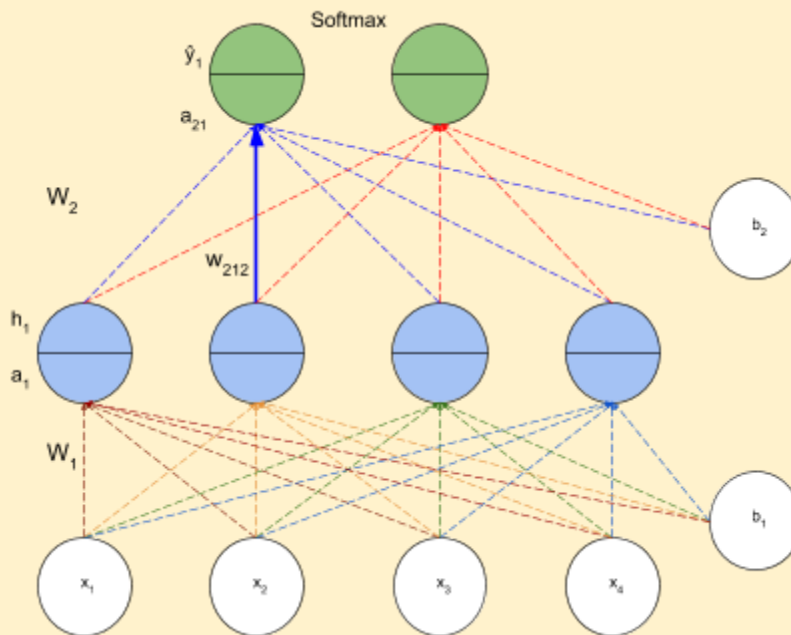


Partial Derivatives with respect to a

Part 1

How do we compute partial derivatives

1. The following neural network will be used to demonstrate the calculations



2. Here are the parameters of the network

a. $b = [0.5 \ 0.3]$

b.

$$W_1 = \begin{bmatrix} 0.1 & 0.3 & 0.8 & -0.4 \\ -0.3 & -0.2 & 0.5 & 0.5 \\ -0.3 & 0 & 0.5 & 0.4 \\ 0.2 & 0.5 & -0.9 & 0.7 \end{bmatrix}$$

c.

$$W_2 = \begin{bmatrix} 0.5 & 0.8 & 0.2 & 0.4 \\ 0.5 & 0.2 & 0.3 & -0.5 \end{bmatrix}$$

d. $x = [2 \ 5 \ 3 \ 3]$ true distribution $y = [1 \ 0]$

PadhAI: Backpropagation - the light math version

One Fourth Labs

3. Now, we want to find the partial derivative w.r.t w_{212} as highlighted in the figure $\frac{\partial L}{\partial w_{212}}$
4.
$$\frac{\partial L}{\partial w_{212}} = \left(\frac{\partial L}{\partial a_{21}}\right) \cdot \left(\frac{\partial a_{21}}{\partial w_{212}}\right) = \left(\frac{\partial L}{\partial \hat{y}_1}\right) \cdot \left(\frac{\partial \hat{y}_1}{\partial a_{21}}\right) \cdot \left(\frac{\partial a_{21}}{\partial w_{212}}\right)$$
5. We will solve the above equation sequentially
 - a. Consider square error loss function L
 - b.
$$\frac{\partial L}{\partial \hat{y}_1} = \sum_{i=1}^2 (y_i - \hat{y}_i)^2$$
 - i.
$$\frac{\partial L}{\partial \hat{y}_1} = \frac{\partial}{\partial y_1} [(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2]$$
 - ii. Here, the y_2 terms get cancelled, leaving $\frac{\partial}{\partial y_1} [(y_1 - \hat{y}_1)^2] = -2(y_1 - \hat{y}_1)$