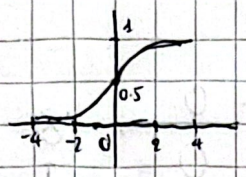
Section 3: Activation Function

- Sigmoid function: $\phi(z) = \frac{1}{1+e^{-z}}$

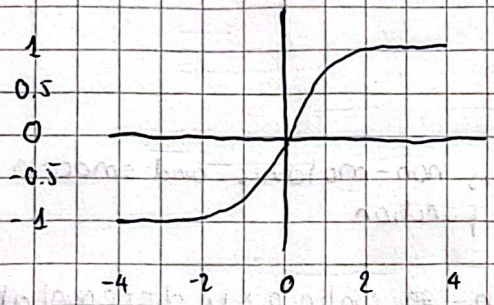


$$y = \frac{1}{1+e^{-x}}$$

$$\frac{dy}{dx} = -\left(\frac{1}{1+e^{-x}}\right)^2 (-e^{-x}) = y(1-y) \leq 0.25$$

Problem in sigmoid: Assume there are 4 layers with sigmoid activate function, then the product of derivatives is $0.2^4 \approx 0.00279$.
→ vanishing gradient

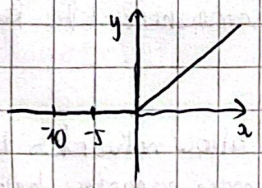
- Tanh Activation: $f(x) = \frac{1-e^{-2x}}{1+e^{-2x}}$



$$\frac{d\tanh(x)}{dx} = 1 - \tanh(x)^2$$

Can be vanish but still better than sigmoid

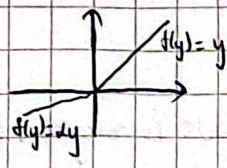
- ReLU (Rectified Linear Unit): $R(x) = \max(0, x)$



$$f(x) = \begin{cases} 0, & x<0 \\ x, & x \geq 0 \end{cases} \rightarrow f'(x) = \begin{cases} 1, & x>0 \\ 0, & \text{otherwise} \end{cases}$$

Problem: Input is negative, output is zero and gradient will die.

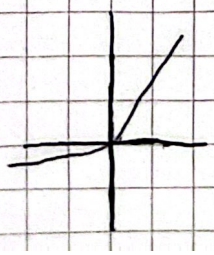
- PreLU (Parametric ReLU):



$$f(x) = \begin{cases} \alpha x, & x<0 \\ x, & x \geq 0 \end{cases} \Rightarrow f'(x) = \begin{cases} \alpha & \text{for } x<0 \\ 1 & \text{for } x \geq 0 \end{cases}$$

- ELU (Exponentially Linear Units): $f(x) = \begin{cases} x, & x>0 \\ \alpha(e^x - 1), & x \leq 0 \end{cases}$

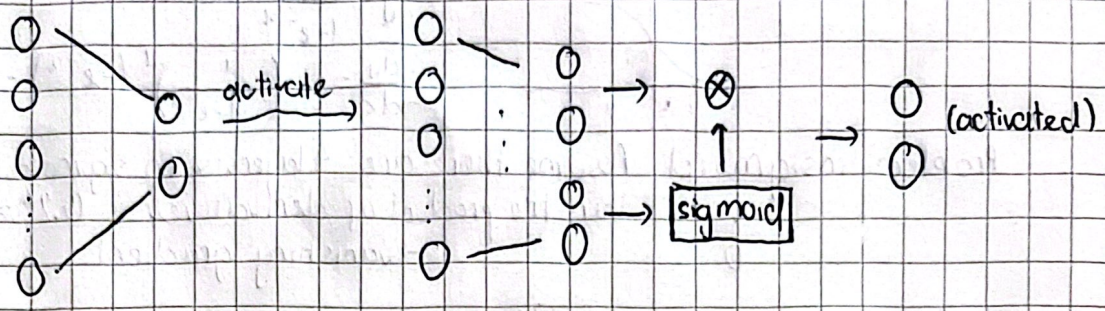
$$\Rightarrow f'(x) = \begin{cases} 1 & \text{if } x>0 \\ f(x) + \alpha, & x \leq 0 \end{cases}$$



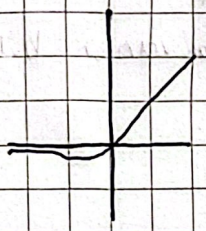
- GLU (Gated Linear Units)

From a layer of dimension d, output double its dimension (dx2)
Then: $(XW_1 + b_1) \otimes \text{sigmoid}(XW_2 + b_2)$

Example: d=2



- Swish : $f(x) = x \cdot \frac{1}{1+e^{-x}}$



Swish is non-motonic and smooth function

- Smooth. ⊛ continuously differentiable everywhere, help gradients flow better than being zero somewhere like ReLU

- Non-monocity : allow networks to present more complex behaviour

- Mish activation:

- Soft plus : ~~$f'(x)$~~ $= \frac{1}{\beta} \log(1 + e^{\beta x})$

Properties: · If $\beta = 1$, the derivative of softplus is sigmoid
· If $\beta \to \infty$, Soft+ becomes ReLU
· $\beta \to 0$, _____ linear

- Mish activation: $\text{Mish} = x \tanh(\text{Softplus}(x))$
$= x \tanh(\ln(1+e^x))$