3.(a) Let $\bar{x}_n = \begin{pmatrix} 1 \\ x_n \end{pmatrix}$

$$L(w,b) = \sum_{n=1}^{m} r_n \left(y_n - \begin{bmatrix} b \\ w \end{bmatrix} \bar{x}_n\right)^2$$

Assume w and x, y have attribute m.

Let $j = 1$ to $m+1$, $\bar{w} = \begin{bmatrix} b \\ w \end{bmatrix}$

$$\frac{\partial L(w,b)}{\partial \bar{w}_j} = \sum_{n=1}^{m} 2 r_n (y_n - \bar{w}\bar{x}_n)(\bar{x}_n)_j$$

When $\frac{\partial L(w,b)}{\partial \bar{w}_j} = \vec{0}$ $\forall j = 1$ to $m+1$, $L(w,b)$ have minimum value.

$$\sum_{n=1}^{m} 2 r_n (y_n - \bar{w}\bar{x}_n)(\bar{x}_n)_j = \vec{0} \quad \forall j$$

$$\sum_{n=1}^{m} r_n (y_n - \bar{w}\bar{x}_n)\bar{x}_n = \vec{0}$$

$$\sum_{n=1}^{m} r_n y_n \bar{x}_n = \sum_{n=1}^{m} r_n \bar{w}\bar{x}_n \bar{x}_n$$

Therefore $A \begin{bmatrix} b \\ w \end{bmatrix} = b'$ $\quad A = \sum_{n=1}^{m} r_n \bar{x}_n \bar{x}_n$ $\quad b' = \sum_{n=1}^{m} r_n y_n \bar{x}_n$

(b) Let $y_n = w^T x_n - b + \varepsilon$

Since $\varepsilon \sim N(0, \sigma)$

$$Pr(y \mid x, w^T, \sigma) = N(y \mid w^T x - b, \sigma)$$

$$= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x - w^T x + b)^2}{2\sigma^2}}$$

So the negative log likelihood is

$$-\log \Pr(y \mid x, w^T, \sigma) = -\sum_{n=1}^{N} \log\left(\frac{1}{\sqrt{2\pi}\,\sigma_n} e^{-\frac{(y_n - w^T x_n + b)^2}{2\sigma_n^2}}\right)$$

$$= -\sum_{n=1}^{N}\left[\log\left(\frac{1}{\sqrt{2\pi}\,\sigma_n}\right) + \log\left(e^{-\frac{(y_n - w^T x_n + b)^2}{2\sigma_n^2}}\right)\right]$$

To minimize $-\log \Pr(y \mid x, w^T, \sigma)$,

$$(w, b) = \operatorname{argmin}_{w,b} -\sum_{n=1}^{N}\left[\log\left(\frac{1}{\sqrt{2\pi}\,\sigma_n}\right) + \log\left(e^{-\frac{(y_n - w^T x_n + b)^2}{2\sigma_n^2}}\right)\right]$$

$$= \operatorname{argmin}_{w,b} -\sum_{n=1}^{N} \log\left(e^{-\frac{(y_n - w^T x_n + b)^2}{2\sigma_n^2}}\right)$$ since first term doesn't have w or b

$$= \operatorname{argmin}_{w,b} \sum_{n=1}^{N} \frac{(y_n - w^T x_n + b)^2}{2\sigma_n^2}$$

Which is the same as objective in question (a).

The variance of our measurement noise is $\sigma_n^2 \propto \frac{1}{r_n}$