FUNDAMENTALS OF MACHINE LEARNING
ASSIGNMENT 1

TEJAL KULKARNI - CS21BTECH11058
DEEPSHIKHA - CS21BTECH11016

Question 3

Part(a):

In heteroscedasticity the variance of errors is not constant. In a linear regression model we have,

$$y_n = w^T x_n + \epsilon_n$$

where,
$y_n = n^{th}$ prediction point
$w =$ Weight matrix
$x_n = n^{th}$input vector
$\epsilon_n = n^{th}$ error term

We have,

$$p(t_n|x_n, w) = N(t_n|w^T x_n, \sigma_n^2)$$

Hence,

$$p(t_n|x_n, w) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(t_n - w^T x_n)^2}{2\sigma_n^2}\right)$$

Since the error variance is difference for each data point we have considered σ_n^2 to be the variance for the nth data point.
Assuming the prior probability is modelled by a gaussian we have,

$$p(w) = N(0, \Sigma)$$

where Σ is the covariance matrix

Part(b): Combining probabilities of each data point to calculate the ML estimate, we have the formula,

$$p(t|x, w) = \prod_{n=1}^{N} N(t_n|w^T x_n, \sigma_n^2)$$

By taking log of the following we get the log likelihood.

$$p(t|x, w) = \sum_{n=1}^{N} \log N(t_n|w^T x_n, \sigma_n^2)$$

hence ML objective function L(w) is

$$L_{ML}(w) = \sum_{n=1}^{N} \log N(t_n|w^T x_n, \sigma_n^2)$$

Date: 8th October 2023.

For MAP estimate,

$$p(w|X,t) \propto p(t|X,w)p(w)$$

Hence,

$$p(w|X,t) = \prod_{n=1}^{N} N(t_n|w^T x_n, \sigma_n^2)N(0,\Sigma)$$

On taking log we get,

$$\begin{aligned} p(w|X,t) &= \sum_{n=1}^{N} \log N(t_n|w^T x_n, \sigma_n^2)N(0,\Sigma) \\ &= \sum_{n=1}^{N} \log N(t_n|w^T x_n, \sigma_n^2) - w^T \Sigma^{-1} w \end{aligned}$$

Hence the MAP objective function is,

$$L_{MAP}(w) = \sum_{n=1}^{N} \log N(t_n|w^T x_n, \sigma_n^2) - w^T \Sigma^{-1} w$$

Part(c):
From Part(b), we got,

$$\begin{aligned} L(w) &= \sum_{n=1}^{N} \log N(t_n|w^T x_n, \sigma_n^2) \\ &= \sum_{n=1}^{N} \log \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp{(-\frac{(t_n - w^T x_n)^2}{2\sigma_n^2})} \\ &= \sum_{n=1}^{N} [\log \frac{1}{\sqrt{2\pi\sigma_n^2}} - \frac{(t_n - w^T x_n)^2}{2\sigma_n^2}] \end{aligned}$$

Hence we have to minimize the following function which is almost same as mean squared error except for the variance factor. Consider r_n as the weighting factor in weighted mean squared error and,

$$r_n \propto \frac{1}{\sigma_n^2}$$

Hence we have sum of squares error function to be,

$$E_d(w) = \sum_{n=1}^{N} \frac{r_n(t_n - w^T x_n)^2}{2}$$

To get the optimal value of w we maximise L(w) and hence upon differentiation of the above equation* with respect to w we will have,

$$\frac{dE_d(w)}{dw} = \sum_{n=1}^{N} r_n x_n (t_n - w^T x_n)$$

Equating it to zero and considering the matrix form of the quation we get,

$$X^T R X w = X^T R t$$

where R is the weighting matrix of size $n \times n$

$$w = (X^T R X)^{-1} X^T R t$$