Q3 a) Heteroscedasticity setting:

Method of weighted least squares can be used when the ordinary least square assumption of constant variance in the errors is violated.

In this setting, the variance of the error term for a single data point is not constant and may depend on input value x_n.

So for linear regression, we assume that t_n is generated from Gaussian (Normal) distribution with mean μ & varying variance σ^2 both of which depends on x_n (μ, σ^2 are functions of x_n)

$$\therefore \quad \underbrace{t_n}_{\text{Observed output}} \sim N(\underbrace{\mu}_{\text{Mean}}, \underbrace{\sigma^2}_{\text{Variance}})$$

here $\mu = W^T x_n$ (for linear regression model)

Now, pdf of Gaussian distribution is :-

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

here $x = t_n$

μ, σ^2 are dependent on x_n

$$\therefore \mathbb{P}(t_n | x_n, W) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t_n-\mu)^2}{2\sigma^2}\right)$$

$$\boxed{\mathbb{P}(t_n | x_n, W) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t_n-W^T x_n)^2}{2\sigma^2}\right)}$$

Likelihood equation for a single data point.

For multivariate Gaussian prior for θ:

$$W \sim N(\mu_0, \Sigma_0)$$

covariance matrix

mean vector of prior distribution

pdf of multivariate Gaussian distribution

$$f(x) = \frac{1}{\sqrt{(2\pi)^n}\sqrt{|\Sigma|}} \exp\left(\frac{-1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

where n is the dimension

here $x = W$, $\mu = \mu_0$, $\Sigma = \Sigma_0$

$$\boxed{\mathbb{P}(W) = \frac{1}{\sqrt{2\pi^n}\sqrt{|\Sigma_0|}} \exp\left(\frac{-1}{2}\frac{(W-\mu_0)^T}{\Sigma_0}(W-\mu_0)\right)}$$

Prior equation

Q3 b) ML objective function

$$\mathbb{P}(t_n | x_n, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t_n-\mu)^2}{2\sigma^2}\right)$$

where μ & σ^2 are dependent on x_n

Dataset size $= N$ i.e. $n=1$ to $n=N$

The likelihood of observing the entire dataset is the product of the likelihoods for all individuals data points

$$P(t, x, \theta) = \prod_{n=1}^{N} P(t_n | x_n, \theta)$$

(assuming that the data points are independent)

log-likelihood :-

$$\log P(t|x,\theta) = \sum_{n=1}^{N} \left[-\frac{1}{2} \log(2\pi\sigma^2) - \frac{(t_n - \mu)^2}{2\sigma^2} \right]$$

ML objective function can be defined as negative log-likelihood and minimize it.

$$J_{ML}(\theta) = -\log P(t|x,\theta)$$

$$= -\sum_{n=1}^{N} \left[-\frac{1}{2} \log(2\pi\sigma^2) - \frac{(t_n-\mu)}{2\sigma^2} \right]$$

MAP (Maximum A Posteriori) objective function.

$$P(\theta) = \frac{1}{\sqrt{(2\pi)^n} \sqrt{|\Sigma_0|}} \exp\left(-\frac{1}{2} \frac{(\theta - \mu_0)^T}{\Sigma_0} (\theta - \mu_0) \right)$$

Posterior distribution $\propto$ likelihood $\times$ prior.

$$P(\theta | t, x) \propto P(t|x, \theta) \cdot P(\theta)$$

$$\log P(\theta | t, x) \propto \log P(t|x,\theta) + \log P(\theta)$$

MAP objective function can be defined as negative log-posterior

$$J_{MAP}(\theta) = -\log P(\theta | t, x)$$

$$J_{MAP}(\theta) = -\left(\sum_{n=1}^{N}\left[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(t_n-\mu)^2}{2\sigma^2}\right]\right)$$

$$-\left(-\frac{1}{2}\log(|\Sigma_0|) - \frac{1}{2}\frac{(\theta-\mu_0)^T(\theta-\mu_0)}{\Sigma_0}\right)$$

Q3 c) ML objective function

$$J_{ML}(w) = -\log P(t|x,w)$$

likelihood for a single data point n:

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

maximizing this.

$$J_{ML}(w) = -\log\left(\prod_{n=1}^{N} P(t_n|x_n,w)\right)$$

$$= -\sum_{n=1}^{N}\log P(t_n|x_n,w)$$

$$J_{ML}(w) = -\sum_{n=1}^{N}\left(-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(t_n-w^T\theta)^2}{2\sigma^2}\right)$$

$$= \frac{1}{2}\cdot\sum_{n=1}^{N}\left(\log(2\pi\sigma^2) + \frac{(t_n-w^T\theta)^2}{\sigma^2}\right)$$

$r_n = \frac{1}{\sigma^2}$ for optimum value.

$\therefore$ the sum of squares error function will become.

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