Name: Mayuresh .R. Dindorkar.
Roll No: CS23MTECH14007.

ML Assignment No. 1

Page No.
Date

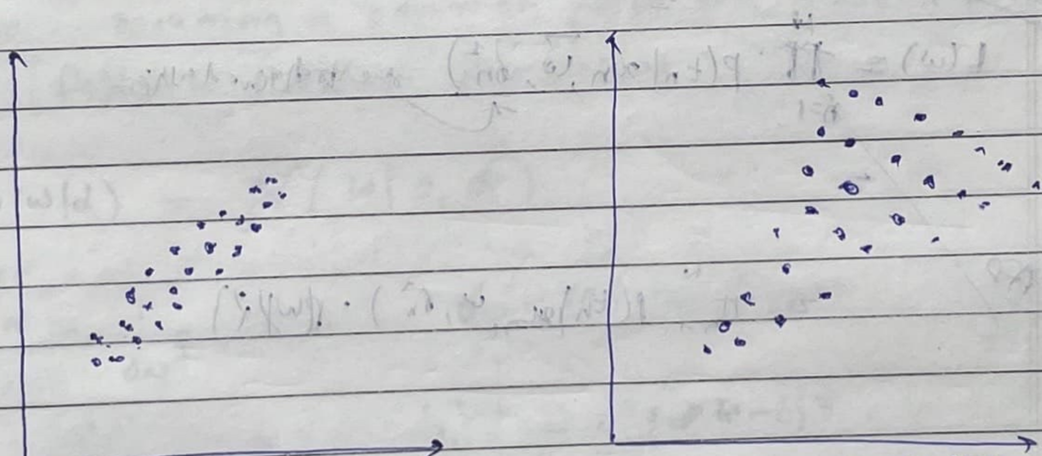
Name: sanyam kaul.
Roll NO: CS23MTECH14011.

Q.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 errors is violated (which is called heteroscedasticity).

Q.3) (a) Derive the expression of likelihood and prior of a heteroscedastic setting for a single data point with input x_n and output t_n.

→ (1) The ordinary least squares method assumes constant variance in errors. ie, the residuals are spread evenly around fitted line. (homoscedastic scenario)

(2) Whereas, in case of heteroscedastic scenario, the residuals show non-uniform spread pattern as we move along x - axis. We use ordinal regression in such case.



Homoscedasticity   Heteroscedasticity

Page No.
Date

- Consider a single data point x_n and t_n be its corresponding output.

Assuming that this data point is drawn independantly from gaussian distribution, the likelihood is given by,

(*) Likelihood equation:-

$$p(t_n \mid \phi(x_n), \vec{\omega}, \delta_n^2) = \mathcal{N}(t_n \mid \omega^T \cdot \phi(x_n), \delta_n^2).$$

Where,
$\phi(x_n)$ = Basis function for x.
$\vec{\omega}$ = weight vector / parameter vector.
δ_n^2 = Variance of error for data point x_n.

$$\boxed{\therefore\ p(t_n \mid \phi(x_n), \vec{\omega}, \delta_n^2) = \frac{1}{\sqrt{2\pi\delta_n^2}} \cdot e^{-\frac{(t_n - \omega^T\phi(x_n))^2}{2\delta_n^2}}}$$

(*) Prior:- Assuming a common prior for model's parameters follow a guassian distribution.

$$p(\omega \mid \alpha) = \mathcal{N}(\omega \mid 0, \alpha^{-1})$$

where,
$\alpha = \dfrac{1}{\delta_\omega^2}$ ...... δ_ω = Variance of prior

$$\therefore p(\omega \mid \alpha) = \left(\frac{\alpha}{2\pi}\right)^{1/2} \cdot e^{-\frac{\alpha}{2}(\omega - 0)^2}$$

$$\boxed{p(\omega \mid \alpha) = \left(\frac{\alpha}{2\pi}\right)^{1/2} \cdot e^{-\frac{\alpha}{2}(\omega^T \omega)}}$$

Page No.
Date

Q.3) (b) Provide expression for objective function that you will consider for ML and MAP estimation of parameters considering a dataset of size N.

→

1) MLE:- The maximum likelihood estimation objective function is given by,

$$w_{ML} = \underset{w}{\arg\max} \prod_{n=1}^{N} p(t_n | x_n, \vec{w}, \sigma_n^2)$$

(*) In heteroscedastic setting, every datapoint with have different variance denoted by σ_n^2.

$$w_{ML} = \underset{w}{\arg\max} \prod_{n=1}^{N} \mathcal{N}(t_n | w^T \cdot \phi(x_n), \sigma_n^2)$$

(*) ~~MAP~~ Lets expand the likelihood equation, to obtain error function $E(w)$.

$$\therefore \log \prod_{n=1}^{N} p(t_n | x_n, \vec{w}, \sigma_n^2) = \log \left[\prod_{n=1}^{N} \mathcal{N}(t_n | w^T \cdot \phi(x_n), \sigma_n^2) \right]$$

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

$$= \sum_{n=1}^{N} \log \left[\left(\frac{1}{2\pi\sigma_n^2} \right)^{1/2} e^{-\frac{(t_n - w^T \phi(x_n))^2}{2\sigma_n^2}} \right]$$

$$\therefore \text{log likelihood} = -\frac{1}{2} \sum_{n=1}^{N} \frac{(t_n - w^T \cdot \phi(x_n))^2}{\sigma_n^2} - \frac{N}{2} \log(2\pi) - \frac{N}{2} \cdot \log(\sigma_n)^2$$

Page No.
Date

We know that, error function $E(w)$ is equal to negative log likelihood.

$$\therefore E(\omega) = \frac{1}{2} \sum_{n=1}^{N} \frac{(t_n - \omega^T.\phi(x_n))^2}{\sigma_n^2} + \frac{N}{2}\log(2\pi) + \frac{N}{2}\log(\sigma_n^2)$$

Neglecting the last 2 terms as they are not dependant on ω.

$$\boxed{\therefore E(\omega) = \frac{1}{2} \sum_{n=1}^{N} \frac{(t_n - \omega^T.\phi(x_n))^2}{\sigma_n^2}}$$

$$\boxed{E(\omega) = \frac{1}{2} \sum_{n=1}^{N} r_n (t_n - \omega^T.\phi(x_n))^2} \quad \text{--- (1)}$$

..where $r_n = \frac{1}{\sigma_n^2}$.= weighting factor

(2) MAP :- The objective function for maximum aposteriori estimation is given by,

$$\omega_{MAP} = \underset{\omega}{\text{argmax}} \text{ (likelihood)} \cdot \text{(prior)}.$$

$$\boxed{\omega_{MAP} = \underset{\omega}{\text{argmax}} \prod_{n=1}^{N} p(t_n | x_n, \vec{\omega}, \sigma_n^2) \cdot p(\omega | \alpha)}$$

$$\therefore \omega_{MAP} = \underset{\omega}{\text{argmax}} \log \left[\prod_{n=1}^{N} p(t_n | x_n, \vec{\omega}, \sigma_n^2) \right] + \log p(\omega | \alpha)$$

$$= \underset{\omega}{\text{argmin}} - \log \left[\prod_{n=1}^{N} p(t_n | x_n, \vec{\omega}, \sigma_n^2) \right] - \log p(\omega | \alpha).$$

$$= \underset{\omega}{\text{argmin}} - \sum_{n=1}^{N} \log p(t_n | x_n, \vec{\omega}, \sigma_n^2) - \log p(\omega | \alpha) \quad \text{---- (2)}$$

Lets solve $\log p(\omega|\alpha)$:-

$$\log p(\omega|\alpha) = \log\left[\mathcal{N}(0, \omega, \alpha^{-1})\right]$$

$$= \log\left[\left(\frac{\alpha}{2\pi}\right)^{1/2} e^{-\frac{\alpha}{2}(\omega^T.\omega - 0)^2}\right]$$

$$= \frac{N}{2}\log\left(\frac{\alpha}{2\pi}\right) - \frac{\alpha}{2}(\omega^T\omega)$$

$$\therefore \boxed{\log p(\omega|\alpha) = \frac{-\alpha}{2}(\omega^T.\omega)} \quad \text{------} (3)$$

From (3),

$$\therefore \omega_{MAP} = \underset{\omega}{\text{argmin}} - \Sigma \log p(t_n | x_n, \vec{\omega}, \delta_n^2) - \log p(\omega|\alpha)$$

From (1) & (2), we get;

$$\boxed{\omega_{MAP} = \frac{1}{2}\sum_{n=1}^{N} r_n(\omega^T.\phi(x_n) - t_n)^2 + \frac{\alpha}{2}(\omega^T.\omega)}$$

Page No.
Date

Q.3 (C): Show that ML objective will result in a dataset in which each data point t_n is associated with weighting factor $r_n > 0$, so the sum of squares error function becomes

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

Find the expression for solution ω that minimizes the error function.

→

ML objective function for heteroscedastic conditions is given by,

$$\omega_{ML} = \underset{\omega}{\arg\max} \prod_{n=1}^{N} p(t_n \mid x_n, \vec{\omega}, \sigma_n^2)$$

σ_n denotes variance for datapoint x_n.

Assuming the data is independantly drawn from gaussian distribution, likelihood is given by,

$$\prod_{n=1}^{N} p(t_n \mid x_n, \vec{\omega}, \sigma_n^2) = \prod_{n=1}^{N} \mathcal{N}(t_n \mid \omega^T\cdot\phi(x_n), \sigma_n^2)$$

$$\text{log likelihood (LL)} = \sum_{n=1}^{N} \log\left[\left(\frac{1}{2\pi\sigma_n^2}\right)^{1/2} e^{-\frac{(t_n - \omega^T\cdot\phi(x_n))^2}{2\sigma_n^2}}\right]$$

$$\therefore LL = -\frac{1}{2}\sum_{n=1}^{N}\frac{(t_n - \omega^T\cdot\phi(x_n))^2}{\sigma_n^2} - \frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma_n)^2$$

We know that, error function $E(\omega)$ is equal to negative log likelihood.

$$\therefore E(\omega) = \frac{1}{2}\sum_{n=1}^{N}\frac{(t_n - \omega^T\cdot\phi(x_n))^2}{\sigma_n^2} + \frac{N}{2}\log(2\pi) + \frac{N}{2}\log(\sigma_n)^2.$$

Omitting last 2 terms as they doesnot depend on ω.

Page No.
Date

$$\therefore E(\omega) = \frac{1}{2} \sum_{n=1}^{N} \frac{(t_n - \omega^T \phi(x_n))^2}{\sigma_n^2}$$

Let, weighting factor, $r_n = 1/\sigma_n^2$.

$$\boxed{\therefore E(\omega) = \frac{1}{2} \sum_{n=1}^{N} r_n (t_n - \omega^T \cdot \phi(x_n))^2}$$

To find w_{ML} that minimizes the error function, we take gradient of LL and equate it to zero.

$\therefore \nabla(LL) = 0$.

Solving this we get, $\boxed{w_{ML} = (\Phi^T R \Phi)^{-1} \Phi^T \cdot R \cdot t}$

where Φ = design matrix,

R = diagonal matrix of weighting factors

t = ~~output~~ vector of outputs.

(*) Observations:-

1) For each data point x_n, the weighing factor r_n is inversely proportional to variance of error for that datapoint.

2) The Maximum likelihood objective function for weighted least squares gives sum of squares error function where every datapoint x_n is associated with corresponding r_n. (where r_n = weighting factor).