統計學習初論（105-2）

Self-Assessment Homework 2

國立台灣大學資管系

Question 1 (Exercise 3.7)

Consider the Bayesian analysis of linear regression. Let the prior distribution of slope vector w follow a normal distribution: $p(w) = N(w|m_0, S_0)$. The likelihood function of observations $\boldsymbol{t} = (t_1\ t_2 \ldots, t_N)^T$, and $X = (x_1\ x_2 \ldots\ x_N)^T$ is $p(\boldsymbol{t}|X, w, \beta) = \prod_{n=1}^{N} N(t_n|w^T\phi(x_n), \beta^{-1})$. To simply the problem, assume that β is a known scalar. The posterior distribution of w is $p(w|\boldsymbol{t}) \propto p(\boldsymbol{t}|X, w, \beta)p(w)$. Show that $p(w|\boldsymbol{t}) = N(w|m_N, S_N)$, where $m_N = S_N(S_0^{-1}m_0 + \beta\Phi^T\boldsymbol{t})$ and $S_N^{-1} = S_0^{-1} + \beta\Phi^T\Phi$.

Question 2 (Exercise 3.10)

We know that in a regression model with training data $\boldsymbol{t} = (t_1\ t_2 \ldots, t_N)^T$, and $X = (x_1\ x_2 \ldots\ x_N)^T$, the posterior distribution of the slope vector w is $p(w|\boldsymbol{t}) = N(w|m_N, S_N)$, where $m_N = S_N(S_0^{-1}m_0 + \beta\Phi^T\boldsymbol{t})$ and $S_N^{-1} = S_0^{-1} + \beta\Phi^T\Phi$. Given an new data point with input feature x_a, the predictive distribution for its outcome t_a is: $p(t_a|x_a, \boldsymbol{t}) = \int N(t_a|w^T\phi(x_a),\ \beta^{-1})N(w|m_N, S_N)dw$.

Show that $p(t_a|x_a, \boldsymbol{t}) = N\left(t_a\middle|m_N^T\phi(x_a),\ \frac{1}{\beta} + \phi(x_a)^T S_N\phi(x_a)\right)$.

Question 3 (Exercise 3.14)

Show that when $\alpha = 0$, the kernel function $k(x, x') = \beta\phi(x)^T S_N\phi(x')$ has the property that $\sum_{n=1}^{N} k(x, x_n) = 1$.

Question 4 (Exercise 3.18)

Consider the evident function of Bayesian regression $\mathrm{p}(\mathbf{t}|\alpha, \beta) = \int p(\boldsymbol{t}|\boldsymbol{X}, w, \beta)p(w|\alpha)dw = \left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}}\left(\frac{\alpha}{2\pi}\right)^{\frac{M}{2}}\int \exp[-E(w)]\,dw$, where $E(w) = \beta E_D(w) + \alpha E_W(w) = \frac{\beta}{2}\|t - \Phi w\|^2 + \frac{\alpha}{2}w^Tw$. Show that $E(w) = \frac{\beta}{2}\|t - \Phi m_N\|^2 + \frac{\alpha}{2}m_N^T m_N + \frac{1}{2}(w - m_N)^T A(w - m_N)$, where $A = \alpha I + \beta\Phi^T\Phi$ and $m_N = \beta A^{-1}\Phi^T\boldsymbol{t}$.