Зважена лінійна регресія — це регресія, у якій ми по-різному оцінюємо помилку для кожного з навчальних прикладів. Для навчання зваженої лінійної регресії нам потрібно мінімізувати функцію втрат виду:

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{m} \omega^{(i)}(\theta^\top x^{(i)} - y^{(i)})^2$$

а) Покажіть, що функція втрат зваженої лінійної регресії $J(\theta)$ також може бути записана у такому вигляді: $J(\theta) = (X\theta - \vec{y})^\top W(X\theta - \vec{y})$ де $X, \vec{y}$ визначені так само, як на лекції, а W — діагональна матриця. Поясніть, що таке матриця W та якими будуть її елементи.

Відповідь:

Використаємо той факт, що для будь-якого вектора $z : z^\top z = \sum_i {z_i}^2$, тоді

$J(\theta) = \sum_{i=1}^{m} \omega^{(i)}(\theta^\top x^{(i)} - y^{(i)})^2 = (X\theta - \vec{y})^\top W(X\theta - \vec{y})$, де $W = \begin{bmatrix} \omega^{(1)} & 0 & \ddots & 0 \\ 0 & \omega^{(2)} & \ddots & 0 \\ \ddots & \ddots & \ddots & \ddots \\ 0 & \ddots & \ddots & \omega^{(m)} \end{bmatrix}$,

діагональна матриця розмірністю $m \times m$

/---/

b) Якщо всі $\omega^{(i)} = 1$, тоді, як ми бачили на лекції, нормальне рівняння має вигляд: $X^\top X\theta = X^\top \vec{y}$, а значення θ, що мінімізує функцію втрат і дає найвищу точність передбачення: $\theta = (X^\top X)^{-1} X^\top \vec{y}$

Виведіть вираз для знаходження θ у зваженій лінійній регресії. Знайдіть градієнт $\nabla_\theta J(\theta)$ і, прирівнявши його до нуля, виведіть нормальне рівняння для знаходження θ. Вираз буде залежати від X,W і $\vec{y}$

Відповідь:

$\frac{\partial}{\partial\theta} J(\theta) = \frac{1}{2} \frac{\partial}{\partial\theta}(X\theta - Y)^\top W(X\theta - Y) = \frac{1}{2} \frac{\partial}{\partial\theta}(\theta^\top X^\top WX\theta - \theta^\top X^\top WY - Y^\top WX\theta + Y^\top WY) = (X^\top WX\theta$

$\frac{\partial}{\partial\theta} J(\theta) = 0$, тоді $(X^\top WX\theta - X^\top WY) = 0$, отже $\theta = (X^\top WX)^{-1}(X^\top WY)$