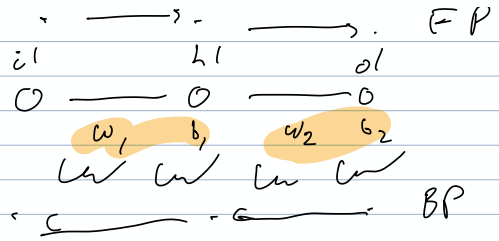


i1 = temp

o1 = how happy I am

$h_1 = NLU(w_1 \cdot i_1 + b_1)$

NLU = non-linear unit

$o_1 = h_1 w_2 + b_2$

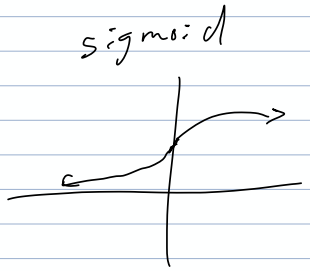
$f(x) = \frac{1}{1+e^{-x}}$ (sigmoid)

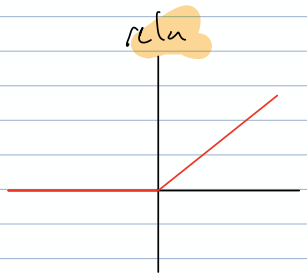
Ex

$h_1 = Relu(w_1 \cdot i_1 + b_1)$

$o_1 = h_1 w_2 + b_2$

$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{otherwise} \end{cases}$ (relu)





$i_1 = 3$

$w_1 = 4$

$b_1 = -13$

$w_2 = 6$

$b_2 = 7$

temp	happiness
60	-100
74	100
3	-200

$relu(4 \cdot 3 - 13) = 0 = h_1$

$0 \cdot w_2 + b_2 = 7$

1. Make network (choose NLU's, hidden layers)
   (b) randomly initialize w's, b's
2. Choose a loss function

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

- outputs a scalar

3. Backprop

$h_1 = ReLU(w_1 \cdot i_1 + b_1)$

$o_1 = h_1 w_2 + b_2$ $\quad L = \frac{1}{2}(o_1 - y)^2$

$\frac{\partial o_1}{\partial w_2} = h_1$ $\quad \frac{\partial o_1}{\partial b_2} = 1$ $\quad \frac{\partial L}{\partial o_1} = o_1 - y$

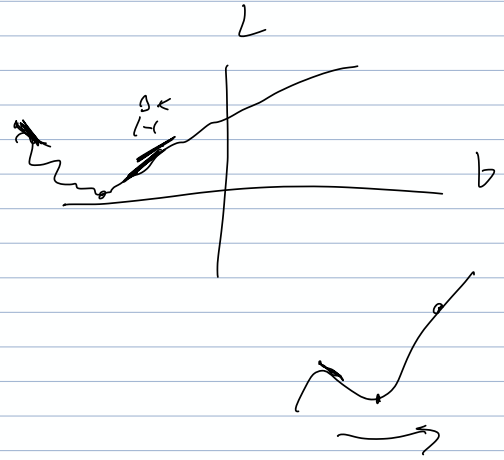
$= 7 - (-200)$

$= 207$

$L = MSE = 207^2$

$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial o_1} \cdot \frac{\partial o_1}{\partial w_2}$

$= 207 \cdot 0 = 0$



$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial o_1} \cdot \frac{\partial o_1}{\partial b_2}$

$= 207 \cdot 1 = 207$

- We should decrease b to decrease L (if $\frac{\partial L}{\partial b_2} > 0$)
- $\uparrow b_2$ to $\downarrow L$ (if $\frac{\partial L}{\partial b_2} < 0$)

$$b_2' = b_2 - \frac{\partial L}{\partial b_2} \cdot \varepsilon$$

ε = learning rate

$= 1e^{-3}$

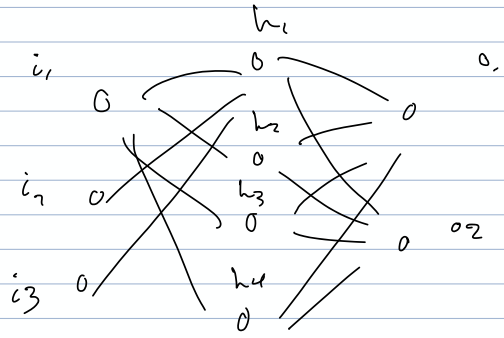
$$\frac{\partial o_1}{\partial h_1} = w_2$$

$$\frac{\partial L}{\partial h_1} = \frac{\partial L}{\partial o_1} \cdot \frac{\partial o_1}{\partial h_1}$$

$$= 287.6$$

$$\frac{\partial h_1}{\partial b_1} =$$

$$\frac{\partial L}{\partial h_1} \cdot \frac{\partial h_1}{\partial b_1} = \frac{\partial L}{\partial b_1}$$



$$h_1 = NCh(w_{i_1,h_1} \cdot i_1 + w_{i_2,h_1} \cdot i_2 + w_{i_3,h_1} \cdot i_3 + b_{h_1})$$

↑
MLP