(3) Least squares assumes constant variance for the problem

$$y_i = w^T x_i + \epsilon_i \;;\; \epsilon_i \text{ gaussian with } \mu=0, \sigma^2$$

If the variance (σ^2) is not same for all the cases then uses weighted least squared approach. Let us consider the variance as $\left(\frac{\sigma^2}{r_i}\right)$ for weighted least squares.

Loss function,

$$L(w) = \frac{1}{\sqrt{2\pi\left(\frac{\sigma^2}{r_i}\right)}} \exp\left(-\frac{1}{2\left(\frac{\sigma^2}{r_i}\right)}\left(y_i - w^T x_i\right)^2\right)$$

For 'N' ~~observations~~ independent observations

$$L(y|x,w) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\left(\frac{\sigma^2}{r_i}\right)}} \exp\left(\frac{-1}{2\left(\frac{\sigma^2}{r_i}\right)}\left(y_i - w^T x_i\right)^2\right)$$

$$\cancel{\log\left(L(y|x,w)\right) = \sum_{i=1}^{N}\left[\frac{1}{2\pi\left(\frac{\sigma^2}{r_i}\right)}\right]} + \cancel{a}$$

$$\log\left(L(y|x,w)\right) = \sum_{i=1}^{N}\left[\log\left(\frac{1}{2\pi\left(\frac{\sigma^2}{r_i}\right)}\right) + \frac{-1}{2\left(\frac{\sigma^2}{r_i}\right)}\left(y_i - w^T x_i\right)^2\right]$$

$$\boxed{L(y|x,w) = \sum_{i=}^{N}\left[-\frac{\log(2\pi)}{2} - \log\left(\frac{\sigma^2}{r_i}\right)\frac{1}{2} - \frac{r_i}{2\sigma^2}\left(y_i - w^T x_i\right)^2\right]}$$

optimised 'w^*' for maximum likelihood is

$$w^* = \arg_w \max \left(+ L(y \mid x, w) \right)$$

$$\therefore w^* = \arg_w \min \left(\sum_{i=1}^{N} \left(\frac{1}{2} \log(2\pi) + \log(\sigma^2) - \log(r_i) \right) + \frac{r_i}{2\sigma^2} (y_i - w^T x_i)^2 \right)$$

$$w^* = \arg_w \min \left(\frac{r_i}{2\sigma^2} \sum_{i=1}^{N} (y_i - w^T x_i)^2 \right)$$

weight least squared apprach loss in terms of matrix can be written as

$$L = \frac{1}{2} \sum_{i=1}^{N} r_i (y_i - w^T x_i)^2$$

$$\frac{\partial L}{\partial w} = \frac{\partial}{\partial} \left[\frac{r_i}{2} \sum_{i=1}^{N} (y_i - w^T x_i)(y_i - w^T x_i)^T \right]$$

$$= \frac{r_i}{2} \sum_{i=1}^{N} \frac{\partial}{\partial 2} \left[(y_i - w^T x_i)(y_i - w^T x_i)^T \right]$$

$$= \frac{r_i}{2} \left[-2(x_i^T y_i) + 2 x_i x_i^T w \right] = 0$$

$$\cancel{\gamma}_i x_i^T y_i = x_i x_i^T w$$

$$\therefore \boxed{w^* = \left(x_i \cancel{\gamma}_i x_i^T \right)^{-1} r_i x_i^T y_i}$$

Solving for the prior of the weighted least ~~to~~ Squared Problem. Assuming the weights $(w_0, w_1, \dots w_n)$ to follow a normal distribution of mean $\mu = \mu_i$ and $\sigma^2 = \tau_i^2$; $N(\mu_{i}, \dots$

The error function follows inverse gamma function, this is because the error variance cannot be negative.

Inverse gamma function,

$$f(x, \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-(\beta/x)}$$

Since priors $w_0, w_1, \dots w_n$ being normally distributed

$$L(w_0, w_1, \dots w_n) = \frac{1}{\sqrt{2\pi}\tau_i} \exp\left(\frac{-1}{2\tau_i^2}(\beta_i - \mu_i)^2\right)$$

For MAP estimation expression,

$$\ln(\text{prior}) = \sum_{i=1}^{N}\left[\left(\frac{-1}{2}\ln(2\pi) - \ln(\tau_i) - \frac{1}{2\tau_i^2}(w_i - \mu_i)^2\right]\right.$$

$$+ \alpha \ln(\beta) - \ln(\Gamma(\alpha)) - (\alpha+1)\ln(\sigma^2) - \frac{\beta}{\sigma^2}$$

$$MAP = \arg\max\left(\ln(\text{likelihood}) + \ln(\text{prior})\right)$$

$$MAP = \arg\min\left(\sum_{i=1}^{N} \log(r_i) + \frac{r_i}{2\sigma^2}\left(y_i - w^T x_i\right)^2 \right)$$

$$+ \arg\min\left(\sum_{i=1}^{N} \ln(\tau_i) + \frac{1}{2\tau_i^2}\left(w_i - \mu_i\right)^2 + \alpha \ln(\beta) \right.$$

$$\left. - \ln(\Gamma(\alpha)) - (\alpha+1)\ln(\sigma^2) - \frac{\beta}{\sigma^2} \right)$$