Q2) Weighted Least Squares Regression

Consider linear regression and let $c_1, \dots, c_n > 0$ be known weights. Determine the solution of $\min_{w,b} \sum_{i=1}^{n} c_i (y_i - w^T x_i - b)^2$.
Express your solution in terms of the matrix $C = \mathrm{diag}(c_1, \dots, c_n)$ and an appropriate data matrix X.

We first need to reduce the problem to ordinary least squares. Therefore, we must first eliminate $b \rightarrow$

$\sum_{i=1}^{n} c_i (y_i - w^T x_i - b)^2$ We first take the partial derivative with respect to b and

$$\frac{\partial}{\partial b} \sum_{i=1}^{n} c_i (y_i - w^T x_i - b)^2 = 0 \Rightarrow 2 \sum_{i=1}^{n} c_i (y_i - w^T x_i - b) = 0 \Rightarrow$$

$$\Rightarrow \sum_{i=1}^{n} c_i (y_i - w^T x_i - b) = 0 \Rightarrow \sum_{i=1}^{n} c_i y_i - w^T \sum_{i=1}^{n} c_i x_i - \sum_{i=1}^{n} c_i b = 0 \Rightarrow$$

$$\Rightarrow \sum_{i=1}^{n} c_i y_i - w^T \sum_{i=1}^{n} c_i x_i = b \sum_{i=1}^{n} c_i \Rightarrow \frac{\sum_{i=1}^{n} c_i y_i - w^T \sum_{i=1}^{n} c_i x_i = b \sum_{i=1}^{n} c_i}{\sum_{i=1}^{n} c_i} \Rightarrow$$

$$\Rightarrow \bar{y}_c = \frac{\sum_{i=1}^{n} c_i y_i}{\sum_{i=1}^{n} c_i} \; \& \; \bar{x}_c = \frac{\sum_{i=1}^{n} c_i x_i}{\sum_{i=1}^{n} c_i}$$ So we have $\Rightarrow$

$\Rightarrow b = \bar{y}_c - w^T \bar{x}_c$ We can now plug this back into our original equation:

$$\sum_{i=1}^{n} c_i (y_i - w^T x_i - \bar{y}_c + w^T \bar{x}_c)^2 \Rightarrow \sum_{i=1}^{n} c_i (y_i - \bar{y}_c + w^T(\bar{x}_c - x_i)) \Rightarrow$$

$\Rightarrow \sum_{i=1}^{n} c_i (y_i - \bar{y}_c - w^T(x_i - \bar{x}_c))$ We then define $\tilde{y}_i = y_i - \bar{y}_c$ and $\tilde{x}_i = x_i - \bar{x}_c$. We therefore have $\sum_{i=1}^{n} c_i (\tilde{y}_i - w^T \tilde{x}_i)^2$

This is now reduced to the ordinary least squares.

We then represent as matrix form as $\sum_{i=1}^{2} c_i(\tilde{y}_i - w^T\tilde{x}_i)^2$ to get $C(\tilde{y} - w^T\tilde{x})^2$, we then take the L2 norm to get $\Rightarrow C\|\tilde{y} - \tilde{x}w\|^2$ we then expand the square to get $\Rightarrow (\tilde{Y} - \tilde{X}w)^T C(\tilde{Y} - \tilde{X}w) \Rightarrow$

$\Rightarrow \tilde{Y}^T C\tilde{Y} - \tilde{Y}^T C\tilde{X}w - w^T\tilde{X}^T C\tilde{y} + w^T\tilde{X}^T C\tilde{X}w \Rightarrow$

We then take the $\nabla L(w)$ and set it equal to zero

$$\nabla L(w) = \frac{\partial}{\partial w}\tilde{Y}^T C\tilde{Y} - \frac{\partial}{\partial w} 2\tilde{Y}^T C\tilde{X}w + \frac{\partial}{\partial w} w^T\tilde{X}^T C\tilde{X}w \Rightarrow$$

$$\Rightarrow 2(X^TCX - X^TC\tilde{y}) = 0 \Rightarrow X^TCX = X^TC\tilde{y} \Rightarrow$$

$\Rightarrow$ therefore $w = (X^TCX)^{-1}X^TC\tilde{y}$