3. (a). As know that in a heteroscedastic setting,
The Variance of Error is "Not constant" and also
It varies with each point.

The likelihood function as $P(t_n | x_n, W, \sigma_n^2)$
Where, The Notations of the likelihood function are,

t_n = Observed Output
x_n = Input for data point n.
W = Parameters of the Model.
σ_n^2 = Variance of the Error

Now, Assuming that the Errors are normally distributed with mean $y(x_n, W)$ and Variance σ_n^2,

Now, The likelihood for a single data point is,

$$P(t_n | x_n, W, \sigma_n^2) = N(t_n | x_n, W), \sigma_n^2)$$

As know that,
Here $N(\mu, \sigma^2)$ represents mean μ and Variance σ^2.

$$P(t_n | x_n, W, \sigma_n^2) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(t_n - y(x_n, W))^2}{2\sigma_n^2}\right)$$

(b). To provide the objective function for the ML and MAP estimation of the Parameters Considering a data set of size N.

For ML Estimation:
Let us Assume data points are independent.

The joint likelihood for all N data

$$L(w, \sigma_1^2, \sigma_2^2, \dots \sigma_N^2) = \prod_{n=1}^{N} p(t_n | x_n, w, \sigma_n^2)$$

$$-\ln(L) = -\ln\left(\prod_{n=1}^{N} p(t_n | x_n, w, \sigma_n^2)\right)$$

$$= -\sum_{n=1}^{N} \ln p(t_n | x_n, w, \sigma_n^2).$$

$$= -\sum_{n=1}^{N} \ln\left(\frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(t_n - y(x_n, w))^2}{2\sigma_n^2}\right)\right)$$

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

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

Here, The first term Constant minimising

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

Which minimises the Error function, thus reducing the Error.

Now, The Error function for ML estimate is:-

$$E(w, \sigma_1^2, \sigma_2^2, \dots \sigma_N^2) = \frac{1}{2}\sum_{n=1}^{N}\left(\frac{(t_n - y(x_n, w))^2}{\sigma_n^2}\right)$$

Now for ~~Maximum~~ MAP Estimation:-

Now Introducing priors over the parameter w and the Variances σ_n^2.

Let $p(w)$ be the prior for w. $\therefore$ Bayes Theorem

$$E_{MAP}(w) = \arg\max \prod_{x_i \in x} p(x_i | w) \cdot p(w).$$

The objective function for MAP Estimation is the Negative log posterior:

$$E_{MAP}(W, \sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2) = -\ln p(\omega) + E(\omega, \sigma_1^2, \sigma_2^2, \ldots \sigma_N^2)$$

(c). To show that the ML objective results in weighting factors $\gamma_n > 0$,

$$E(w, \sigma_1^2, \sigma_2^2, \ldots \sigma_N^2) = \frac{1}{2} \sum_{n=1}^{N} \left(\frac{(t_n - y(x_n, w))^2}{\sigma_n^2} \right)$$

Let $y(x_n, w) = w^T \phi(x_n)$, $\gamma_n = \frac{1}{\sigma_n^2}$

We observe that $\gamma_n > 0$

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

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

$$\therefore E_D(W) = \frac{1}{2} \| (y - X^T w) \|^2$$

Now, To Minimise the Error function.

$$\frac{\partial}{\partial w}(E_D(\omega)) = 0$$

$$\frac{\partial}{\partial w}\left(\frac{1}{2} \| (y - X^T w) \|^2 \right) = 0$$

$$\frac{1}{2}\left[\frac{d}{d\omega} ((y - X^T \omega)^T (y - X^T w)) \right] = 0$$

$$\frac{1}{2}(-2(y - X^T w)^T X^T) = 0$$

$$(X(y - X^T w))^T = 0$$

$$Xy - XX^T w = 0$$

$$w_{ML} = (XX^T)^{-1} Xy.$$

Where $y = \left[\frac{t_1}{\sqrt{\gamma_1}}, \frac{t_2}{\sqrt{\gamma_2}}, \ldots, \frac{t_N}{\sqrt{\gamma_N}}\right]^T$

Now, The Design Matrix.

$$X = [x_1, x_2, \ldots, x_N]_{(D+1 \times N)}.$$

$$x_i = \left[\frac{1}{\sqrt{\gamma_i}}, \frac{\phi(x_{i1})}{\sqrt{\gamma_i}}, \ldots, \frac{\phi(x_{iD})}{\sqrt{\gamma_i}}\right]^T$$