Machine Learning 1 - Homework Week 2

Minh Ngo 10897402[1]

September 18, 2015

[1] University of Amsterdam
minh.ngole@student.uva.nl

Task 3.2

$$p(\mathbf{w}|\mathbf{m_N}, \mathbf{S_N}) = \mathcal{N}(\mathbf{w}|\mathbf{m_N}, \mathbf{S_N}) \tag{1}$$

$$\mathbf{S_N} = (\alpha\mathbf{I} + \beta\mathbf{\Phi}^T\mathbf{\Phi})^{-1} \tag{2}$$

$$\begin{aligned}\mathbf{m_N} &= \beta(\alpha\mathbf{I} + \beta\mathbf{\Phi}^T\mathbf{\Phi})\mathbf{\Phi}^T\mathbf{t} \\ &= \beta\mathbf{S_N}\mathbf{\Phi}^T\mathbf{t}\end{aligned} \tag{3}$$

$$\Leftrightarrow \beta\mathbf{\Phi}^T\mathbf{t} = {\mathbf{S_N}}^{-1}\mathbf{m_N} \tag{4}$$

$$\begin{aligned}&p(t_*|\boldsymbol{\phi_*}, \mathbf{w}, \beta)p(\mathbf{w}|\mathbf{m_N}, \mathbf{S_N}) \\ &= \mathcal{N}(t_*|{\boldsymbol{\phi_*}}^T\mathbf{w}, \frac{1}{\beta})\mathcal{N}(\mathbf{w}|\mathbf{m_N}, \mathbf{S_N}) \\ &= \sqrt{\frac{\beta}{2\pi}}exp\bigg(-\frac{\beta}{2\pi}(t_* - {\boldsymbol{\phi_*}}^T\mathbf{w})^2\bigg)\frac{1}{\sqrt{(2\pi)^N|\mathbf{S_N}|}}exp\bigg(-\frac{1}{2}(\mathbf{w} - \mathbf{m_N})^T{\mathbf{S_N}}^{-1}(\mathbf{w} - \mathbf{m_N})\bigg)\end{aligned} \tag{5}$$

$$Let\ c_1 = \sqrt{\frac{\beta}{2\pi}}\frac{1}{\sqrt{(2\pi)^N|\mathbf{S_N}|}} \tag{6}$$

$$\begin{aligned} E1 &= (t_* - \boldsymbol{\phi_*}^T\mathbf{w})^2 \\ &= t_*^2 - 2t_*\boldsymbol{\phi_*}^T\mathbf{w} + \mathbf{w}^T\boldsymbol{\phi_*}\boldsymbol{\phi_*}^T\mathbf{w} \\ &= t_*^2 - 2t_*\boldsymbol{\phi_*}^T\mathbf{w} + \mathbf{w}^T\boldsymbol{\phi_*}\boldsymbol{\phi_*}^T\mathbf{I}\mathbf{w} \end{aligned} \tag{7}$$

$$\begin{aligned} E2 &= (\mathbf{w} - \mathbf{m_N})^T\mathbf{S_N}^{-1}(\mathbf{w} - \mathbf{m_N}) \\ &= \mathbf{w}^T\mathbf{S_N}^{-1}\mathbf{w} - 2\mathbf{w}^T\mathbf{S_N}^{-1}\mathbf{m_N} + \mathbf{m_N}^T\mathbf{S_N}^{-1}\mathbf{m_N} \end{aligned} \tag{8}$$

$$Let\ c_2 = exp\bigg(-\frac{1}{2}(\beta t_*^2 + \mathbf{m_N}^T\mathbf{S_N}^{-1}\mathbf{m_N})\bigg) \tag{9}$$

$$\Rightarrow (5) = c_1c_2exp\bigg(-\frac{1}{2}(\beta\mathbf{w}^T\boldsymbol{\phi_*}\boldsymbol{\phi_*}^T\mathbf{I}\mathbf{w} + 2\beta t_*\boldsymbol{\phi_*}^T\mathbf{w} + \mathbf{w}^T\mathbf{S_N}^{-1}\mathbf{w} - 2\mathbf{w}^T\mathbf{S_N}^{-1}\mathbf{m_N})\bigg) \tag{10}$$

$$\begin{aligned} E3 &= \beta\mathbf{w}^T\boldsymbol{\phi_*}\boldsymbol{\phi_*}^T\mathbf{I}\mathbf{w} + 2\beta t_*\boldsymbol{\phi_*}^T\mathbf{w} + \mathbf{w}^T\mathbf{S_N}^{-1}\mathbf{w} - 2\mathbf{w}^T\mathbf{S_N}^{-1}\mathbf{m_N} \\ &= \mathbf{w}^T(\mathbf{S_N}^{-1} + \beta\boldsymbol{\phi_*}\boldsymbol{\phi_*}^T\mathbf{I})\mathbf{w} - 2\beta\mathbf{w}^T(t_*\boldsymbol{\phi_*} + \mathbf{S_N}^{-1}\mathbf{m_N}) \\ &= \mathbf{w}^T(\mathbf{S_N}^{-1} + \beta\boldsymbol{\phi_*}\boldsymbol{\phi_*}^T\mathbf{I})\mathbf{w} - 2\beta\mathbf{w}^T(t_*\boldsymbol{\phi_*} + \mathbf{S_N}^{-1}\mathbf{S_N}\boldsymbol{\Phi}^T\mathbf{t}) \\ &= \mathbf{w}^T(\mathbf{S_N}^{-1} + \beta\boldsymbol{\phi_*}\boldsymbol{\phi_*}^T\mathbf{I})\mathbf{w} - 2\beta\mathbf{w}^T(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t}) \end{aligned} \tag{11}$$

$$\begin{aligned} Let\ \mathbf{S_{S+1}^{-1}} &= \mathbf{S_N}^{-1} + \beta\boldsymbol{\phi_*}\boldsymbol{\phi_*}^T\mathbf{I} \\ \mathbf{S_{S+1}^{-1}}\mathbf{m_{N+1}} &= \mathbf{S_N}^{-1}\mathbf{m_N} + \beta\boldsymbol{\phi_*}^T t_* \end{aligned} \tag{12}$$

$$\begin{aligned} \Leftrightarrow \mathbf{m_{N+1}} &= \mathbf{S_{S+1}}(\mathbf{S_N}^{-1}\mathbf{m_N} + \beta\boldsymbol{\phi_*}^T t_*) \\ &= (\mathbf{S_N}^{-1} + \beta\boldsymbol{\phi_*}\boldsymbol{\phi_*}^T\mathbf{I})^{-1}(\mathbf{S_N}^{-1}\mathbf{m_N} + \beta\boldsymbol{\phi_*}^T t_*) \end{aligned} \tag{13}$$

$$\begin{aligned} \Rightarrow E3 &= \mathbf{w}^T\mathbf{S_{S+1}^{-1}}\mathbf{w} - 2[\mathbf{w}^T]\mathbf{S_{S+1}^{-1}}[\mathbf{S_{S+1}}\beta(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t})] \\ &+ [\mathbf{S_{S+1}}\beta(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t})]^T\mathbf{S_{S+1}^{-1}}[\mathbf{S_{S+1}}\beta(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t})] \\ &- [\mathbf{S_{S+1}}\beta(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t})]^T\mathbf{S_{S+1}^{-1}}[\mathbf{S_{S+1}}\beta(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t})] \\ &= [\mathbf{w} - \beta\mathbf{S_{S+1}}(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t})]^T\mathbf{S_{S+1}^{-1}}[\mathbf{w} - \beta\mathbf{S_{S+1}}(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t})] \\ &- [\mathbf{S_{S+1}}\beta(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t})]^T\mathbf{S_{S+1}^{-1}}[\mathbf{S_{S+1}}\beta(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t})] \end{aligned} \tag{14}$$

$$
\begin{aligned}
\Rightarrow (11) &= c_1 c_2 exp\bigg(-\frac{1}{2}E3\bigg) \\
&= c_1 c_2 \sqrt{(2\pi)^N |\mathbf{S_{S+1}}|}\frac{1}{\sqrt{(2\pi)^N |\mathbf{S_{S+1}}|}} exp\bigg(-\frac{1}{2}E3\bigg)
\end{aligned} \tag{15}
$$

$$
Let\ c_3 = \sqrt{(2\pi)^N |\mathbf{S_{S+1}}|} exp\bigg(\frac{1}{2}[\mathbf{S_{S+1}}\beta(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t})]^T \mathbf{S_{S+1}^{-1}}[\mathbf{S_{S+1}}\beta(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t})]\bigg) \tag{16}
$$

$$
\begin{aligned}
\Rightarrow (15) &= c_1 c_2 c_3 \frac{1}{\sqrt{(2\pi)^N |\mathbf{S_{S+1}}|}} \times \\
&\times exp\bigg(-\frac{1}{2}[\mathbf{w} - \beta\mathbf{S_{S+1}}(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t})]^T \mathbf{S_{S+1}^{-1}}[\mathbf{w} - \beta\mathbf{S_{S+1}}(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t})]\bigg) \\
&= c\mathcal{N}(\mathbf{w}|\beta\mathbf{S_{S+1}}(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t}), \mathbf{S_{S+1}^{-1}}) \\
&where\ c = c_1 c_2 c_3
\end{aligned} \tag{17}
$$

$$
\begin{aligned}
c &= \sqrt{\frac{\beta}{2\pi}} \frac{1}{\sqrt{(2\pi)^N |\mathbf{S_N}|}} \sqrt{(2\pi)^N |\mathbf{S_{S+1}}|} \times \\
&\times exp\bigg(-\frac{1}{2}(\beta t_*^2 + \mathbf{m_N}^T \mathbf{S_N}^{-1}\mathbf{m_N} - \\
&\qquad - [\mathbf{S_{S+1}}\beta(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t})]^T \mathbf{S_{S+1}^{-1}}[\mathbf{S_{S+1}}\beta(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t})])\bigg) \\
&= \sqrt{\frac{\beta}{2\pi}} \sqrt{\frac{|\mathbf{S_{S+1}}|}{|\mathbf{S_N}|}} exp\bigg(\frac{1}{2}E4\bigg) \\
where\ E4 &= \beta t_*^2 + \mathbf{m_N}^T \mathbf{S_N}^{-1}\mathbf{m_N} - [\mathbf{S_{S+1}}\beta(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t})]^T \mathbf{S_{S+1}^{-1}}[\mathbf{S_{S+1}}\beta(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t})]
\end{aligned} \tag{18}
$$

$$
\begin{aligned}
det(\mathbf{I} + \mathbf{u}\mathbf{v}^T) &= 1 + \mathbf{u}^T\mathbf{v} \\
\Rightarrow \sqrt{\frac{|\mathbf{S_{S+1}}|}{|\mathbf{S_N}|}} &= \frac{1}{\sqrt{|(\beta\boldsymbol{\phi_*}\boldsymbol{\phi_*}^T\mathbf{I} + \mathbf{S_N}^{-1})\mathbf{S_N}|}} \\
&= \frac{1}{\sqrt{\beta\boldsymbol{\phi_*}\boldsymbol{\phi_*}^T\mathbf{S_N} + \mathbf{I}}} \\
&= \frac{1}{1 + \beta\boldsymbol{\phi_*}^T(\boldsymbol{\phi_*}\mathbf{S_N}^T)} \\
&= \frac{1}{\sqrt{1 + \beta\boldsymbol{\phi_*}^T\mathbf{S_N}\boldsymbol{\phi_*}}} \\
\Leftrightarrow \sqrt{\frac{\beta}{2\pi}}\sqrt{\frac{|\mathbf{S_{S+1}}|}{|\mathbf{S_N}|}} &= \frac{1}{\sqrt{\frac{1}{\beta} + \boldsymbol{\phi_*}^T\mathbf{S_N}\boldsymbol{\phi_*}}}
\end{aligned}
\tag{19}
$$

$$
\mathbf{m_{N+1}}^T\mathbf{S_{S+1}^{-1}}\mathbf{m_{N+1}} = (\mathbf{S_N}^{-1}\mathbf{m_N} + \beta\boldsymbol{\phi_*}^T t_*)^T\mathbf{S_{S+1}}(\mathbf{S_N}^{-1}\mathbf{m_N} + \beta\boldsymbol{\phi_*}^T t_*) \tag{20}
$$

$$
\begin{aligned}
\Rightarrow E4 &= \beta t_*^2 + \mathbf{m_N}^T\mathbf{S_N}^{-1}\mathbf{m_N} - [\mathbf{S_{S+1}}\beta(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t})]^T\mathbf{S_{S+1}^{-1}}[\mathbf{S_{S+1}}\beta(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t})] \\
&= \beta t_*^2 + \mathbf{m_N}^T\mathbf{S_N}^{-1}\mathbf{m_N} - \mathbf{m_{N+1}}^T\mathbf{S_{S+1}^{-1}}\mathbf{m_{N+1}}
\end{aligned}
\tag{21}
$$

$$
\begin{aligned}
(21) &= \beta t_*^2 - (\mathbf{S_N}^{-1}\mathbf{m_N} + \beta\boldsymbol{\phi_*}^T t_*)^T\mathbf{S_{S+1}}(\mathbf{S_N}^{-1}\mathbf{m_N} + \beta\boldsymbol{\phi_*}^T t_*) + \mathbf{m_N}^T\mathbf{S_N}^{-1}\mathbf{m_N} \\
&= t_*(\beta - \beta^2\boldsymbol{\phi_*}\mathbf{S_{S+1}}\boldsymbol{\phi_*}^T)t_* - 2t_*(\beta\boldsymbol{\phi_*}\mathbf{S_{S+1}}\mathbf{S_N}^{-1}\mathbf{m_N}) + \\
&\quad + (\mathbf{m_N}^T\mathbf{S_N}^{-1}\mathbf{m_N} - \mathbf{m_N}^T(\mathbf{S_N}^{-1})^T\mathbf{S_{S+1}}\mathbf{S_N}^{-1}\mathbf{m_N})
\end{aligned}
\tag{22}
$$

$$
\begin{aligned}
Let\ \beta_*(\boldsymbol{\phi}) &= \beta - \beta^2\boldsymbol{\phi_*}\mathbf{S_{S+1}}\boldsymbol{\phi_*}^T \\
\beta_*(\boldsymbol{\phi})y_* &= \beta\boldsymbol{\phi_*}\mathbf{S_{S+1}}\mathbf{S_N}^{-1}\mathbf{m_N} \\
\beta_*(\boldsymbol{\phi})y_*^2 &= \mathbf{m_N}^T\mathbf{S_N}^{-1}\mathbf{m_N} - \mathbf{m_N}^T(\mathbf{S_N}^{-1})^T\mathbf{S_{S+1}}\mathbf{S_N}^{-1}\mathbf{m_N}
\end{aligned}
\tag{23}
$$

$$
(\mathbf{A} + \mathbf{C}\mathbf{B}\mathbf{C}^T)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}(\mathbf{B}^{-1} - \mathbf{C}^T\mathbf{A}^{-1}\mathbf{C})\mathbf{C}^T\mathbf{A}^{-1} \tag{24}
$$

$$\Rightarrow \beta_*(\boldsymbol{\phi}) = \beta - \beta\boldsymbol{\phi_*}(\mathbf{S_N}^{-1} + \beta\boldsymbol{\phi_*}\boldsymbol{\phi_*}^T\mathbf{I})^{-1}\boldsymbol{\phi_*}\beta$$
$$= \frac{1}{\frac{1}{\beta} + \boldsymbol{\phi_*}\mathbf{S_N}\boldsymbol{\phi_*}^T} \tag{25}$$

$$\begin{aligned}\Rightarrow y_* &= \frac{1}{\beta_*(\boldsymbol{\phi})}(\beta\boldsymbol{\phi_*}\mathbf{S_{S+1}}\mathbf{S_N}^{-1}\mathbf{m_N}) \\ &= (\frac{1}{\beta} + \boldsymbol{\phi_*}\mathbf{S_N}\boldsymbol{\phi_*}^T)\beta\boldsymbol{\phi_*}(\beta\boldsymbol{\phi_*}\boldsymbol{\phi_*}^T\mathbf{I} + \mathbf{S_N}^{-1})^{-1}\mathbf{S_N}^{-1}\mathbf{m_N} \\ &= \boldsymbol{\phi_*}[((1 + \beta\boldsymbol{\phi_*}\mathbf{S_N}\boldsymbol{\phi_*}^T)(1 + \beta\boldsymbol{\phi_*}\mathbf{S_N}\boldsymbol{\phi_*}^T)^{-1}\mathbf{S_N}^{-1}]\mathbf{m_N} \\ &= \boldsymbol{\phi_*}[\mathbf{S_N}\mathbf{S_N}^{-1}((1 + \beta\boldsymbol{\phi_*}\mathbf{S_N}\boldsymbol{\phi_*}^T)(1 + \beta\boldsymbol{\phi_*}\mathbf{S_N}\boldsymbol{\phi_*}^T)^{-1}\mathbf{S_N}^{-1}]\mathbf{m_N} \\ &= \boldsymbol{\phi_*}[\mathbf{S_N}\mathbf{S_N}^{-1}]\mathbf{m_N} = \boldsymbol{\phi_*}\mathbf{m_N}\end{aligned} \tag{26}$$

$$\Rightarrow (22) = (t_* - y_*)^2\beta_*(\boldsymbol{\phi}) = E4 \tag{27}$$

$$\begin{aligned}\Rightarrow c &= \sqrt{\frac{\beta}{2\pi}}\sqrt{\frac{|\mathbf{S_{S+1}}|}{|\mathbf{S_N}|}}exp\left(\frac{1}{2}E4\right) \\ &= \frac{1}{\sqrt{\frac{1}{\beta} + \boldsymbol{\phi_*}^T\mathbf{S_N}\boldsymbol{\phi_*}}}exp\left(-\frac{1}{2}(t_* - y_*)^2\beta_*(\boldsymbol{\phi})\right) \\ &= \frac{1}{\sqrt{\frac{1}{\beta} + \boldsymbol{\phi_*}^T\mathbf{S_N}\boldsymbol{\phi_*}}}exp\left(-\frac{1}{2}\frac{(t_* - \boldsymbol{\phi_*}\mathbf{m_N})^2}{\frac{1}{\beta} + \boldsymbol{\phi_*}\mathbf{S_N}\boldsymbol{\phi_*}^T}\right) \\ &= \frac{1}{\sqrt{\frac{1}{\beta} + \boldsymbol{\phi_*}^T\mathbf{S_N}\boldsymbol{\phi_*}}}exp\left(-\frac{1}{2}\frac{(t_* - \boldsymbol{\phi_*}\mathbf{m_N})^2}{\frac{1}{\beta} + \boldsymbol{\phi_*}^T\mathbf{S_N}\boldsymbol{\phi_*}}\right) \\ &= \mathcal{N}(t_*|\boldsymbol{\phi_*}\mathbf{m_N}, \frac{1}{\beta} + \boldsymbol{\phi_*}^T\mathbf{S_N}\boldsymbol{\phi_*})\end{aligned} \tag{28}$$

$$\begin{aligned}\Rightarrow p(t_*|\boldsymbol{\phi_*}, \mathbf{w}, \beta)p(\mathbf{w}|\mathbf{m_N}, \mathbf{S_N}) &= \mathcal{N}(t_*|\boldsymbol{\phi_*}\mathbf{m_N}, \frac{1}{\beta} + \boldsymbol{\phi_*}^T\mathbf{S_N}\boldsymbol{\phi_*})\times \\ &\quad \times \mathcal{N}(\mathbf{w}|\beta\mathbf{S_{S+1}}(t_*\boldsymbol{\phi_*} + \boldsymbol{\Phi}^T\mathbf{t}), \mathbf{S_{S+1}^{-1}})\end{aligned} \tag{29}$$

$$
\begin{aligned}
p(t_*|\boldsymbol{\phi_*}, \boldsymbol{\Phi}, \alpha, \beta) &= \int p(t_*|\boldsymbol{\phi_*}, \mathbf{w}, \beta) p(\mathbf{w}|\mathbf{m_N}, \mathbf{S_N}) d\mathbf{w} \\
&= \mathcal{N}(t_*|\boldsymbol{\phi_*}\mathbf{m_N}, \frac{1}{\beta} + \boldsymbol{\phi_*}^T \mathbf{S_N} \boldsymbol{\phi_*}) \times \\
&\times \int \mathcal{N}(\mathbf{w}|\beta \mathbf{S_{S+1}}(t_* \boldsymbol{\phi_*} + \boldsymbol{\Phi}^T \mathbf{t}), \mathbf{S_{S+1}^{-1}}) d\mathbf{w} \\
&= \mathcal{N}(t_*|\boldsymbol{\phi_*}\mathbf{m_N}, \frac{1}{\beta} + \boldsymbol{\phi_*}^T \mathbf{S_N} \boldsymbol{\phi_*})
\end{aligned}
\tag{30}
$$