

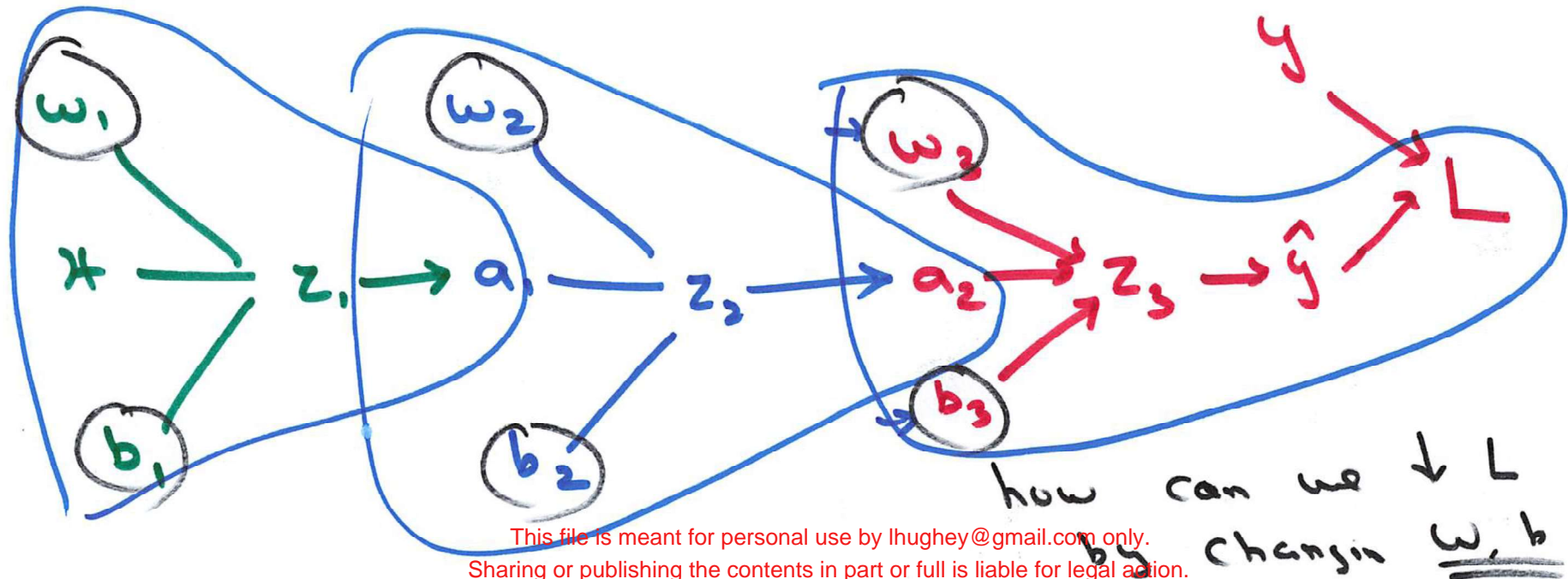
$$\begin{cases} z_1 = x w_1 + b_1 \\ a_1 = \sigma(z_1) \end{cases}$$

$$\begin{cases} z_2 = a_1 w_2 + b_2 \\ a_2 = \sigma(z_2) \end{cases}$$

$$\begin{cases} z_3 = a_2 w_3 + b_3 \\ \hat{y} = \sigma(z_3) \end{cases} \leftarrow$$

$$L(w) = \mathbb{E}(\hat{y} - y)^2$$

∇L ?



how can we $\downarrow L$
by changing w, b

Slope of L with respect to w_3 } = $\frac{\partial L}{\partial w_3}$

$$\frac{\partial L}{\partial w_3} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_3} \cdot \frac{\partial z_3}{\partial w_3}$$

\downarrow \downarrow \searrow
 $2(\hat{y} - y)$ $\sigma'(z_3)$ a_2

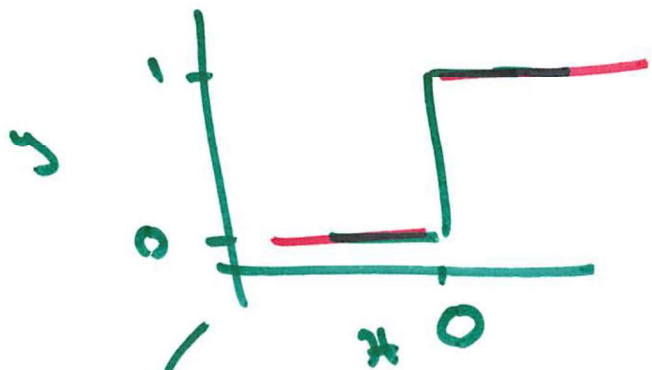
$$\frac{\partial a_2}{\partial w_2}$$

$$y = f\left(\sum \underline{w_i} x_i + b\right)$$

non linear func

$$y = f\left(\sum \underline{w_i} \underline{x_i} + b\right)$$

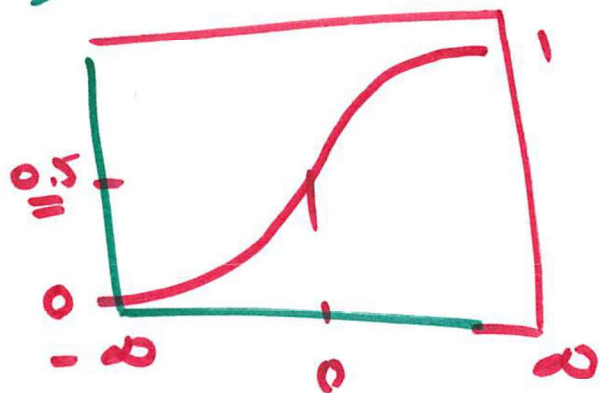




$$y = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

~~$y = 0$~~ $y = \underline{\underline{\sigma(x)}}$

$$\sigma' = \sigma - 2 \left(\frac{\text{slope of } \sigma}{q} \right)$$



Sigmoid $\rightarrow 0$ to 1 $\sigma'(x)$

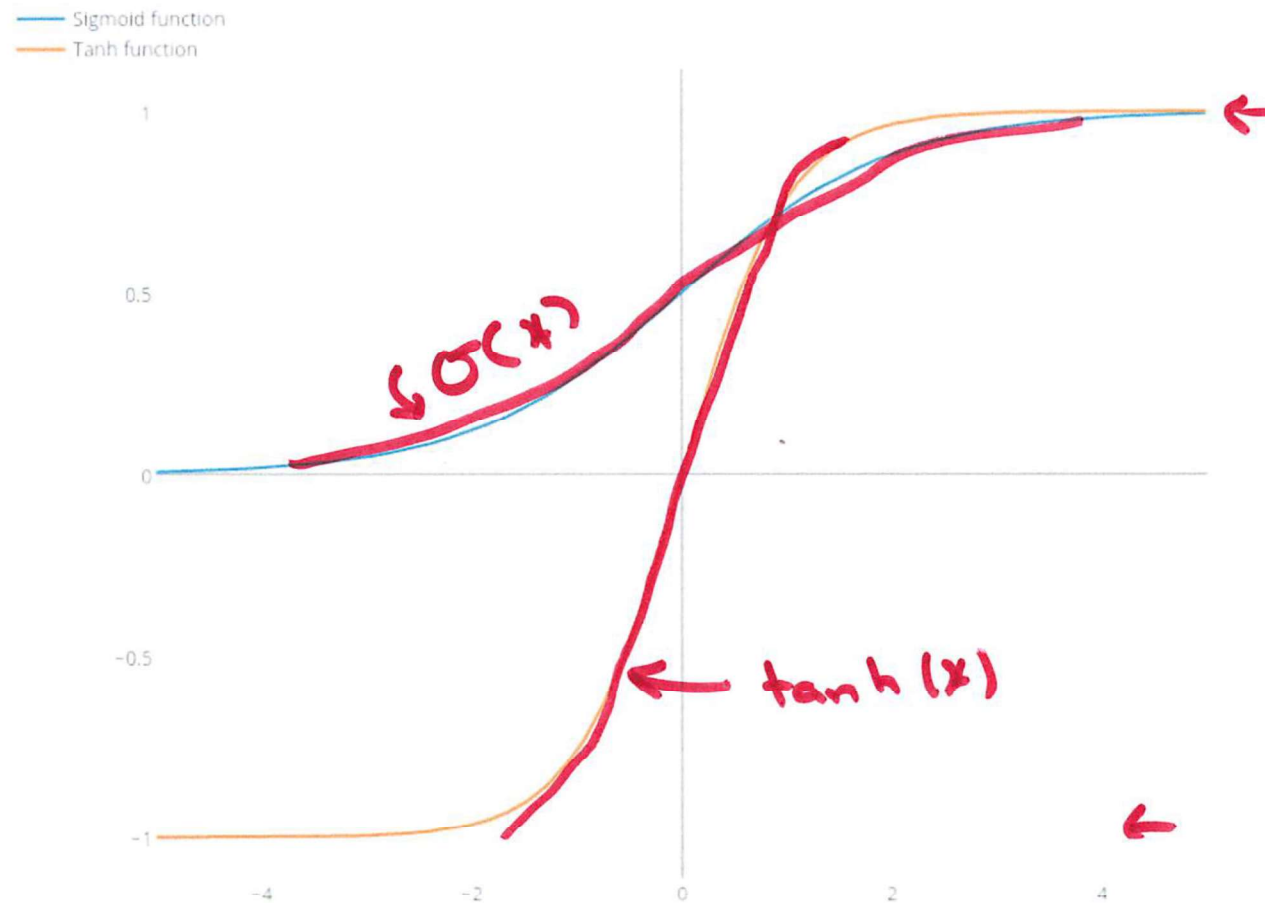
$$y = \sigma(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1}$$

$$\underline{\underline{2\sigma(x) - 1}} \rightarrow -1 \text{ to } 1$$

$$y = \tanh(x) = 2\sigma(2x) - 1 \rightarrow -1 \text{ to } 1$$

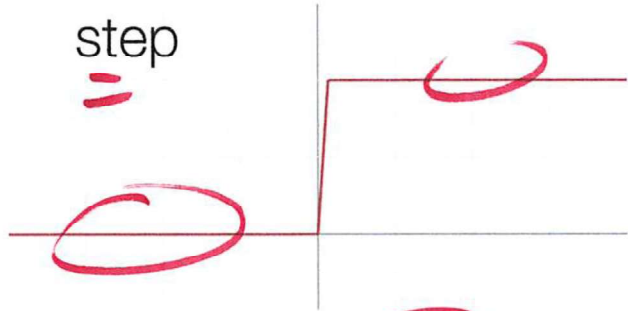
$$\hookrightarrow \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1}$$

$$\tanh(x) = 2\sigma(2x) - 1$$

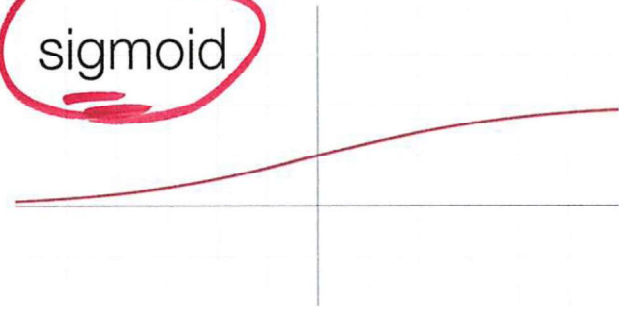


Activation Functions

step



sigmoid



ReLU

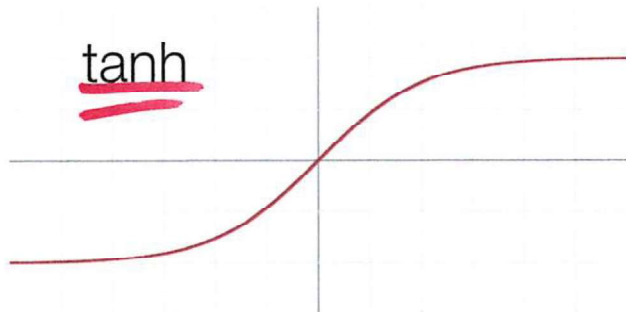


$$y = f(x)$$

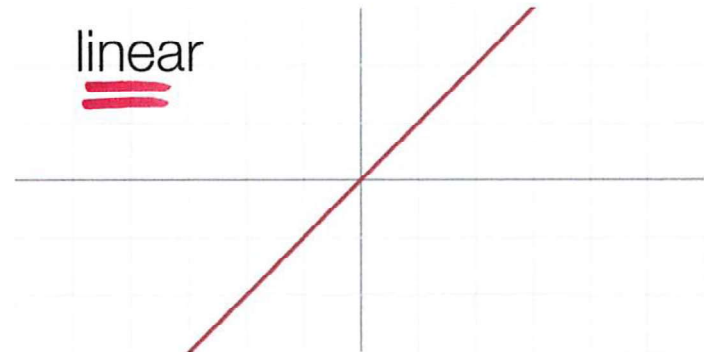
$$y = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases}$$

$$y = \max(0, x)$$

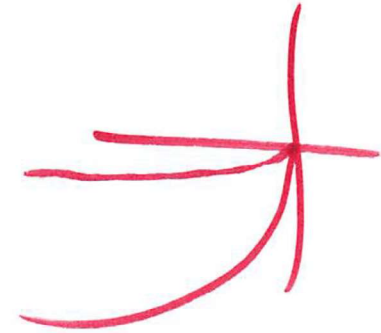
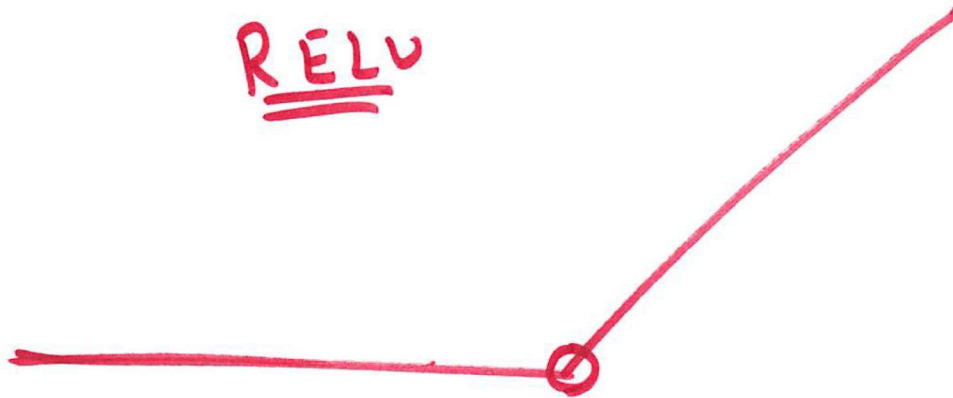
tanh



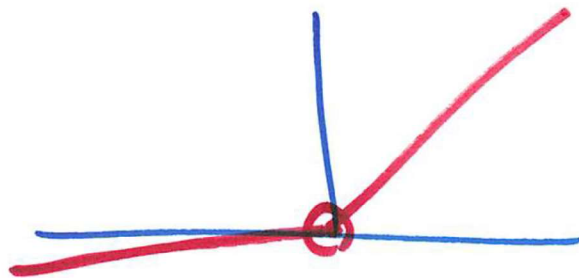
linear



ReLU

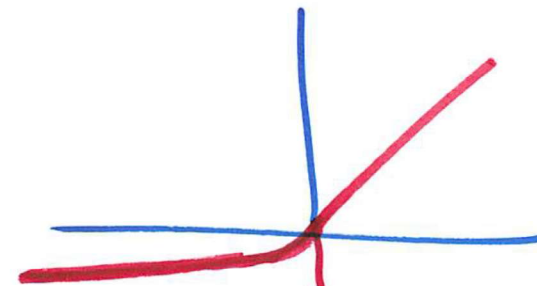


leaky ReLU



$$y = \begin{cases} x & \text{if } x \geq 0 \\ 0.001x & \text{if } x < 0 \end{cases}$$

EReLU



$$y = \begin{cases} x & \text{if } x \geq 0 \\ x(e^x - 1) & \text{if } x < 0 \end{cases}$$

Hidden → ReLU, Leaky ReLU, ERelu
Sigmoid, tanh, ~~linear~~

Output
layer

Regression → linear

Classification

binary

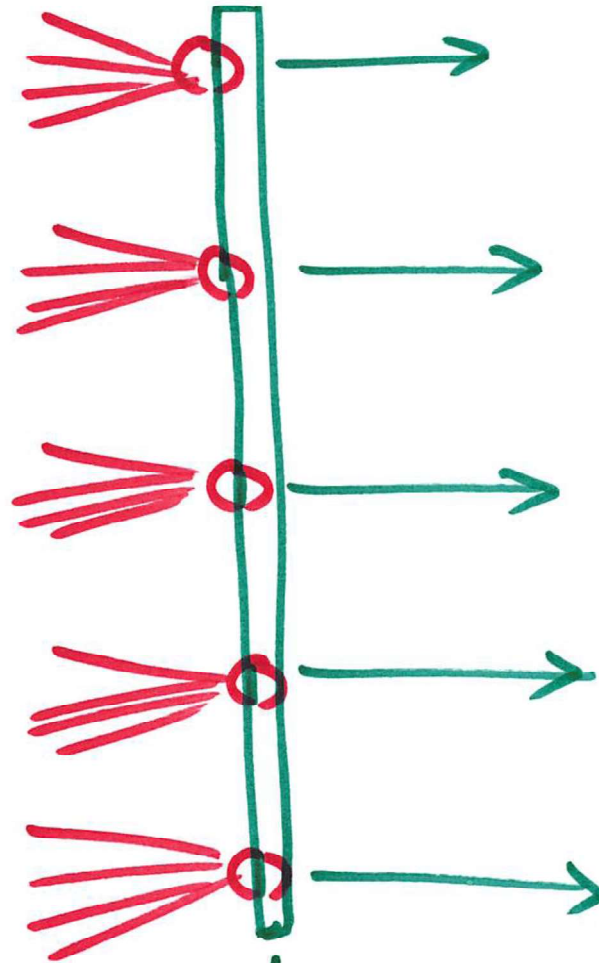
Sigmoid/tanh

multiclass

A	B	C	D	E
.

Softmax ①

$f(\dots) \rightarrow$



Softmax activation

$$z_i = (\sum w_i x + b)$$

~~$$a_i = f(z_i)$$~~