Example for a simple linear regression

1. Draw the network scheme
2. Draw computation graph. Provide the MSE part
3. Write dimensions of each term
4. Relate the nodes from the network with the graph nodes

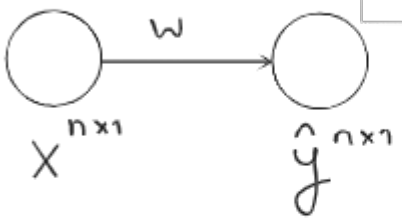
5. Write down the paths for chained partial derivatives
6. Compute the derivatives based on the graph
7. Write the dimensions of each term
8. Reorder and transpose accordingly
9. Rewrite with the terms from the network scheme as the final formula

Architecture

$\hat{y} = wX$

$Loss = MSE(\hat{y}, y)$

Network scheme



Chain of derivatives

$$\frac{\partial MSE}{\partial w} = \frac{\partial MSE}{\partial d} \times \frac{\partial d}{\partial c} \times \frac{\partial c}{\partial b} \times \frac{\partial b}{\partial a} \times \frac{\partial a}{\partial w}$$

$$\frac{1}{n} \quad \sum^{n} \quad 2b^{n\times 1} \quad 1 \quad X^{n\times 1}$$

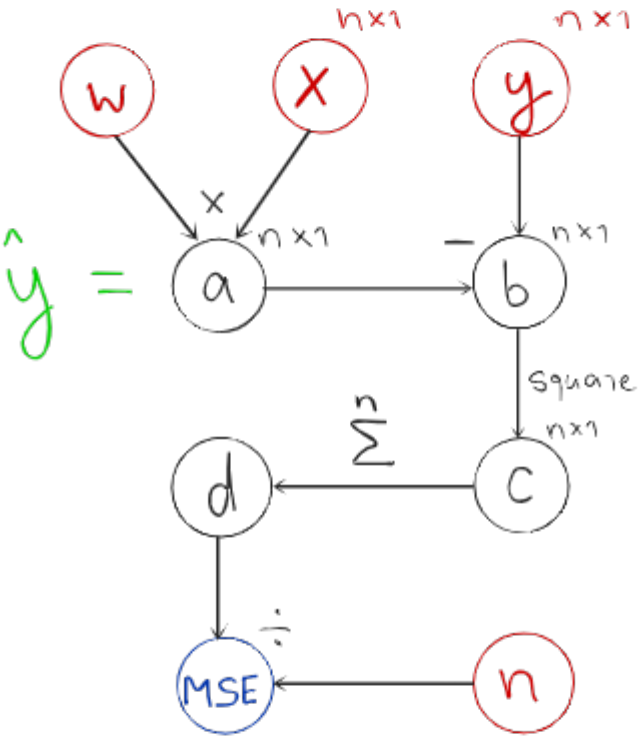
$$\frac{1}{n/2} \quad \sum^{n} \quad b^{T^{1\times n}} \quad X^{n\times 1}$$

$$= \frac{1}{n/2} \sum^{n} \left[b^T X \right]^{1\times 1}$$

$$= \frac{2}{n} b^T X \quad = \frac{2}{n} (\hat{y} - y)^T X$$

$$\frac{\partial MSE}{\partial w} = \frac{2}{n} (\hat{y} - y)^T X$$

Computation graph



See next page for numerical computation

Compute with numbers

Given:

$w_0 = 2$

$X = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$

$y = \begin{bmatrix} 6 \\ 12 \end{bmatrix}$

$\alpha = 0.05$

1) Forward pass $\hat{y} = wX$

$\hat{y} = 2 \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$

2) Compute error

$MSE = \frac{1}{2} \frac{1}{2} \sum \left(\begin{bmatrix} 4 \\ 8 \end{bmatrix} - \begin{bmatrix} 6 \\ 12 \end{bmatrix} \right)^2 = \frac{1}{4} (4 + 16) = 5$

3) Compute gradient $\frac{\partial MSE}{\partial w} = \frac{2}{n} (\hat{y} - y)^T X$

$\frac{\partial MSE}{\partial w} = \frac{2}{2} \left(\begin{bmatrix} 4 \\ 8 \end{bmatrix} - \begin{bmatrix} 6 \\ 12 \end{bmatrix} \right)^T \begin{pmatrix} 2 \\ 4 \end{pmatrix} = [-2 \ -4] \begin{bmatrix} 2 \\ 4 \end{bmatrix} = -20$

4) Propagate the error $w_1 = w_0 - \alpha \frac{\partial MSE}{\partial w}$

$w_1 = 2 - 0.05 \cdot (-20) = 2 + 1 = 3$

5) Forward pass $\hat{y} = wX$

$\hat{y} = 3 \cdot \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 6 \\ 12 \end{bmatrix}$

6) Compute error $MSE = \frac{1}{2n} \sum^{n} (\hat{y} - y)^2$

$MSE = \frac{1}{2} \frac{1}{2} \sum \left(\begin{bmatrix} 6 \\ 12 \end{bmatrix} - \begin{bmatrix} 6 \\ 12 \end{bmatrix} \right)^2 = 0$

7) Celebrate!