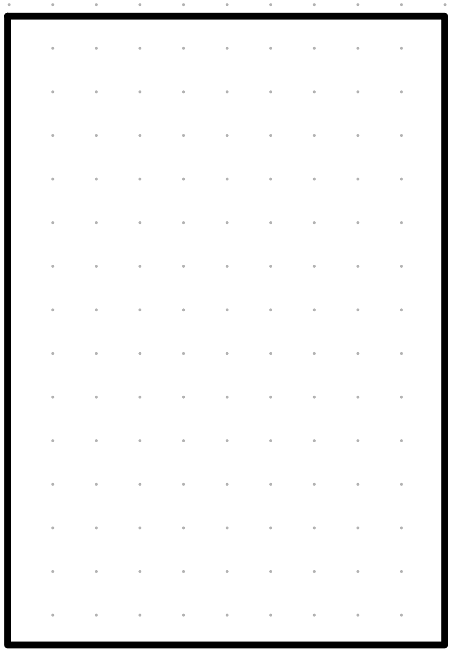
$$\frac{\partial J(\theta)}{\partial \theta_j} = \sum_{i=1}^{N} (\hat{y}_i - y_i) x_i^j$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = \sum_{i=1}^{N} (\hat{y}_i - y_i) x_i^j$$

MATRIX X

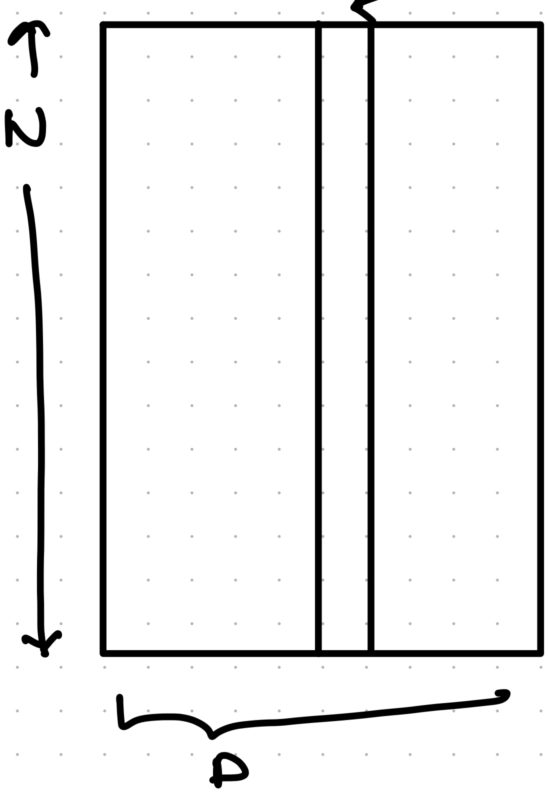


$$\frac{\partial J(\theta)}{\partial \theta_j} = \sum_{i=1}^{N} (\hat{y}_i - y_i) x_i^j$$

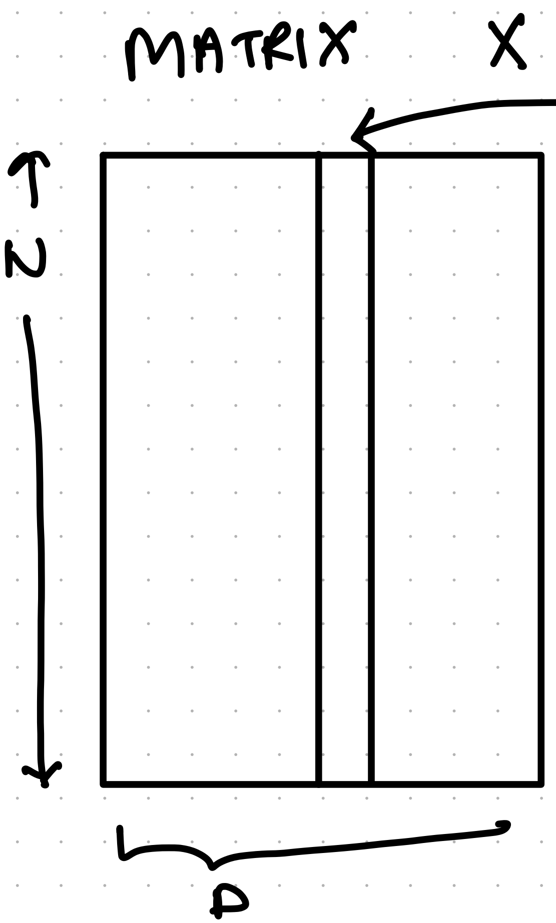
MATRIX X

j^{th} column of X

$= X^j$

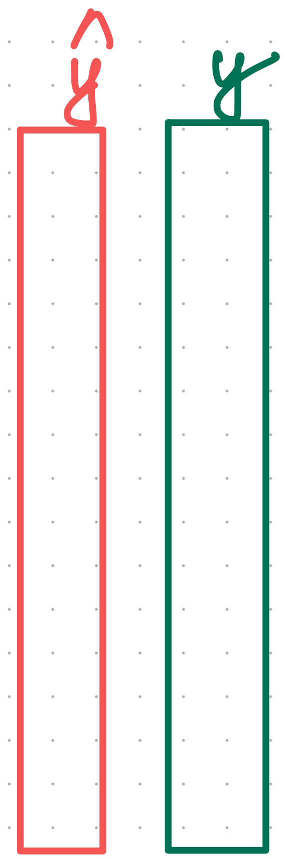


$$\frac{\partial J(\theta)}{\partial \theta_j} = \sum_{i=1}^{N} (\hat{y}_i - y_i)\, x_i^j$$

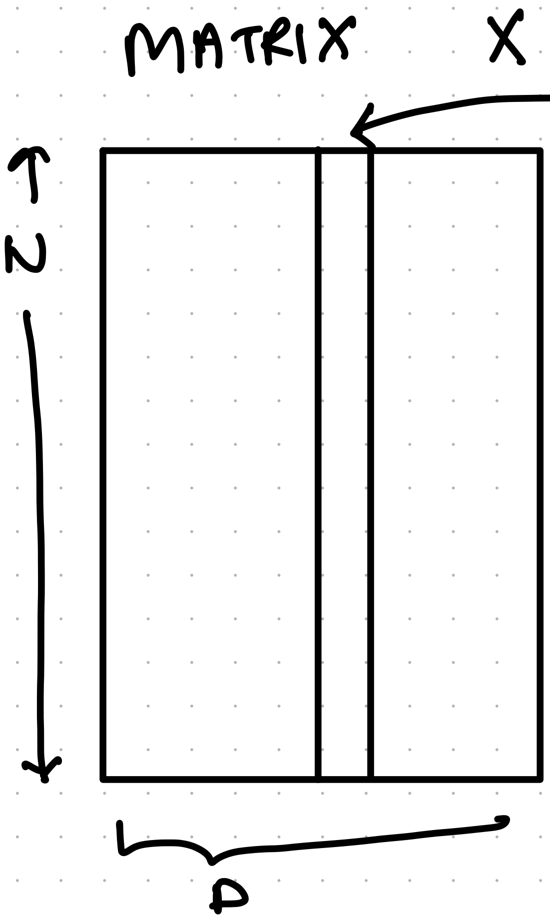


j^{th} column of X

$= X^j$

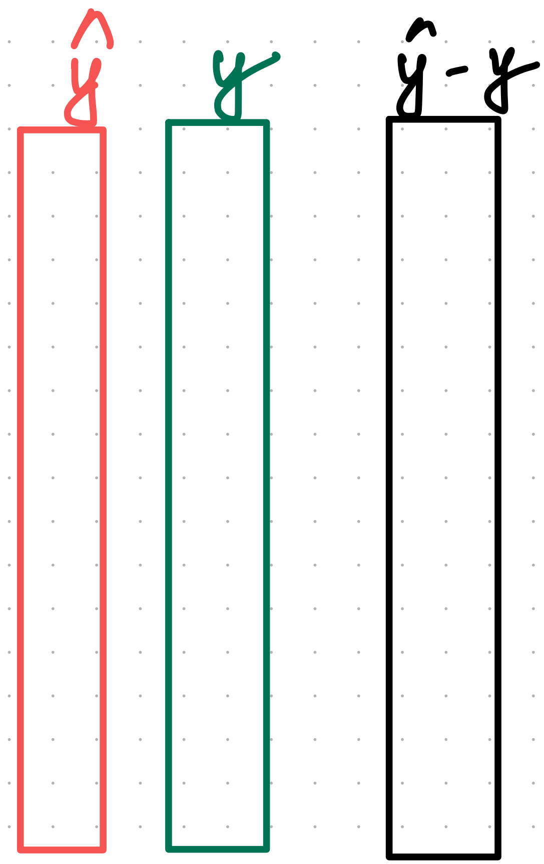


$$\frac{\partial J(\theta)}{\partial \theta_j} = \sum_{i=1}^{N} (\hat{y}_i - y_i)\, x_i^j$$



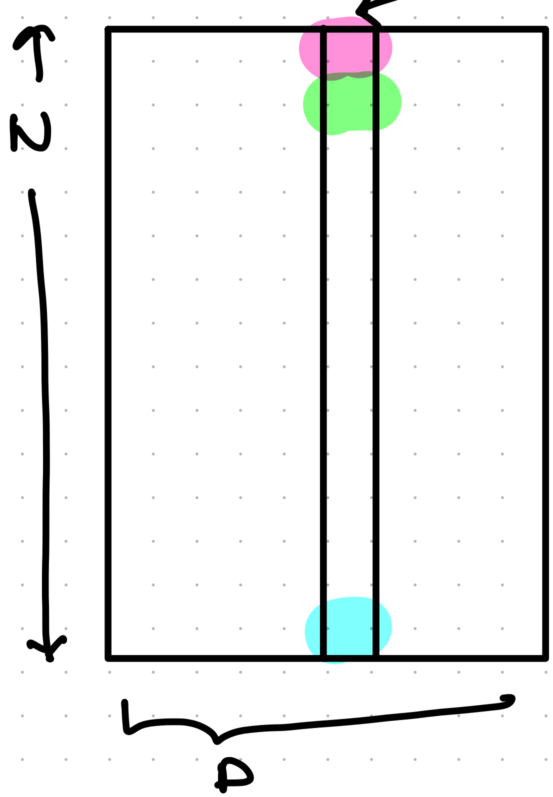
j^{th} column of X

$= X^j$



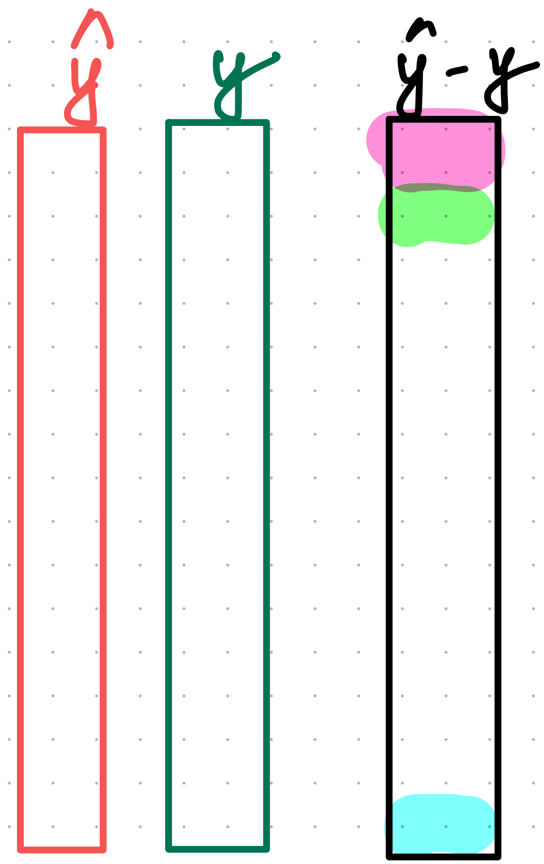
$$\frac{\partial J(\theta)}{\partial \theta_j} = \sum_{i=1}^{N} (\hat{y}_i - y_i) x_i^j$$

MATRIX X

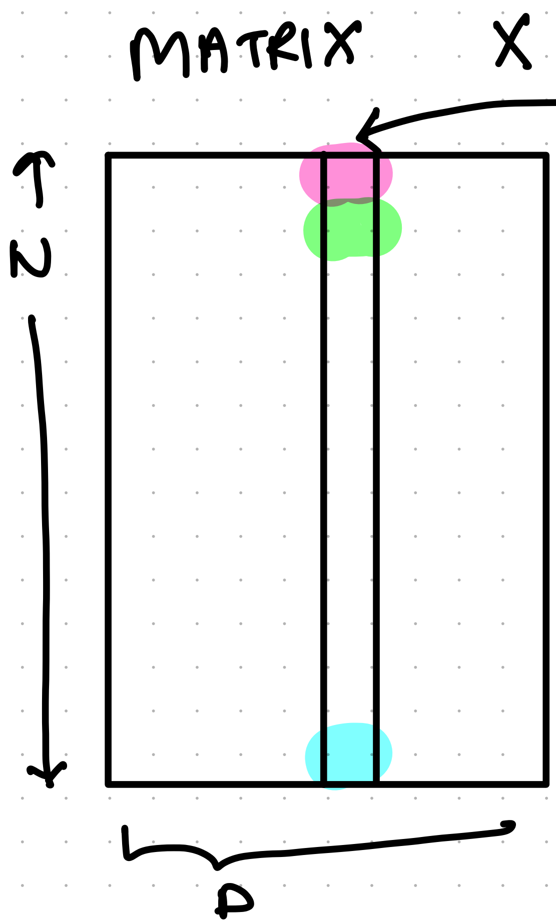


j^{th} column of X

$= X^j$

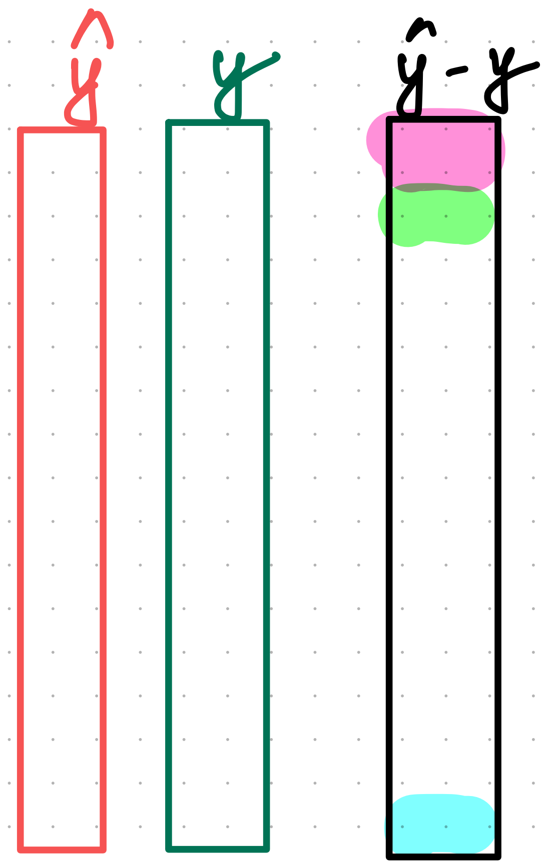


$$\frac{\partial J(\theta)}{\partial \theta_j} = \sum_{i=1}^{N} (\hat{y}_i - y_i)\, x_i^j = \underset{1\times N}{X^{j^T}} (\hat{y} - y)$$



j^{th} column of X

$= X^j$



$$\frac{\partial J(\theta)}{\partial \theta_j} = \sum_{i=1}^{N} (\hat{y}_i - y_i)\, x_i^j = \underset{1\times N}{x^{j^T}} (\hat{y} - y)$$

$$\begin{bmatrix} \frac{\partial J(\theta)}{\partial \theta_1} \\ \frac{\partial J(\theta)}{\partial \theta_2} \\ \vdots \\ \frac{\partial J(\theta)}{\partial \theta_D} \end{bmatrix} = \begin{pmatrix} x^{1^T} (\hat{y} - y) \\ x^{2^T} (\hat{y} - y) \\ \vdots \\ x^{D^T} (\hat{y} - y) \end{pmatrix}$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = \sum_{i=1}^{N} (\hat{y}_i - y_i)\, x_i^j = \underset{1\times N}{x^{j^T}} (\hat{y} - y)$$

$$\begin{bmatrix} \frac{\partial J(\theta)}{\partial \theta_1} \\ \frac{\partial J(\theta)}{\partial \theta_2} \\ \vdots \\ \frac{\partial J(\theta)}{\partial \theta_D} \end{bmatrix} = \begin{pmatrix} x^{1^T} (\hat{y} - y) \\ x^{2^T} (\hat{y} - y) \\ \vdots \\ x^{D^T} (\hat{y} - y) \end{pmatrix} = X^T (\hat{y} - y)$$