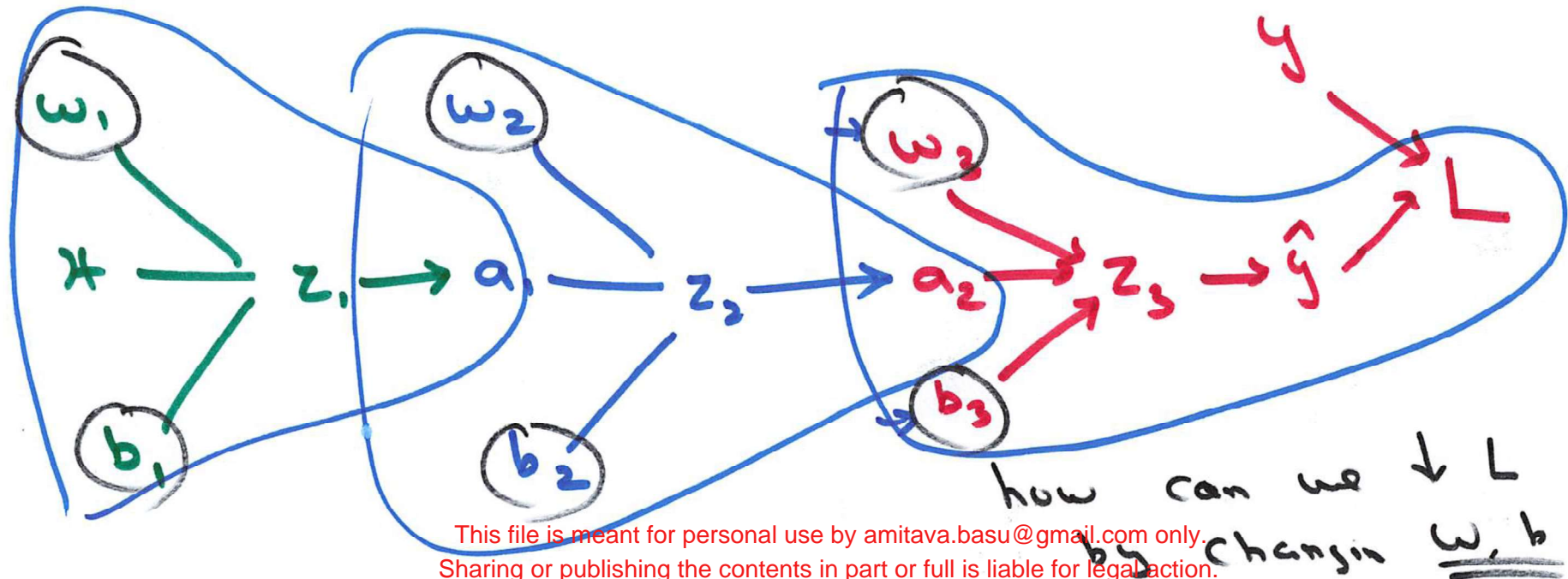


$$L(w) = \mathbb{E}(\hat{y} - y)^2$$

∇L ?



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

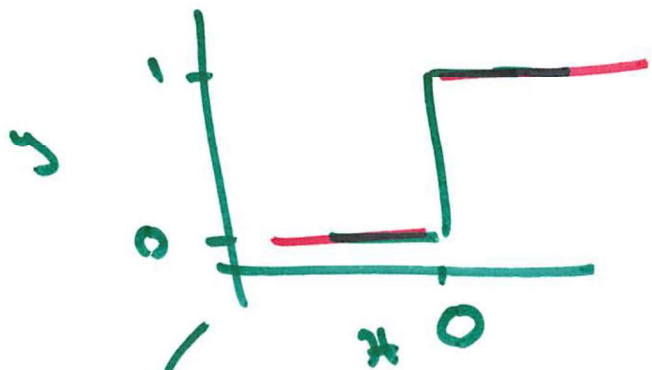
$$\frac{\partial a_2}{\partial w_2}$$

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

non linear func

$$a = f(z)$$

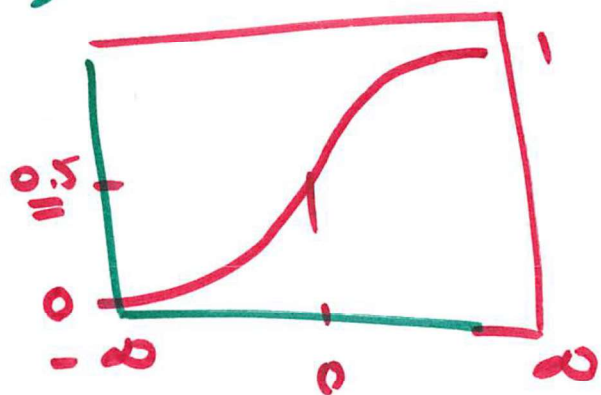




$$y = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

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

$$\underline{\underline{f'}} = \underline{\underline{f''}} - 2 \underline{\underline{(\text{slope of } f)}}$$



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

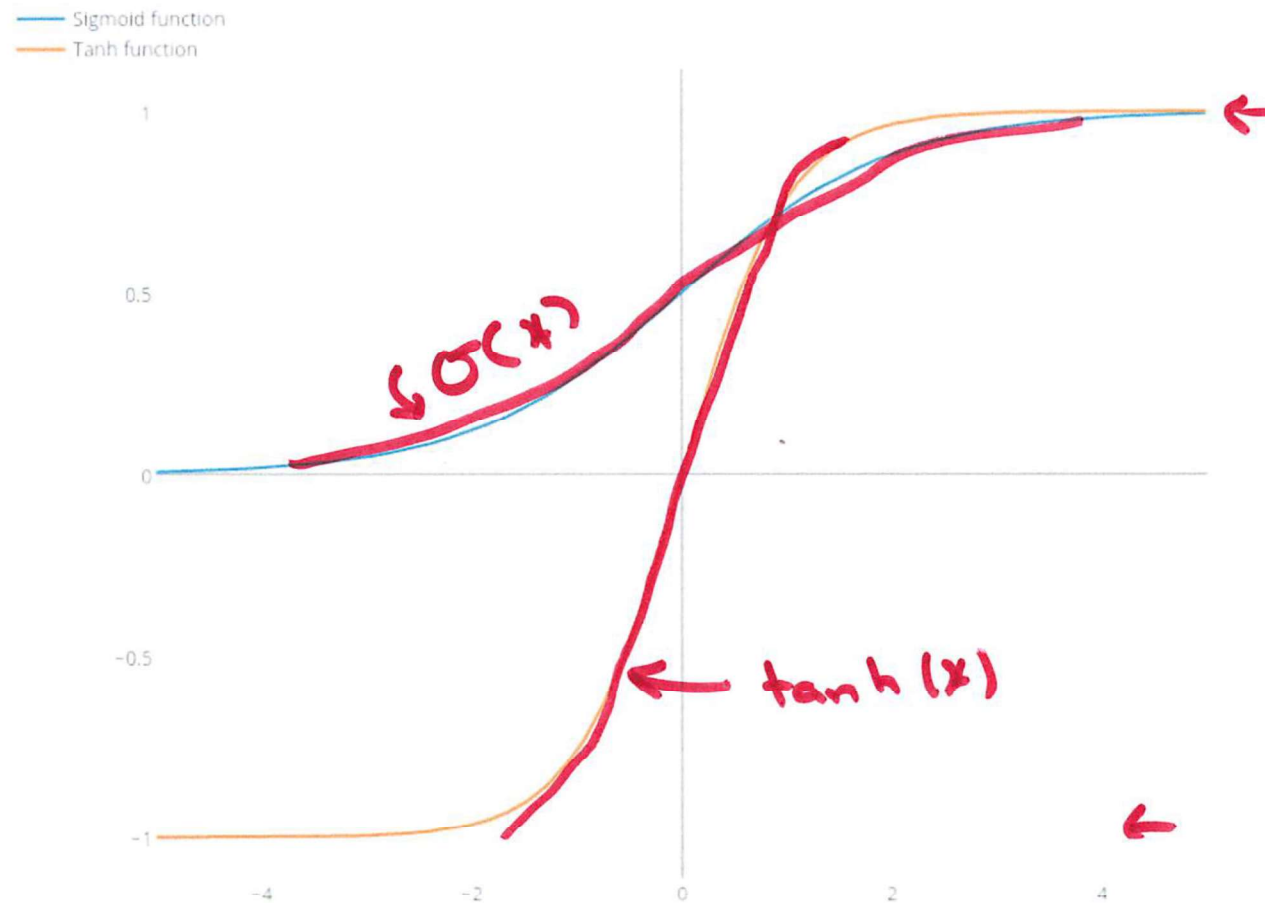
$$y = \sigma(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1}$$

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

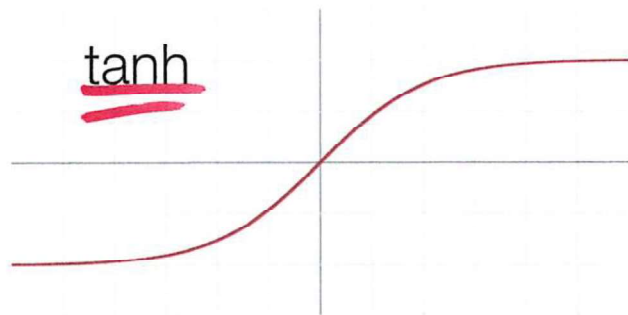
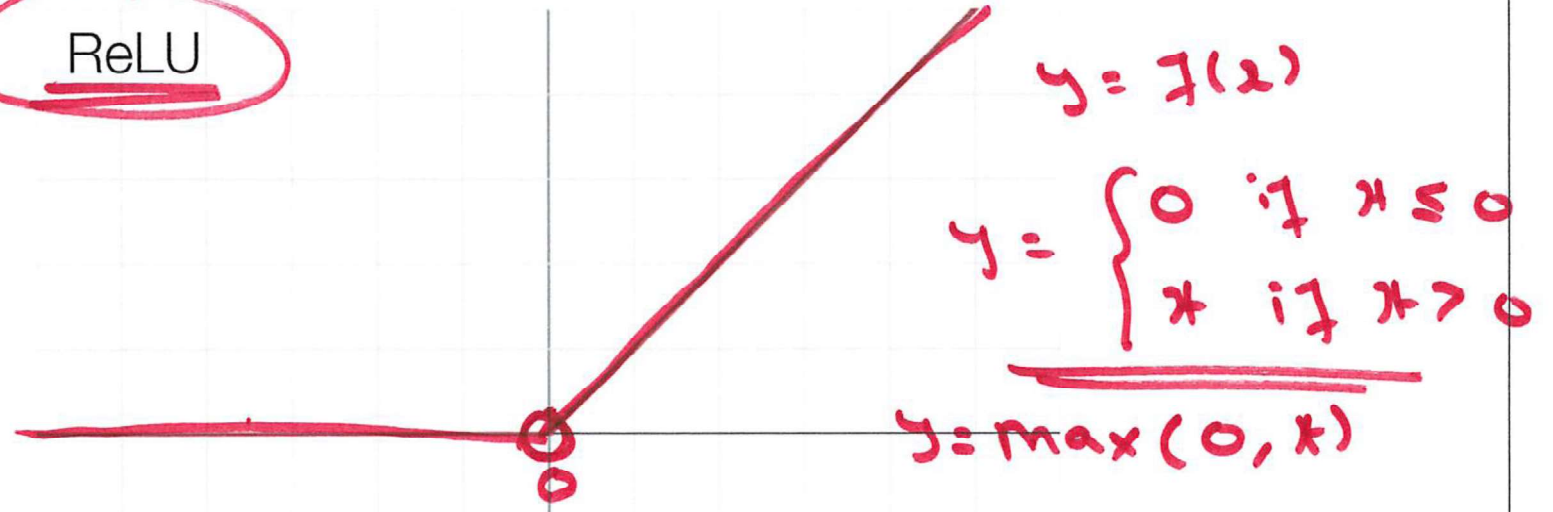
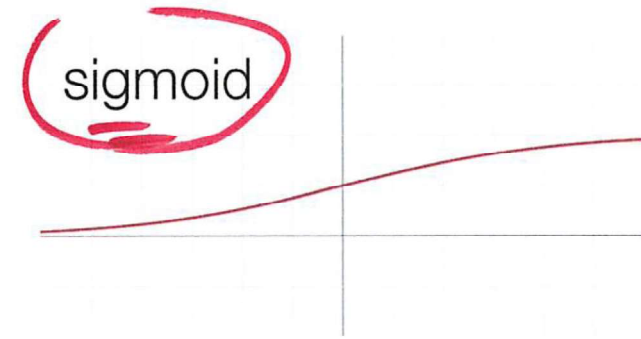
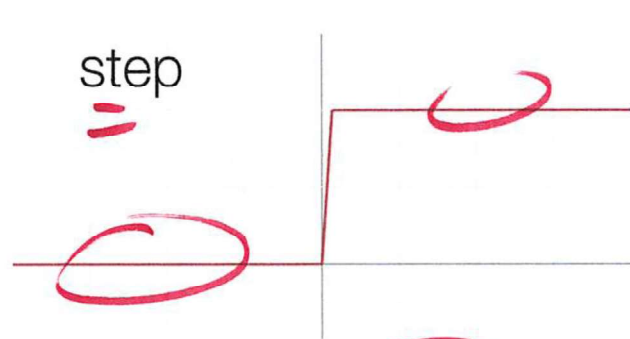
$$y = \tanh(x) = 2\sigma(2x) - 1 \rightarrow -1 \text{ to } 1$$

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

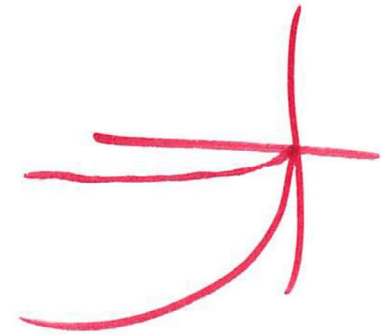
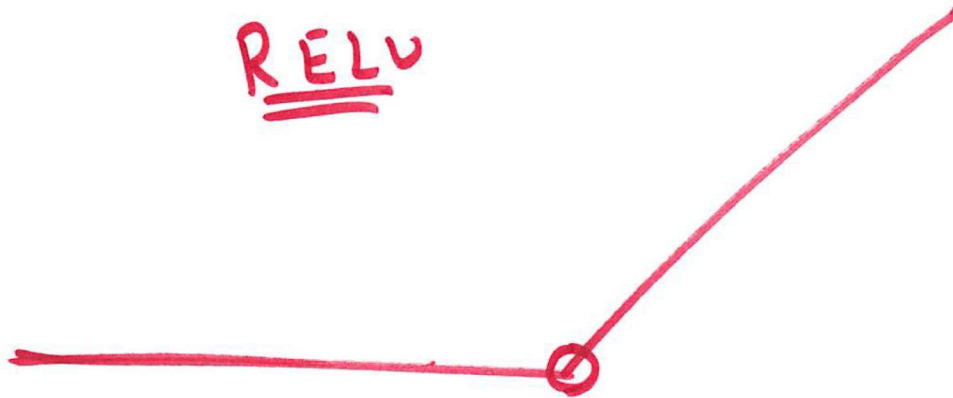
$$\tanh(x) = 2\sigma(2x) - 1$$



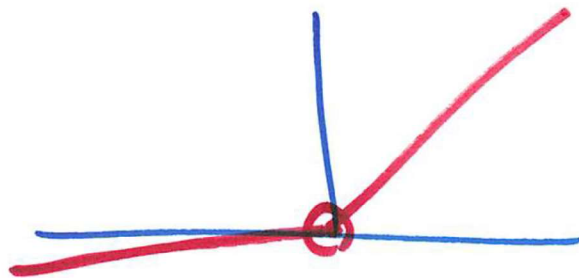
Activation Functions



ReLU

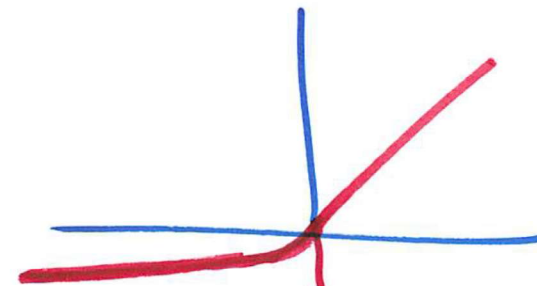


leaky ReLU



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

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

Classifiers

binary

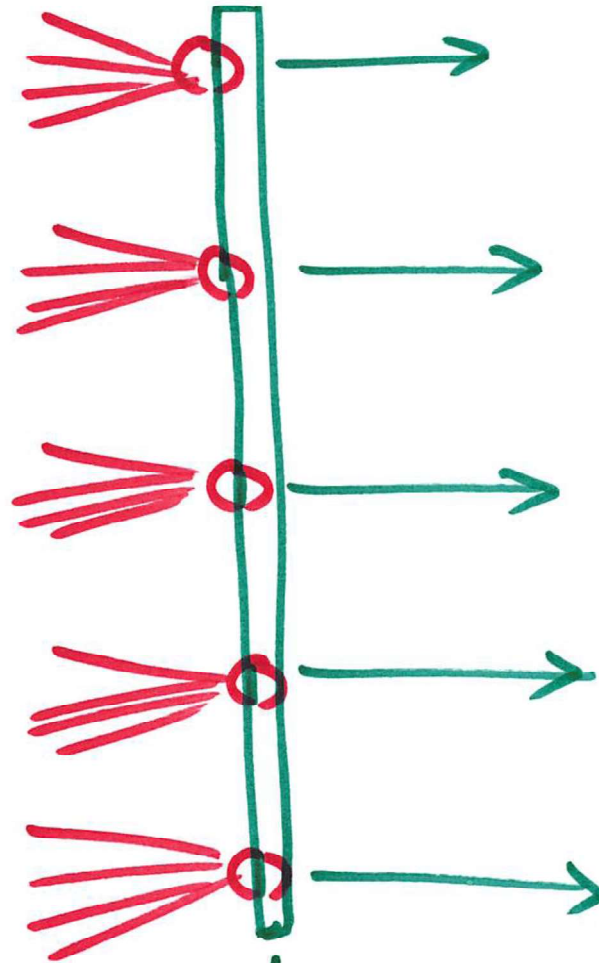
Sigmoid/tanh

multiclass

A	B	C	D	E
.

Softmax ①

$\gamma(\dots) \rightarrow$



Softmax activation

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

~~$$a_i = \gamma(z_i)$$~~