Assignment - 1

Challa Akshay Santoshi, CS21BTECH11012
Cheekatla Hema Sri, CS21BTECH11013

October 8, 2023

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 heteroscedast icity).

a

The expression of likelihood and prior for a heteroscedastic setting for a single data point with input x_n and output t_n:

We have the linear model for heteroscedastic setting is given as follows

$$\mathbf{t} = \mathbf{w}^T \phi + \epsilon \implies \epsilon = \mathbf{t} - \mathbf{w}^T \phi \tag{1}$$

Here we are taking the following assumptions

- The error ϵ is in Gaussian distribution with mean 0 and variance σ^2 and it is i.i.d
- The data is taken from Gaussian distribution

Since we are assuming gaussian distribution, now the likelihood function for the given observation can be written as follows

$$\mathcal{L}(w) = p(t_n|x_n, w) \tag{2}$$

$$= \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\{\frac{(t_n - \mathbf{w}^T\phi)^2}{2\sigma_n^2}\} \tag{3}$$

Where w is the parameter vector and ϕ is the feature vector that are taken from the equation (1) and σ_n^2 is the variance of the datapoint $\{x_n, t_n\}$

Now let us see prior distribution over the polynomial coefficients $\mathbf{w}$. As we have assumed guassian distribution, the expression is as follows

$$p(\mathbf{w}|\alpha) = \mathcal{N}(\mathbf{w}|0, \alpha^{-1}\mathbf{I}) \tag{4}$$

$$= (\frac{\alpha}{2\pi})^{\frac{M+1}{2}} e^{-\frac{\alpha}{2}\mathbf{w}\mathbf{w}^T} \tag{5}$$

Where α is the precision of the distribution and M is the order of the polynomial and $\mathbf{w}$ is the parameter vector

b

Expression for objective function for ML on considering a data set of size N:

In the above problem we have derived likelihood for a single data point. Now we have N data points with input $\mathbf{x}$ and outputs $\mathbf{t}$. For simplicity, we are assuming that these data points are in gaussian distribution and are i.i.d. Hence the likelihood can be written as follows

$$\mathcal{L}(\mathbf{w}) = p(\mathbf{t}|\mathbf{x}, w) \tag{6}$$
$$= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp \frac{(t_n - w^T\phi_n)^2}{2\sigma_n^2} \tag{7}$$
$$\tag{8}$$

(symbols represent the same as in the previous problem) Now consider negative log of the above expression

$$-log(\mathcal{L}(\mathbf{w})) = -log(\prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp \frac{(t_n - w^T\phi_n)^2}{2\sigma_n^2}) \tag{9}$$
$$= -\sum_{n=1}^{N} [\frac{1}{2} ln2\pi + ln\sigma_n + \frac{(t_n - w^T\phi_n)^2}{2\sigma_n^2}] \tag{10}$$
$$\tag{11}$$

Since the first two terms are constants, the negative log likelihood depends on the third term only. Hence the objective function of ML is written as follows

$$\frac{1}{2}\sum_{n=1}^{N}\left(\frac{1}{\sigma_n^2}(t_n - w^T\phi_n)^2 + ln\sigma_n^2\right) \tag{12}$$

Expression for objective function for MAP estimation:
Using Bayes' theorem, the posterior distribution for w is proportional to the product of the prior distribution and the likelihood function

$$p(\mathbf{w}|\mathbf{x}, \mathbf{t}, \alpha, \beta) \propto p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta)p(\mathbf{w}|\alpha) \tag{13}$$

By taking the negative logarithm of the above expression, we can find the MAP which is the minimum of the following is resultant where we ignored the constant terms and substituted the likelihood and prior functions that are computed previously in the above equation

$$\frac{1}{2}\sum_{n=1}^{N}\frac{1}{\sigma_n^2}(t_n - \mathbf{w}^T\phi_n)^2 + \frac{\alpha}{2}\mathbf{w}^T\mathbf{w} \tag{14}$$

c

In the previous problem, we have derived the following ML objective

$$E_D(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\frac{1}{\sigma_n^2}(t_n - \mathbf{w}^T\phi_n)^2 \tag{15}$$
$$\text{let } r_n = \frac{1}{\sigma_n^2}, r_n > 0 \tag{16}$$
$$\implies E_D(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N} r_n(t_n - \mathbf{w}^T\phi_n)^2 \tag{17}$$

Since r_n is inverse of variance of the data point, which varies with each data point, we can say that the ML objective is associated with the weighted factor for each data point and the above is sum-of-square error function. Now let us find the expression for w that minimizes the above error function. For that, take derivative with respect to $\mathbf{w}$ and equate it to zero.

$$\nabla E_D(\mathbf{w}) = \sum_{n=1}^{N} r_n(t_n - \mathbf{w}\phi_n)(-\phi_n) = 0 \tag{18}$$

$$\implies \sum_{n=1}^{N} r_n t_n \phi_n = \sum_{n=1}^{N} r_n \phi_n^T \phi_n \mathbf{w} \tag{19}$$

$$\implies \mathbf{w} = (\sum_{n=1}^{N} r_n \phi_n^T \phi_n)^{-1}(\sum_{n=1}^{N} r_n t_n \phi) \tag{20}$$

$$\implies \mathbf{w} = (\Phi^T R \Phi)^{-1}(\Phi^T R \mathbf{t}) \tag{21}$$

Where R is an $n \times n$ matrix such that,

$$R_{ij} = \begin{cases} r_n, & i = j \\ 0, & i \neq j \end{cases} \tag{22}$$