Sigmoid function:

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

Net:

inputs = [2, 3, 1]

$$w_1 = \begin{bmatrix} 0.5 & 0.3 & -0.1 \\ 0.2 & 0.4 & 0.6 \\ -0.3 & 0.5 & 0.2 \end{bmatrix}$$

$b_1 = [0.1 \quad 0.2 \quad 0.3]$

$w_2 = [0.8 \quad -0.5 \quad 0.4]$

$b_2 = 0.3$

actual = 1

First layer: (hidden layer)

$$w_1 = \begin{bmatrix} 0.5 & 0.3 & -0.1 \\ 0.2 & 0.4 & 0.6 \\ -0.3 & 0.5 & 0.2 \end{bmatrix} \qquad x = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$$

classic matrix multiplication

$$z_1 = \underbrace{\begin{bmatrix} 0.5 & 0.3 & -0.1 \\ 0.2 & 0.4 & 0.6 \\ -0.3 & 0.5 & 0.2 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \end{bmatrix}}_{z_1 = \text{np.dot}(w_1, x) + b_1} = \begin{bmatrix} 1.9 \\ 2.4 \\ 1.4 \end{bmatrix}$$

$$h = \sigma(z_1) = \underbrace{\begin{bmatrix} \frac{1}{1+e^{-1.9}} \\ \frac{1}{1+e^{-2.4}} \\ \frac{1}{1+e^{-1.4}} \end{bmatrix}}_{\text{hidden layer activation}} = \begin{bmatrix} 0.86989 \\ 0.916827 \\ 0.802183 \end{bmatrix} \begin{matrix} \leftarrow \text{hidden nueron 1} \\ \leftarrow \text{hidden nueron 2} \\ \leftarrow \text{hidden nueron 3} \end{matrix}$$

Second Layer:

$$z_2 = w_2 \cdot h = \overbrace{\begin{bmatrix} 0.8 \\ -0.5 \\ 0.4 \end{bmatrix} \cdot \begin{bmatrix} 0.869 \\ 0.916 \\ 0.802 \end{bmatrix}}^{\text{sum(multiplication)}} + [0.3] = 0.559 + 0.3 = \underline{0.859}$$

← predicted value

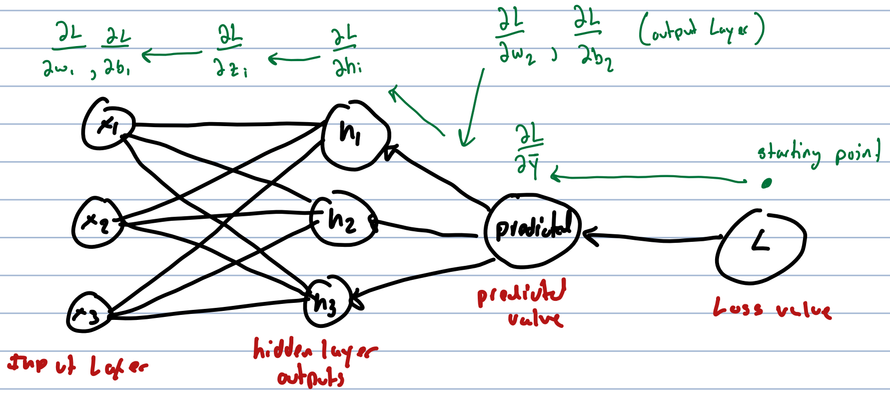
predicted = 0.859

loss = $(1 - 0.859)^2 = 0.02006$

Back propogation

We must move backward:

Let predicted = $\bar{y}$, actual = y



$\bar{y} = w_{21}h_1 + w_{22}h_2 + w_{23}h_3 + b_2$

$\frac{\partial \bar{y}}{\partial h} = w$

First output layer:

① $\frac{\partial L}{\partial \bar{y}} = -2(y-\bar{y})$

② $\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \bar{y}} \cdot h$

③ $\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial \bar{y}}$

Next, hidden layer:

② $\frac{\partial L}{\partial h} = \frac{\partial L}{\partial \bar{y}} \cdot \frac{\partial \bar{y}}{\partial h_i}$

$= \frac{\partial L}{\partial \bar{y}} \cdot w_2$

③ $\frac{\partial L}{\partial z_i} = \frac{\partial L}{\partial h} \cdot \frac{\partial h}{\partial z_i}$

$= \frac{\partial L}{\partial h} \cdot \sigma'(z)$

$\frac{\partial h}{\partial z} = \sigma'(z)$

④ $\frac{\partial L}{\partial w_i} = \frac{\partial L}{\partial z_i} \cdot \frac{\partial z_i}{\partial w_i}$

$= \frac{\partial L}{\partial z_i} \cdot x$

$z_i = w_i x + b$

$\frac{\partial z_i}{\partial w_i} = x$

$\frac{\partial z_i}{\partial b_i} = 1$

⑤ $\frac{\partial L}{\partial b_i} = \frac{\partial L}{\partial z_i} \cdot \frac{\partial z_i}{\partial b_i}$

$= \frac{\partial L}{\partial z_i}$

After calculation of gradients:

$$\left.\begin{aligned} w_1^{new} &= w_1 - \alpha \frac{\partial L}{\partial w_1} \\ b_1^{new} &= b_1 - \alpha \frac{\partial L}{\partial b_1} \end{aligned}\right\} \text{Hidden Layer}$$

$$w_2^{new} = w_2 - \alpha \frac{\partial L}{\partial w_2}$$

$$b_2^{new} = b_2 - \alpha \frac{\partial L}{\partial b_2}$$

α: learning rate