HW3

12011923 张旭东

Q1

Define

$$R = diag(r_1, r_2, \ldots, r_N) \tag{1}$$

Then

$$E_D(\mathbf{w}) = \frac{1}{2}(t - \varnothing\mathbf{w})^{\mathrm{T}} R (t - \varnothing\mathbf{w}) \tag{2}$$

Let

$$\frac{dE_D(\mathbf{w})}{d\mathbf{w}} = 0 \tag{3}$$

We can get

$$\varnothing^{\mathrm{T}} R \varnothing \mathbf{w} + \mathbf{w}^{\mathrm{T}} \varnothing^{\mathrm{T}} R \varnothing = t^{\mathrm{T}} R \varnothing + \varnothing^{\mathrm{T}} R t \tag{4}$$

Then

$$\mathbf{w}^* = (\varnothing^{\mathrm{T}} R \varnothing)^{-1} \varnothing^{\mathrm{T}} R t \tag{5}$$

(1) According to the formula **(5)**, we can get

$$\beta(t|\mathbf{x}, \mathbf{w}, \beta) = \prod_{n=1}^{N} N(t_n | \mathbf{w}^{\mathrm{T}} \varnothing(\mathbf{x_n}), \beta^{-1}) \tag{6}$$

$$\ln p(t|\mathbf{w}, \beta) = \frac{N}{2} \ln \beta - \frac{N}{2} \ln(2\pi) - \beta E_D^{'}(\mathbf{w}) \tag{7}$$

$$E_D^{'}(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} (t_n - \mathbf{w}^{\mathrm{T}} \varnothing(\mathbf{x_n}))^2 \tag{8}$$

Therefore, r_n can be regarded as β parameter particular to the data point $(\mathbf{x_n}, t_n)$

(2) r_n can be regarded as an effective number of replicated observations of data point $(\mathbf{x_n}, t_n)$. If r_n is positive integer, it's more clear to be interpreted this way.

Q2

$$\ln p(\mathbf{w}, \beta|t) = \ln p(\mathbf{w}, \beta) + \sum_{n=1}^{N} \ln p(t_n|\mathbf{w}^{\mathrm{T}}\varnothing(\mathbf{x_n}), \beta^{-1}) \tag{9}$$

$$= \frac{M}{2}\ln\beta - \frac{1}{2}\ln\mathbf{S_0} - \frac{\beta}{2}(\mathbf{w} - \mathbf{m_0})^{\mathrm{T}}\mathbf{S_0}^{-1}(\mathbf{w} - \mathbf{m_0}) - b_0\beta \tag{10}$$

$$+(a_0 - 1)\ln\beta + \frac{N}{2}\ln\beta - \frac{\beta}{2}\sum_{n=1}^{N}(\mathbf{w}^{\mathrm{T}}\varnothing(\mathbf{x_n}) - t_n)^2 + C \tag{11}$$

where C is a constant.

$$p(\mathbf{w}, \beta|\mathbf{t}) = p(\mathbf{w}|\beta, \mathbf{t})p(\beta|\mathbf{t}) \tag{12}$$

consider the first dependence on $\mathbf{w}$, we have

$$\ln p(\mathbf{w}|\beta, \mathbf{t}) = -\frac{\beta}{2}\mathbf{w}^{\mathrm{T}}[\varnothing^{\mathrm{T}}\varnothing + \mathbf{S_0^{-1}}]\mathbf{w} + \mathbf{w}^{\mathrm{T}}[\beta\mathbf{S_0^{-1}}\mathbf{m_0} + \beta\varnothing^{\mathrm{T}}t] + C \tag{13}$$

Therefore, $p(\mathbf{w}|\beta, t)$ is a Gaussian distribution with mean and covariance given by

$$\mathbf{m_N} = \mathbf{S_N}[\mathbf{S_0}^{-1}\mathbf{m_0} + \varnothing^{\mathrm{T}}t] \tag{14}$$

$$\beta\mathbf{S_N}^{-1} = \beta(\mathbf{S_0}^{-1} + \varnothing^{\mathrm{T}}\varnothing) \tag{15}$$

$$\mathbf{S_N} = (\mathbf{S_0}^{-1} + \varnothing^{\mathrm{T}}\varnothing)^{-1} \tag{16}$$

$$\ln p(\beta|t) = -\frac{\beta}{2}\mathbf{m_0}^{\mathrm{T}}\mathbf{S_0}^{-1}\mathbf{m_0} + \frac{\beta}{2}\mathbf{m_N}^{\mathrm{T}}\mathbf{S_N}^{\mathrm{T}}\mathbf{m_N} + \frac{N}{2}\ln\beta - b_0\beta + (a_0 - 1)\ln\beta - \frac{\beta}{2}\sum_{n=1}^{N}t_n^2 + C \tag{17}$$

We recognize the log of Gamma distribution

$$\Gamma(\beta|a_N, b_N) = \frac{b_N^{a_N}\beta^{a_N - 1}e^{-b_N\beta}}{\Gamma(a_N)} \tag{18}$$

$$\log(\Gamma(\beta|a_N, b_N)) = a_N\log b_N + (a_N - 1)\log\beta - b_N\beta - log(\Gamma(a_N)) \tag{19}$$

$$a_N = a_0 + \frac{N}{2} \tag{20}$$

$$b_N = b_0 + \frac{1}{2}(\mathbf{m_0}^{\mathrm{T}}\mathbf{S_0}^{-1}\mathbf{m_0} + \mathbf{m_N}^{\mathrm{T}}\mathbf{S_N}^{-1}\mathbf{m_N} + \sum_{n=1}^{N}t_n^2) \tag{21}$$

Q3

$$\begin{aligned}
\int \exp\left[E(\mathbf{w})\right] d\mathbf{w} &= \int \exp[-E(\mathbf{m_N}) - \frac{1}{2}(\mathbf{w}-\mathbf{m_N})^\intercal \mathbf{A}(\mathbf{w}-\mathbf{m_N})] d\mathbf{w} \\
&= \exp[-E(\mathbf{m_N})] \int \exp[-\frac{1}{2}(\mathbf{w}-\mathbf{m_N})^\intercal \mathbf{A}(\mathbf{w}-\mathbf{m_N})] d\mathbf{w} \\
&= \exp[-E(\mathbf{m_N})](2\pi)^{\frac{\mathbf{m}}{2}}|\mathbf{A}|^{-\frac{1}{2}} \int \frac{1}{(2\pi)^{\frac{\mathbf{m}}{2}}} \frac{1}{|\mathbf{A}|^{-\frac{1}{2}}} \exp[-\frac{1}{2}(\mathbf{w}-\mathbf{m_N})^\intercal \mathbf{A}(\mathbf{w}-\mathbf{m_N})] d\mathbf{w} \\
&= \exp[-E(\mathbf{m_N})](2\pi)^{\frac{\mathbf{m}}{2}}|\mathbf{A}|^{-\frac{1}{2}}
\end{aligned} \tag{22}$$

$$\begin{aligned}
p(\mathbf{t}|\alpha,\beta) &= (\frac{\beta}{2\pi})^{\frac{N}{2}}(\frac{\alpha}{2\pi})^{\frac{M}{2}} \int \exp[E(\mathbf{w})] d\mathbf{w} \\
&= (\frac{\beta}{2\pi})^{\frac{N}{2}}(\frac{\alpha}{2\pi})^{\frac{M}{2}} \exp[E(\mathbf{m_N})](2\pi)^{\frac{M}{2}}|\mathbf{A}|^{-\frac{1}{2}}
\end{aligned} \tag{23}$$

$$\ln p(\mathbf{t}|\alpha,\beta) = \frac{M}{2}\ln\alpha + \frac{N}{2}\ln\beta - E(\mathbf{m_N}) - \frac{1}{2}\ln|\mathbf{A}| - \frac{N}{2}\ln(2\pi) \tag{24}$$

Q4

$$\frac{\partial F(a)}{\partial a} = \sum_i (Y_i - aX_i)(-X_i) = \sum_i aX_i - X_iY_i \tag{25}$$

$$a = \frac{\sum_i X_iY_i}{\sum_i X_i^2} \tag{26}$$

Q5

$$\begin{aligned}
\log p(y|\theta) &= y\log\theta - \theta - \sum_{i=0}^{y}\log i \\
&= \sum_{i=1}^{n}(y_i\log\theta - \theta - \log y_i!) \\
&= \sum_{i=1}^{n}(y_i\log\theta - \log y_i!) - n\theta
\end{aligned} \tag{27}$$

Q6

$$\log f_X(x) = \alpha\log\lambda + (\alpha-1)\log(x) + \lambda x - \log\Gamma(\alpha) \tag{28}$$

$$g(\lambda) = n\alpha\log(\lambda) + (\alpha-1)\log(\prod_{i=1}^{n}X_i) - \lambda\sum_{i=1}^{n}X_i - n\log\Gamma(\alpha) \tag{29}$$

$$\frac{dg(\lambda)}{d\lambda} = \frac{n\alpha}{\lambda} - \sum_{i=1}^{n} X_i \tag{30}$$

$$\hat{\lambda} = \frac{\alpha}{\frac{1}{n}\sum_{i=1}^{n} X_i} \tag{31}$$