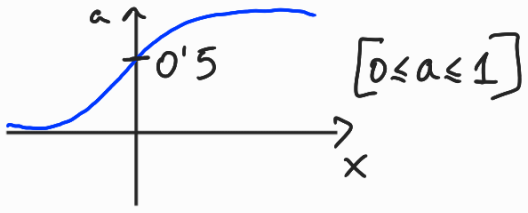
- Sigmoid function: $a(z) = \frac{1}{1+e^{-z}}$

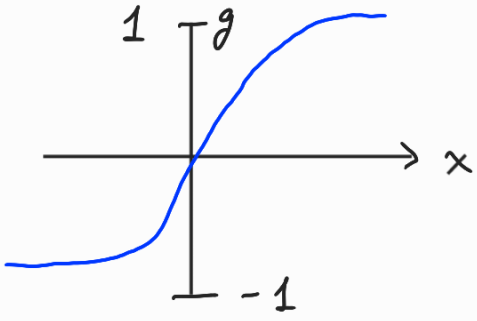


Normally used only in the output layer

$a'(z) = a(1-a)$

- Hyperbolic tangent function:

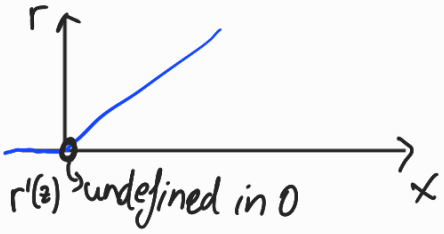
$$g(z^{(i)}) = \tanh(z^{[i]}) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$$



Usually better than sigmoid, because the mean of the activations are closer to 0 by centering data.

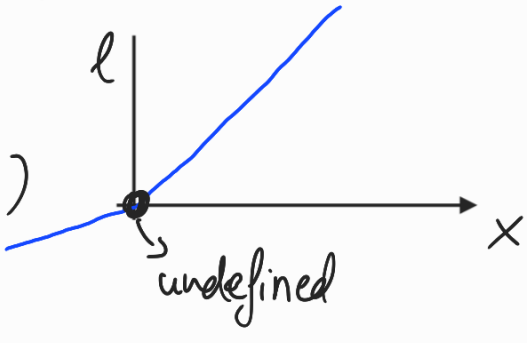
$g'(z) = 1 - (\tanh(z))^2 = 1 - g^2$

- ReLU: $r(z) = \max(0, z)$



$$r(z) = \begin{cases} 0 & \text{if } z < 0 \\ 1 & \text{if } z \geq 1 \end{cases}$$

- Leaky ReLU: $l(z) = \max(0'01z, z)$



$$l'(z) = \begin{cases} 0'01 & \text{if } z < 0 \\ 1 & \text{if } z > 0 \end{cases}$$

How to choose?

- Output between 0 or 1 → Sigmoid
- Learn FAST: ReLUs, because the function slope increases.
- Linear activation functions: for regression tasks in the output layer, but in the hidden layers NO.