Machine Learning (H)
Southern University of Science and Technology
Mengxuan Wu
12212006

Assignment 3

Mengxuan Wu

Question 1

Let R be a diagonal matrix with r_n on the diagonal, then we have

$$E_D(w) = \frac{1}{2}(t - \Phi w)^T R(t - \Phi w)$$

Taking the derivative of $E_D(w)$ with respect to w and setting it to zero, we have

$$\frac{\partial E_D(w)}{\partial w} = \Phi^T R(t - \Phi w) = 0$$

Solving the equation, we have

$$w^* = (\Phi^T R \Phi)^{-1} \Phi^T R t$$

In the data independent noise variance view, R takes the place of precision factor β. Hence, it represents the variance of the noise in each data point.
In the replicated data view, R can be viewed as the number of replicated data points. The error for each replicated data point is r_n times the error for the original data point.

Question 2

The log of the posterior distribution is

$$\begin{aligned}
\ln p(w, \beta | t) &= \ln p(w, \beta) + \sum_{n=1}^{N} \ln p(t_n | w^T \phi(x_n), \beta^{-1}) \\
&= \frac{M}{2} \ln \beta - \frac{\beta}{2}(w - m_0)^T S_0^{-1}(w - m_0) - \frac{1}{2} \ln |S_0| - b_0 \beta \\
&+ (a_0 - 1) \ln \beta + \frac{N}{2} \ln \beta - \frac{\beta}{2} \sum_{n=1}^{N} (t_n - w^T \phi(x_n))^2 + \text{const}
\end{aligned}$$

Consider the dependence of the posterior distribution on w, we have

$$\ln p(w | \beta, t) = \frac{\beta}{2} w^T (\Phi^T \Phi + S_0^{-1}) w - \beta w^T (\Phi^T t + S_0^{-1} m_0) + \text{const}$$

Thus, mean and covariance of the posterior distribution are

$$m_N = S_N(\Phi^T t + S_0^{-1} m_0)$$
$$S_N = (\Phi^T \Phi + S_0^{-1})^{-1}$$

We then consider the dependence of the posterior distribution on β, we have

$$\ln p(\beta|w,t) = -\frac{\beta}{2} m_0^T S_0 m_0 + \frac{\beta}{2} m_N^T S_N m_N + (a_0 + \frac{N}{2} - 1)\ln\beta - b_0\beta - \frac{\beta}{2}\sum_{n=1}^{N} t_n^2 + \text{const}$$

Thus, we have

$$a_N = a_0 + \frac{N}{2}$$
$$b_N = b_0 + \frac{1}{2}\{m_0^T S_0^{-1} m_0 - m_N^T S_N^{-1} m_N + \sum_{n=1}^{N} t_n^2\}$$

Question 3

$$E(w) = E(m_N) + \frac{1}{2}(w - m_N)^T A (w - m_N)$$

where $A = \beta\Phi^T\Phi + \alpha I$.
And we have

$$\int \exp\{-E(w)\}\, dw = \exp\{-E(m_N)\}\,(2\pi)^{M/2}|A|^{-1/2}$$

To integrate the Gaussian distribution, we have

$$\int \frac{1}{(2\pi)^{M/2}|A|^{-1/2}} \exp\left\{-\frac{1}{2}(w - m_N)^T A (w - m_N)\right\} dw = 1$$

Thus, we have

$$\begin{aligned}
\ln p(t|\alpha,\beta) &= \ln\left\{(\frac{\beta}{2\pi})^{N/2}(\frac{\alpha}{2\pi})^{M/2}\int \exp\{-E(w)\}\, dw\right\} \\
&= \ln\left\{(\frac{\beta}{2\pi})^{N/2}(\frac{\alpha}{2\pi})^{M/2}\exp\{-E(m_N)\}\,(2\pi)^{M/2}|A|^{-1/2}\right\} \\
&= \frac{N}{2}\ln\beta + \frac{M}{2}\ln\alpha - E(m_N) - \frac{1}{2}\ln|A| - \frac{N}{2}\ln(2\pi)
\end{aligned}$$

Question 4

The log likelihood function is

$$\ln F(a) = \ln\prod_{i=1}^{n} p(Y_i|X_i,a) = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - aX_i)^2$$

Taking the derivative of $\ln F(a)$ with respect to a and setting it to zero, we have

$$\frac{\partial \ln F(a)}{\partial a} = \frac{1}{\sigma^2}\sum_{i=1}^{n} X_i(Y_i - aX_i) = 0$$

Solving the equation, we have

$$a_{\text{ML}} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$$

Question 5

The log likelihood function is

$$\begin{aligned}\ln L(\theta) &= \ln \prod_{i=1}^{n} p(y_i|\theta) \\ &= \sum_{i=1}^{n} \ln \frac{\theta^{y_i} e^{-\theta}}{y_i!} \\ &= \sum_{i=1}^{n} y_i \ln \theta - n\theta - \sum_{i=1}^{n} \ln y_i!\end{aligned}$$

Question 6

The log likelihood function is

$$\begin{aligned}\ln L(\alpha, \lambda) &= \ln \prod_{i=1}^{n} p(X_i|\alpha, \lambda) \\ &= \ln \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)} \lambda^\alpha X_i^{\alpha-1} e^{-\lambda X_i} \\ &= \sum_{i=1}^{n} \ln \frac{1}{\Gamma(\alpha)} + \alpha \ln \lambda + (\alpha - 1) \ln X_i - \lambda X_i\end{aligned}$$

Taking the derivative of $\ln L(\alpha, \lambda)$ with respect to λ and setting it to zero, we have

$$\frac{\partial \ln L(\alpha, \lambda)}{\partial \lambda} = \frac{n\alpha}{\lambda} - \sum_{i=1}^{n} X_i = 0$$

Solving the equation, we have

$$\lambda_{\mathrm{ML}} = \frac{n\alpha}{\sum_{i=1}^{n} X_i}$$