tilde{\mathbf{w}}^T \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T \tilde{\mathbf{w}} - \frac{1}{2}\tilde{\mathbf{w}}^T \sigma_p^{-2} \tilde{\mathbf{w}}) \\ &= \text{Constant} * \exp(\frac{1}{\sigma^2}\bar{\mathbf{y}}^T \tilde{\mathbf{X}}^T \tilde{\mathbf{w}} - \frac{1}{2}\tilde{\mathbf{w}}^T (\frac{1}{\sigma^2}\tilde{\mathbf{X}} \tilde{\mathbf{X}}^T + \sigma_p^{-2} I)\tilde{\mathbf{w}}) \\ &= \text{Constant} * \exp(-\frac{1}{2}(\tilde{\mathbf{w}} - \bar{\mathbf{a}})^T \mathbf{A} (\tilde{\mathbf{w}} - \bar{\mathbf{a}})), \end{aligned}$$

where $\mathbf{A} = \sigma^{-2}\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T + \sigma_p^{-2} I$ and $\bar{\mathbf{a}} = \sigma^{-2}\mathbf{A}^{-1}\tilde{\mathbf{X}}\bar{\mathbf{y}}$, which indicates:

$$Pr(\tilde{\mathbf{w}}|\tilde{\mathbf{X}}, \bar{\mathbf{y}}) = \text{Norm}_{\tilde{\mathbf{w}}}[\sigma^{-2}\mathbf{A}^{-1}\tilde{\mathbf{X}}\bar{\mathbf{y}}, \mathbf{A}^{-1}]. \tag{2}$$

OTHER SOLUTION

Using the equations provided in the supplementary material, we identify $Pr(\mathbf{y}|\mathbf{x})$ as the likelihood $Pr(\bar{\mathbf{y}}|\tilde{\mathbf{X}}, \tilde{\mathbf{w}})$, and $Pr(\mathbf{x})$ as the prior distribution $Pr(\tilde{\mathbf{w}})$. Then the posterior distribution $Pr(\tilde{\mathbf{w}}|\tilde{\mathbf{X}}, \bar{\mathbf{y}})$ is calculated as $Pr(\mathbf{x}|\mathbf{y})$ in the supplementary material.

2 Problem 2

We use the equations in the supplement material, since

$$Pr(\tilde{\mathbf{w}}|\tilde{\mathbf{X}}, \bar{\mathbf{y}}) = \text{Norm}_{\tilde{\mathbf{w}}}[\sigma^{-2}\mathbf{A}^{-1}\tilde{\mathbf{X}}\bar{\mathbf{y}}, \mathbf{A}^{-1}] \tag{3}$$

$$Pr(y^*|\tilde{\mathbf{w}}, \bar{\mathbf{x}}^*, \tilde{\mathbf{X}}, \bar{\mathbf{y}}) = Pr(y^*|\tilde{\mathbf{w}}, \bar{\mathbf{x}}^*) = \text{Norm}_{y^*}[\tilde{\mathbf{x}}^{*T}\tilde{\mathbf{w}}, \sigma^2]. \tag{4}$$

According to the equations provided, Let $x \leftarrow \tilde{\mathbf{w}}|\tilde{\mathbf{X}}, \bar{\mathbf{y}}$, $y \leftarrow y^*|\bar{\mathbf{x}}^*, \tilde{\mathbf{X}}, \bar{\mathbf{y}}$, then correspondingly we can have:

$$Pr(y^*|\bar{\mathbf{x}}^*, \tilde{\mathbf{X}}, \bar{\mathbf{y}}) = \text{Norm}_{y^*}[\sigma^{-2}\tilde{\mathbf{x}}^{*T}\mathbf{A}^{-1}\tilde{\mathbf{X}}\bar{\mathbf{y}}, \tilde{\mathbf{x}}^{*T}\mathbf{A}^{-1}\tilde{\mathbf{x}}^* + \sigma^2] \tag{5}$$

3 Problem 3

Let $f_i = y_i \log(sig(a_i))$ and $g_i = (1 - y_i)\log(1 - sig(a_i))$, then

$$\nabla_{\tilde{\mathbf{w}}} L = \sum_{i=1}^{P} \left(\nabla_{\tilde{\mathbf{w}}} f_i + \nabla_{\tilde{\mathbf{w}}} g_i\right). \tag{6}$$

Now since $a_i = \bar{\mathbf{x}}_{\mathbf{i}}^T\tilde{\mathbf{w}}$, we have

$$\begin{aligned}\nabla_{\tilde{\mathbf{w}}} f_i &= \frac{y_i}{sig(a_i)} \cdot sig(a_i)(1 - sig(a_i)) \cdot \nabla_{\tilde{\mathbf{w}}} a_i \quad \text{(Chain Rule)}\\ &= (y_i - y_i sig(a_i))\bar{\mathbf{x}}_i,\end{aligned}$$

and

$$\begin{aligned}\nabla_{\tilde{\mathbf{w}}} g_i &= \frac{1 - y_i}{1 - sig(a_i)} \cdot -sig(a_i)(1 - sig(a_i)) \cdot \nabla_{\tilde{\mathbf{w}}} a_i \quad \text{(Chain Rule)}\\ &= (y_i sig(a_i) - sig(a_i))\bar{\mathbf{x}}_i.\end{aligned}$$

So $\nabla_{\tilde{\mathbf{w}}} L = -\sum_{i=1}^{P}(sig(a_i) - y_i)\bar{\mathbf{x}}_i$.

4 Problem 4

Use $\mathbf{g}$ to denote the gradient $\nabla_{\tilde{\mathbf{w}}} L$. Then the (m, n)-th element of $\nabla^2_{\tilde{\mathbf{w}}} L$ is

$$\begin{aligned}(\nabla^2_{\tilde{\mathbf{w}}} L)_{m,n} &= \frac{\partial L}{\partial \tilde{w}_m \partial \tilde{w}_n}\\ &= \frac{\partial g_m}{\partial \tilde{w}_n}\\ &= \frac{\partial - \sum_{i=1}^{P}(sig(\bar{\mathbf{x}}_i^T\tilde{\mathbf{w}}) - y_i)\bar{x}_{i,m}}{\partial \tilde{w}_n}\\ &= \frac{\partial - \sum_{i=1}^{P} sig(\bar{\mathbf{x}}_i^T\tilde{\mathbf{w}})\bar{x}_{i,m}}{\partial \tilde{w}_n}\\ &= -\sum_{i=1}^{P} sig(a_i)(1 - sig(a_i))\bar{x}_{i,n}\bar{x}_{i,m}.\end{aligned}$$

where g_m is the m-th element of $\mathbf{g}$.

Let $\tilde{\mathbf{X}}_i = \bar{\mathbf{x}}_i\bar{\mathbf{x}}_i^T$, then obviously, the (m,n)-th element of $\tilde{\mathbf{X}}_i$ is $\tilde{X}_{i,m,n} = \bar{x}_{i,m}\bar{x}_{i,n} = \bar{x}_{i,n}\bar{x}_{i,m}$, which is the last term in above equation. So,

$$\nabla^2_{\tilde{\mathbf{w}}} L = -\sum_{i=1}^{P} sig(a_i)(1-sig(a_i))\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i^T.$$

5 Problem 5

Combining the prior

$$p(\tilde{\mathbf{w}}) = Norm_{\tilde{\mathbf{w}}_P}[\sigma^{-2}\mathbf{A}_P^{-1}\tilde{\mathbf{X}}_P\bar{\mathbf{y}}_P, \mathbf{A}_P^{-1}]$$

and for notation simplicity, we define $\mathbf{m}_P = \sigma^{-2}\mathbf{A}_P^{-1}\tilde{\mathbf{X}}_P\bar{\mathbf{y}}_P$ and $\mathbf{S}_P = \mathbf{A}_P^{-1}$.

The likelihood is

$$\begin{aligned} p(\bar{y}_{P+1}|\bar{\mathbf{x}}_{P+1}, \tilde{\mathbf{w}}) &= Norm_{\bar{y}_{P+1}}[\tilde{\mathbf{w}}^T\bar{\mathbf{x}}_{P+1}, \sigma^2] \\ &= \left(\frac{\sigma^{-2}}{2\pi}\right)^{1/2} \exp\left(-\frac{\sigma^{-2}}{2}(\bar{y}_{P+1} - \tilde{\mathbf{w}}^T\bar{\mathbf{x}}_{P+1})^2\right). \end{aligned}$$

Then, we obtain a posterior of the form

$$\begin{aligned} &p(\tilde{\mathbf{w}}|\bar{y}_{P+1}, \bar{\mathbf{x}}_{P+1}, \bar{\mathbf{y}}_P, \tilde{\mathbf{X}}_P) \\ &\propto \exp\left(-\frac{1}{2}(\tilde{\mathbf{w}} - \mathbf{m}_P)^T\mathbf{S}_P^{-1}(\tilde{\mathbf{w}} - \mathbf{m}_P) - \frac{\sigma^{-2}}{2}(\bar{y}_{P+1} - \tilde{\mathbf{w}}^T\bar{\mathbf{x}}_{P+1})^2\right) \end{aligned}$$

We can expand the argument of the exponential, omitting the -1/2 factors, as follows

$$\begin{aligned} &(\tilde{\mathbf{w}} - \mathbf{m}_P)^T\mathbf{S}_P^{-1}(\tilde{\mathbf{w}} - \mathbf{m}_P) + \sigma^{-2}(\bar{y}_{P+1} - \tilde{\mathbf{w}}^T\bar{\mathbf{x}}_{P+1}) \\ &= \tilde{\mathbf{w}}^T\mathbf{S}_P^{-1}\tilde{\mathbf{w}} - 2\tilde{\mathbf{w}}^T\mathbf{S}_P^{-1}\mathbf{m}_P + \sigma^{-2}\tilde{\mathbf{w}}^T\bar{\mathbf{x}}_{P+1}\bar{\mathbf{x}}_{P+1}^T\tilde{\mathbf{w}} - 2\sigma^{-2}\tilde{\mathbf{w}}^T\bar{\mathbf{x}}_{P+1}\bar{y}_{P+1} + \text{const} \\ &= \tilde{\mathbf{w}}^T\underbrace{(\mathbf{S}_P^{-1} + \sigma^{-2}\bar{\mathbf{x}}_{P+1}\bar{\mathbf{x}}_{P+1}^T)}_{\mathbf{S}_{P+1}^{-1}}\tilde{\mathbf{w}} - 2\tilde{\mathbf{w}}^T\underbrace{(\mathbf{S}_P^{-1}\mathbf{m}_P + \sigma^{-2}\bar{\mathbf{x}}_{P+1}\bar{y}_{P+1})}_{\mathbf{S}_{P+1}^{-1}\mathbf{m}_{P+1}} + \text{const} \end{aligned}$$

where const denotes remaining terms independent of $\tilde{\mathbf{w}}$.

From this we can read off the desired result directly,

$$p(\tilde{\mathbf{w}}|\bar{y}_{P+1}, \bar{\mathbf{x}}_{P+1}, \bar{\mathbf{y}}_P, \tilde{\mathbf{X}}_P) = Norm_{\tilde{\mathbf{w}}}[\mathbf{m}_{P+1}, \mathbf{S}_{P+1}],$$

where $\mathbf{S}_{P+1} = \mathbf{A}_{P+1}^{-1}$ and $\mathbf{m}_{P+1} = \sigma^{-2}\mathbf{A}_{P+1}^{-1}\tilde{\mathbf{X}}_{P+1}\bar{\mathbf{y}}_{P+1}$.

We can verify that

$$\begin{aligned} &\mathbf{S}_P^{-1} + \sigma^{-2}\bar{\mathbf{x}}_{P+1}\bar{\mathbf{x}}_{P+1}^T \\ &= \mathbf{A}_P + \sigma^{-2}\bar{\mathbf{x}}_{P+1}\bar{\mathbf{x}}_{P+1}^T \\ &= \sigma^{-2}\tilde{\mathbf{X}}_P\tilde{\mathbf{X}}_P^T + \sigma_p^{-2}\mathbf{I} + \sigma^{-2}\bar{\mathbf{x}}_{P+1}\bar{\mathbf{x}}_{P+1}^T \\ &= \sigma^{-2}\tilde{\mathbf{X}}_{P+1}\tilde{\mathbf{X}}_{P+1}^T + \sigma_p^{-2}\mathbf{I} \\ &= \mathbf{A}_{P+1} \\ &= \mathbf{S}_{P+1}^{-1} \end{aligned}$$

and

$$
\begin{aligned}
&\mathbf{S}_P^{-1}\mathbf{m}_P + \sigma^2 \bar{\mathbf{x}}_{P+1}\bar{y}_{P+1} \\
&= \mathbf{A}_P \mathbf{m}_P + \sigma^{-2}\bar{\mathbf{x}}_{P+1}\bar{y}_{P+1} \\
&= \mathbf{A}_P \sigma^{-2}\mathbf{A}_P^{-1}\tilde{\mathbf{X}}_P \bar{\mathbf{y}}_P + \sigma^{-2}\bar{\mathbf{x}}_{P+1}\bar{y}_{P+1} \\
&= \sigma^{-2}\tilde{\mathbf{X}}_P \bar{\mathbf{y}}_P + \sigma^{-2}\bar{\mathbf{x}}_{P+1}\bar{y}_{P+1} \\
&= \sigma^{-2}\tilde{\mathbf{X}}_{P+1}\bar{\mathbf{y}}_{P+1} \\
&= \mathbf{A}_{P+1}\sigma^{-2}\mathbf{A}_{P+1}^{-1}\tilde{\mathbf{X}}_{P+1}\bar{\mathbf{y}}_{P+1} \\
&= \mathbf{A}_{P+1}\mathbf{m}_{P+1} \\
&= \mathbf{S}_{P+1}^{-1}\mathbf{m}_{P+1}
\end{aligned}
$$