3b) Consider linear regression where each data point may have different Gaussian measurement noise.

This implies $y_i \sim N(wx_i - b, \sigma_i^2)$

Each σ_i represents gaussian measurement for each data point.

The probability density fn for the ith response is

$$p(y_i \mid w, b, \sigma_i^2) = \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left[-\frac{1}{2\sigma_i^2}(y_i - wx_i + b)^2\right]$$

The joint pdf of $y_1, \dots, y_m$ is:

$$p(y_1, \dots, y_m \mid w, b, \sigma_1^2, \dots, \sigma_m^2) = \left(\frac{1}{\sqrt{2\pi}}\right)^m \prod_{i=1}^{m} \frac{1}{\sigma_i} \exp\left[\frac{-1}{2} \sum_{i=1}^{m} \frac{(y_i - wx_i + b)^2}{\sigma_i^2}\right]$$

Thus, the log likelihood fn follows:

$$\ell(w, b, \sigma_1^2, \dots, \sigma_m^2) = \frac{-m}{2}\ln(2\pi) - \sum_{i=1}^{m}\ln(\sigma_i) - \frac{1}{2}\sum_{i=1}^{m}\frac{(y_i - wx_i + b)^2}{\sigma_i^2}$$

Maximizing the log likelihood is equivalent to minimizing the negative log likelihood. Since w, b are the only parameters we can alter, we only consider the last term.

So, we can maximize $\ell(w, b, \sigma_1^2, \dots, \sigma_m^2)$ by minimizing

$$L(w, b) = \sum_{i=1}^{m} \frac{1}{\sigma_i^2}(y_i - wx_i + b)^2$$

Let $r_i = \frac{1}{\sigma_i^2}$. Then the negative log likelihood is equivalent to our objective function.

Furthermore, the variance of each measurement noise is

$$\sigma_i^2 = \frac{1}{r_i}$$

A1

3a) $$L(w,b) = \sum_{n=1}^{m} r_n (y_n - wx_n + b)^2$$

We minimize the objective fn by taking partial derivatives w.r.t. w and b. and setting them to zero.

$$\frac{\partial L(w,b)}{\partial w} = \sum_{n=1}^{m} 2r_n(-x_n)(y_n - wx_n + b) \quad (1)$$

$$\frac{\partial L(w,b)}{\partial b} = \sum_{n=1}^{m} 2r_n(y_n - wx_n + b) = 0 \quad (2)$$

Setting (1) to zero:

$$0 = \sum_{n=1}^{m} r_n x_n (y_n - wx_n + b)$$

$$\Rightarrow 0 = \sum_{n=1}^{m} r_n x_n y_n - w\sum_{n=1}^{m} r_n x_n^2 + b\sum_{n=1}^{m} r_n x_n$$

sub (3):

$$\Rightarrow 0 = \sum_{n=1}^{m} r_n x_n y_n - w\sum_{n=1}^{m} r_n x_n^2 + \left(w\bar{x} - \frac{\sum_{n=1}^{m} r_n y_n}{m}\right)\sum_{n=1}^{m} r_n x_n$$

Setting (2) to zero:

$$0 = \sum_{n=1}^{m} r_n (y_n - wx_n + b)$$

$$\Rightarrow 0 = \sum_{n=1}^{m} r_n y_n - \sum_{n=1}^{m} wx_n + mb$$

$$\Rightarrow \hat{b} = \hat{w}\bar{x} - \sum_{n=1}^{m} \frac{r_n y_n}{m} \quad (3)$$

$$\Rightarrow w\left[\sum_{n=1}^{m} r_n x_n^2 - \bar{x}\sum_{n=1}^{m} r_n x_n\right] = \sum_{n=1}^{m} r_n x_n y_n - \frac{\sum_{n=1}^{m} r_n y_n \sum_{n=1}^{m} r_n x_n}{m}$$

$$\hat{w} = \frac{\sum_{n=1}^{m} r_n x_n y_n - \frac{\sum_{n=1}^{m} r_n y_n \sum_{n=1}^{m} r_n x_n}{m}}{\sum r_n x_n^2 - \bar{x}\sum_{n=1}^{m} r_n x_n} \quad (4)$$

Thus, (3) and (4) are a closed form solution for the objective fn.