Machine Learning

Homework 4 solution - Shyam Arumugaswamy

2. Let dataset be $Z_i \epsilon (x_i, y_i)$
   and t1,t2,…ti be the weighted coefficients in a diagonal matrix T,
   so the weighted sum of squared error function is defined as

   $$E_{weighted}(W) = \frac{1}{2} (Z - \phi W)^T T(Z - \phi W)$$

   $$= \frac{1}{2} [Z^T T (Z - \phi W) - W^T \phi^T T(Z - \Phi w)]$$

   now differentiating wrt W

   $$0 = -\phi^T T Z + \phi^T T \phi W_{ml}$$

   $$W_{ml} = (\phi^T T \phi)^{-1} (\phi^T T Z)$$

   if we take T matrix as identity matrix

   $$W_{ml} = (\phi^T \phi)^{-1} (\phi^T T Z)$$

   Likelihood of a univariate gaussian function was of the form

   $$P(Z \mid X, W, \beta) = \prod^N_{n=1} N (Z_n \mid W^T \phi(X_n), \beta^{-1})$$

   Applying log
   $$\ln(P(Z \mid X, W, \beta)) = \prod^N_{n=1} \ln(N (Z_n \mid W^T \phi(X_n), \beta^{-1})$$

   $$= -(N/2) \ln 2\pi - \frac{1}{2} \beta\sum(Z_n - W^T \phi(X_n)^2) + (N/2) \ln\beta$$

   where $\frac{1}{2} \sum(Z_n - W^T \phi(X_n)^2) = E^\wedge(W)$ ----→ Eq 1

   Comparing the eqs 1 of $E^\wedge(W)$ and the given eq,
   Ti can be considered as a inverse variance parameter that scales β and it can be regarded as replicated observations of dataset for all positive values of Ti

3. We define design matrix as $\phi^\wedge = (\frac{\phi}{\sqrt{\lambda I}})$ and $y^\wedge = (\frac{y}{0M})$

   for ridge regression $(y - \phi w)^T (y - \phi w) + \lambda w^T w$

4. Taking log of posterior distribution

$\ln p(w, \beta \mid Z) = \ln p(w, \beta) + \sum \ln p(Yi \mid W^T \phi(x_i), \beta^{-1})$

$= M/2 \ln \beta - ½ \ln |S_0| - \beta/2 (W - M_0)^T S_0^{-1}(W - M_0) - b_0\beta + (a_0 - 1) \ln \beta + N/2 \ln \beta - B/2 \sum (W^T \phi(x_i) - Yi)^2 + \text{const}$

posterior can be written as using product rule
$p(w, \beta \mid Z) = p(w \mid \beta , Z)\, p(\beta \mid Z)$

now
$\ln p(w \mid \beta , Z) = -\beta/2\, W^T [\phi^T \phi + S_0^{-1}]W + W^T[\beta S_0^{-1} M_0 + \beta \phi^T Z] + \text{const}$

Thus $p(w \mid \beta , Z)$ is a gaussian distribution with mean and covariance given by

$M_N = S_N [S_0^{-1} M_0 + \phi^T Z]$
$S_N^{-1} = \beta[\phi^T \phi + S_0^{-1}]$

$\ln p(\beta \mid Z) = -\beta/2\, M_0^T S_0^{-1} M_0 + \beta/2\, M_N^T S_N^{-1} M_N + N/2 \ln\beta - b_0 \beta + (a_0 - 1) \ln\beta - \beta/2 \sum Z_n^2 + \text{const}$

this is gamma distribution , so coefficients of β and lnβ will give us

$a_N = a_0 + N/2$

$b_N = b_0 + 1/2(M_0^T S_0^{-1} M_0 - M_N^T S_N^{-1} M_N + \sum Z_n^2)$