ASHUTOSH SINGH — Exercise 1 — 8DS385.

: LINEAR REGRESSION :

(A) $y = X\beta + e$ [y = Response vector, X = predictor matrix, β = weights vector, e = error].

$e = y - X\beta$, $\Rightarrow \sum e_i = \sum (y_i - x_i^T\beta)$

if we trust some observations more than others, we introduce this preference using weights

$\Rightarrow \sum e_i^2 = \sum (y_i - x_i^T\beta)^2$ [sum squared error].

Introducing weights :

$$\sum e_i^2 = \sum_{i=1}^{N} w_i (y_i - x_i^T\beta)^2 = \text{weighted sum squared errors.}$$

we find a β such that the above error is minimized for given data points.

$$\Rightarrow \hat{\beta} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{N} \frac{w_i}{2} (y_i - x_i^T\beta)^2$$

$$= \arg\min_{\beta \in \mathbb{R}^p} \left[\frac{1}{2} \sum_{i=1}^{N} y_i w_i y_i - \sum_{i=1}^{N} x_i^T \beta w_i y_i + \frac{1}{2} \sum_{i=1}^{N} x_i^T \beta w_i x_i^T \beta \right]$$

$$= \underset{\beta \in \mathbb{R}^p}{\arg\min} \left[\frac{1}{2} y^T W y - X^T \beta^T W y + \frac{1}{2} X^T \beta^T W X \beta \right]$$

$$\beta = \underset{\beta \in \mathbb{R}^p}{\arg\min} \left[\frac{1}{2} (y - X\beta)^T W (y - X\beta) \right]$$

using calculus, to minimize β, we set

$$\nabla(\hat{\beta}) = 0$$

$$\nabla(\beta) = 0 - X^T W y + \frac{2}{2} X^T W X \hat{\beta} = 0$$

$$\Rightarrow \boxed{(X^T W X) \hat{\beta} = X^T W y}$$