Q2.1

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m} w^{(i)} \left(\theta^T x^{(i)} - y^{(i)}\right)^2$$

$$= \frac{1}{2}\sum_{i=1}^{m} \left(\sqrt{w^{(i)}}\cdot\theta^T x^{(i)} - \sqrt{w^{(i)}}\, y^{(i)}\right)^2$$

Because $\sum_i A_i^2 = A^T A$

$$= \frac{1}{2}\left(W^{\frac{1}{2}} X\theta - W^{\frac{1}{2}} y\right)^T \left(W^{\frac{1}{2}} X\theta - W^{\frac{1}{2}} y\right)$$

$$= \frac{1}{2}\left(\theta^T X^T W^{\frac{1}{2}T} - y^T W^{\frac{1}{2}T}\right)\left(W^{\frac{1}{2}} X\theta - W^{\frac{1}{2}} y\right)$$

Because $W^{\frac{1}{2}T} W^{\frac{1}{2}} = W$,

$$= \frac{1}{2}\left(\theta^T X^T W X\theta - \theta^T X^T W y - y^T W X\theta + y^T W y\right)$$

$$= \frac{1}{2}\left[\theta^T X^T W (X\theta - y) - y^T W (X\theta - y)\right]$$

$$= \frac{1}{2}\left(\theta^T X^T - y^T\right) W (X\theta - y)$$

$$= \frac{1}{2}(X\theta - y)^T W (X\theta - y)$$

W is a m×m diagnal matrix $\begin{bmatrix} w_1 & & 0 \\ & w_2 & \\ 0 & & \ddots \; w_m \end{bmatrix}$ with m diagnal terms that represent weighting terms for (non-negative) least square error of each traing data.