Q.3

a) → We know that likelihood for ordinary least square given by

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

→ We know that σ^2 is constant but for weighted least square σ^2 will be differant for each point.

→ So likelihood function for the heteroscedastic will be $P(t_n \mid x_n, w)$

$$P(t_n \mid x_n, w) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(t_n - w\phi(x_n))^2}{2\sigma_n^2}\right)$$

← (1)

Here σ_n is not constant

For ordinal least square prior is of the form $P(w)$ by in this prior will be of the for $P(w, \sigma_n^2)$

b. we know the eq.

$$y_n = w^T \phi(x_n) + \epsilon^* \quad -(2)$$

→ where ϵ^* is not constant

→ let $\epsilon^* = \frac{\epsilon}{\sqrt{\gamma_n}}$ where, ϵ is constant and $\sqrt{\gamma_n}$ is some weight (parameter)

From the eq(2)

$$y_n = w^T \phi(x_n) + \epsilon^*$$

$$y_n = w^T \phi(x_n) + \frac{\epsilon}{\sqrt{\gamma_n}}$$

$$\sqrt{\gamma_n}\, y_n = \sqrt{\gamma_n}\, w^T \phi(x_n) + \epsilon \rightarrow (3)$$

Now eq(3) becomes OLS for each data poin t_n is associated with weighting factor $\gamma_n > 0$

From the eq(3) objective function for ML given by

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

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{\gamma_n (t_n - w\phi(x_n))^2}{2\sigma^2}\right)$$

for all N data points

we got $= \prod_{n=1}^{N} \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) \exp\left(\frac{\gamma_n (t_n - w\phi(x_n))^2}{2\sigma^2}\right)$

$$E_{ML}(w, \sigma_n^2) = \frac{1}{2}\sum_{n=1}^{N}\left[\log(2\pi\sigma^2) + \frac{\gamma_n(t_n - w\phi(x_n))^2}{2\sigma^2}\right] \quad -(4)$$

$$E_{MAP}(w, \sigma^2) = E_{ML}(w, \sigma_n^2) * P(w, \sigma_n^2)$$

c. From the equation 34 we can see that sum of squares error func.

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

where dataset in which each data point t_n is associated with weighting factor $\gamma_n > 0$

differ tiating with respect to w from eq (4). we can give that

$$\frac{d}{dw} E_{ML}(w) = \frac{1}{2}\frac{d}{dw}\left(\sum_{n=1}^{N}\left[\log(2\pi\sigma^2) + \frac{r_n(t_n - w\phi(x_n))^2}{\sigma^2}\right]\right)$$

$$= R^T(Y - W^T X)^T X = 0$$

(where R^T is vector of weights r_i and $r_i > 0$)

$$= (YR - W^T XR)^T X = 0$$

$$= (YRX^T - W^T XRX^T) = 0$$

$$YRX^T = W^T XRX^T$$

$$YRX^T(XRX^T)^{-1} = W^T$$

$$\therefore \boxed{W = (X^T R X)^{-1} X^T R Y}$$

This is the solution of W that minimizes this error function.