Considering n features & m training examples,

$$\hat{y} = w_1 x_1 + w_2 x_2 + w_3 x_3 + \cdots + w_n x_n + b$$

Loss = $L(\hat{y}, y) = (\hat{y} - y)^2$

Cost = MSE = $J(w, b) = \frac{1}{m} \sum_{i=1}^{m} \left[\hat{y}^{(i)} - y^{(i)}\right]^2$

For convenience,

$$J(w, b) = \frac{1}{2m} \sum_{i=1}^{m} \left[\hat{y}^{(i)} - y^{(i)}\right]^2$$

$$\cancel{= \frac{1}{2m} \sum_{i=1}^{m}}$$

minimising cost wrt w_j $(j \in [1, m])$:

$$\frac{\partial J}{\partial w_j} = \frac{1}{m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right) \frac{\partial \left(\hat{y}^{(i)} - y^{(i)}\right)}{\partial w_j}$$

$$= \frac{1}{m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right) x_j^{(i)}$$

wrt b :

$$\frac{\partial J}{\partial b} = \frac{1}{m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right) \frac{\partial \left(\hat{y}^{(i)} - y^{(i)}\right)}{\partial b}$$

$$= \frac{1}{m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right)$$

Date:
Mon. Tue. Wed. Thu. Fri. Sat. Sun

Notes

Updating :

$$w_j' = w_j - \alpha \frac{\partial J}{\partial w_j}$$

$$b' = b - \alpha \frac{\partial J}{\partial b}$$

If we take vector X for m examples & n features we get $X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & & & \\ x_{m1} & \cdots & \cdots & x_{mn} \end{bmatrix}_{m \times n}$

weight matrix $W = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}_{n \times 1}$

$\therefore \hat{y} = XW + b$

where $\hat{y}$ is a $m \times 1$ vector

Date:
Mon. Tue. Wed. Thu. Fri. Sat. Sun

Notes

so,

$$J(w, b) = \frac{1}{2m} (\hat{y} - y)^2$$

$$= \frac{1}{2m} (\hat{y} - y)^T (\hat{y} - y)$$

$$= \frac{1}{2m} (Xw + b - y)^T (Xw + b - y)$$

Using matrix calculus:

$$\frac{\partial J}{\partial w} = \frac{1}{m} (Xw + b - y)^T X$$

$$\frac{\partial J}{\partial b} = \frac{1}{m} (Xw + b - y)^T \cdot 1$$

$$= \frac{1}{m} \Sigma (\hat{y} - y)$$

Updating:

$$w' = w - \alpha \left(\frac{\partial J}{\partial w} \right)^T$$

$$= w - \alpha \frac{1}{m} X^T (\hat{y} - y)$$

$$b' = b - \alpha \left(\frac{\partial J}{\partial b} \right)$$