FOML Assignment 1

Vishal Vijay Devadiga (CS21BTECH11061)
Abhay Kumar (BM21BTECH11001)

Question 3

The method of ordinary least squares assumes that there is constant variance in the errors (which is called **homoscedasticity**). The method of weighted least squares can be used when the ordinary least squares assumption of constant variance in the errors is violated (which is called **heteroscedasticity**).

Part A

Derive the expression of likelihood and prior for a heteroscedastic setting for a single data point with input x_n and output t_n.

Solution

In a heteroscedastic setting, the variance of the error is not constant. Let the variance of the error be σ_n^2.

$$y_i|x_i \sim \mathcal{N}(w^T x_i, \sigma_n^2)$$

For a single data point, the likelihood function is given by:

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

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

Part B

Provide the expression for the objective function that you will consider for the **ML** and **MAP** estimation of the parameters considering a data set of size N.

Solution

Objective function for **ML estimation** is the likelihood function. For a data set of size N, the likelihood function is given by:

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

$$\implies L(w) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(t_n - w^T x_n)^2}{2\sigma_n^2}\right)$$

$$\implies \log L(w) = \sum_{n=1}^{N} \log\left(\frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(t_n - w^T x_n)^2}{2\sigma_n^2}\right)\right)$$

$$\implies \log L(w) = \sum_{n=1}^{N} \left(-\frac{1}{2}\log 2\pi\sigma_n^2 - \frac{(t_n - w^T x_n)^2}{2\sigma_n^2}\right)$$

$$\implies \log L(w) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\sum_{n=1}^{N}\log\sigma_n^2 - \frac{1}{2}\sum_{n=1}^{N}\frac{(t_n - w^T x_n)^2}{\sigma_n^2}$$

Maximizing likelihood is equivalent to maximizing log likelihood. Therefore, the objective function for **ML estimation** is given by:

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

Since the other terms are constant, they can be ignored.

Objective function for **MAP estimation** is the posterior distribution. For a data set of size N, the posterior distribution is given by:

$$L(w) = p(w|t, x) = \frac{p(t|x, w)p(w)}{p(t|x)}$$

where $p(w)$ is the prior distribution.

$$\implies L(w) \propto p(t|x,w)p(w)$$

$$\implies L(w) \propto \prod_{n=1}^{N} \mathcal{N}(t_n|w^T x_n, \sigma_n^2)p(w)$$

$$\implies L(w) \propto \left(\prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(t_n - w^T x_n)^2}{2\sigma_n^2} \right) \right) p(w)$$

$$\implies \log L(w) \propto \sum_{n=1}^{N} \log \left(\frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(t_n - w^T x_n)^2}{2\sigma_n^2} \right) \right) + \log p(w)$$

$$\implies \log L(w) \propto \sum_{n=1}^{N} \left(-\frac{1}{2} \log 2\pi\sigma_n^2 - \frac{(t_n - w^T x_n)^2}{2\sigma_n^2} \right) + \log p(w)$$

$$\implies \log L(w) \propto -\frac{N}{2} \log 2\pi - \frac{1}{2} \sum_{n=1}^{N} \log \sigma_n^2 - \frac{1}{2} \sum_{n=1}^{N} \frac{(t_n - w^T x_n)^2}{\sigma_n^2} + \log p(w)$$

Maximizing posterior is equivalent to maximizing log posterior. Therefore, the objective function for **MAP estimation** is given by:

$$\boxed{-\frac{1}{2} \sum_{n=1}^{N} \frac{(t_n - w^T x_n)^2}{\sigma_n^2} + \log p(w)}$$

Since the other terms are constant, they can be ignored.

Part C

Show that the ML objective will result in a data set in which each data point t_n is associated with a weighting factor $r_n > 0$, so that the sum-of-squares error function becomes:

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

Find an expression for the solution w that minimizes this error function.

Solution

The ML objective function is given by:

$$\sum_{n=1}^{N} \frac{(t_n - y(x_n, w))^2}{\sigma_n^2}$$

where $y(x_n, w) = w^T \phi(x_n)$.

To minimize the sum of squares error function, we need to minimize the weighted sum of squares error function:

$$E_D(w) = \frac{1}{2} \sum_{n=1}^{N} \frac{(t_n - y(x_n, w))^2}{\sigma_n^2}$$

$$\implies E_D(w) = \frac{1}{2} \sum_{n=1}^{N} r_n (t_n - y(x_n, w))^2$$

where $r_n = \dfrac{1}{\sigma_n^2}$.

Let:

$$R = \begin{bmatrix} r_1 & 0 & \dots & 0 \\ 0 & r_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & r_N \end{bmatrix}$$

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

where:

$$t = [t_1, t_2, \dots, t_N]^T$$

$$\Phi = \begin{bmatrix} \phi(x_1)^T \\ \phi(x_2)^T \\ \vdots \\ \phi(x_N)^T \end{bmatrix}$$

$$\phi(x_n) = [\phi_1(x_n), \phi_2(x_n), \dots, \phi_M(x_n)]^T$$

To find the solution w that minimizes the error function, we need to find the value of w for which the derivative of the error function is zero.

$$\nabla_w E_D(w) = -\Phi^T R(t - \Phi w) = 0$$

$$\implies -\Phi^T R t + \Phi^T R \Phi w = 0$$

$$\implies \Phi^T R \Phi w = \Phi^T R t$$

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