One hidden layer
Neural Network

Derivatives of
activation functions

deeplearning.ai

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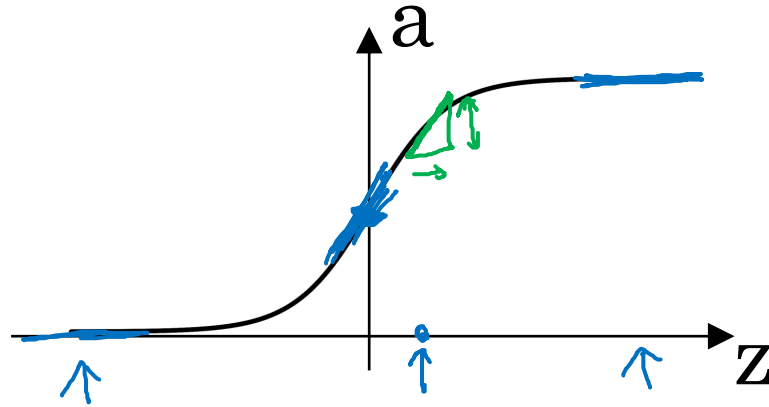
$$g(z) = \frac{1}{1 + e^{-z}}$$

$$a = g(z) = \frac{1}{1+e^{-z}}$$

$$z = 10. \quad g(z) \approx 1$$
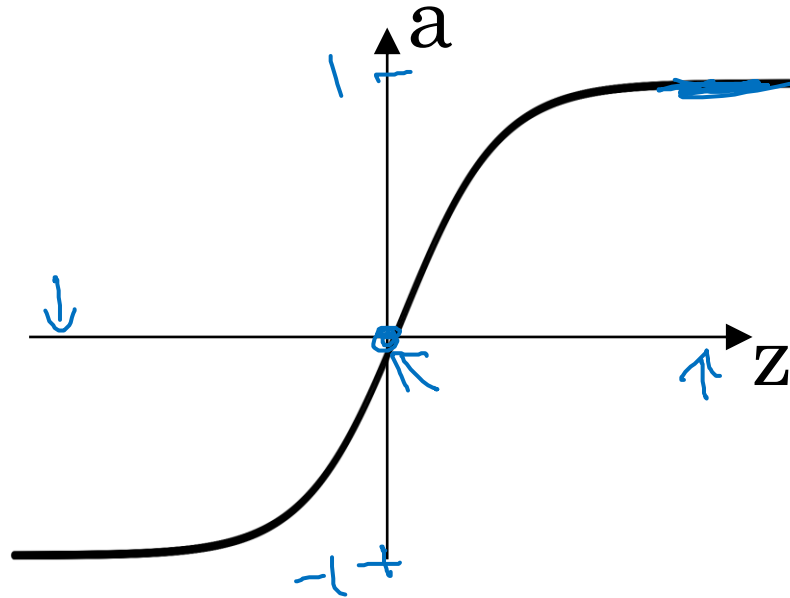$$\frac{d}{dz} g(z) \approx 1 \, (1-1) \approx 0$$

$$z = -10 \quad g(z) \approx 0$$
$$\frac{d}{dz} g(z) \approx 0 \cdot (1-0) \approx 0$$

$$z = 0 \quad g(z) = \frac{1}{2}$$
$$\frac{d}{dz} g(z) = \frac{1}{2}(1-\frac{1}{2}) = \frac{1}{4}$$

$$g'(z) = \boxed{\frac{d}{dz} g(z)} = \text{slope of } g(x) \text{ at } z$$

$$= \frac{1}{1+e^{-z}} \left(1 - \frac{1}{1+e^{-z}} \right)$$

$$= g(z) \left(1 - g(z) \right) \leftarrow$$

$$= \boxed{a \, (1-a)}$$

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

Andrew Ng

Tanh activation function

$g(z) = \tanh(z)$

$$= \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

$g'(z) = \frac{d}{dz} g(z) = $ slope of $g(z)$ at z

$= 1 - (\tanh(z))^2 \leftarrow$

$a = g(z),$ $g'(z) = 1 - a^2$

$z = 10$ $\tanh(z) \approx 1$
$g'(z) \approx 0$
$z = -10$ $\tanh(z) \approx -1$
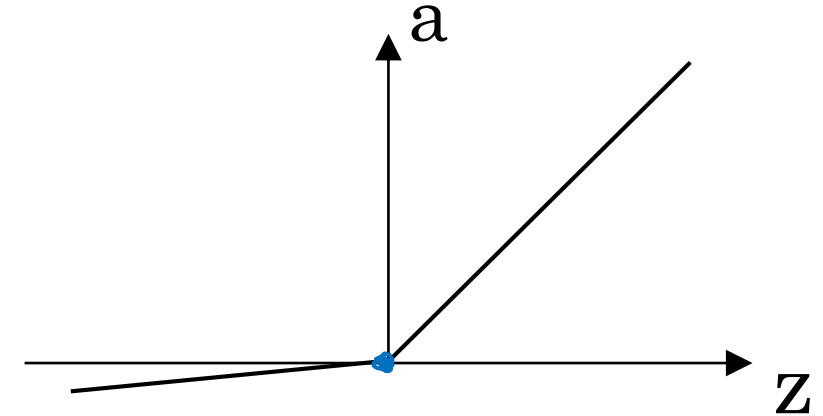$g'(z) \approx 0$
$z = 0$ $\tanh(z) = 0$
$g'(z) = 1$

ReLU and Leaky ReLU
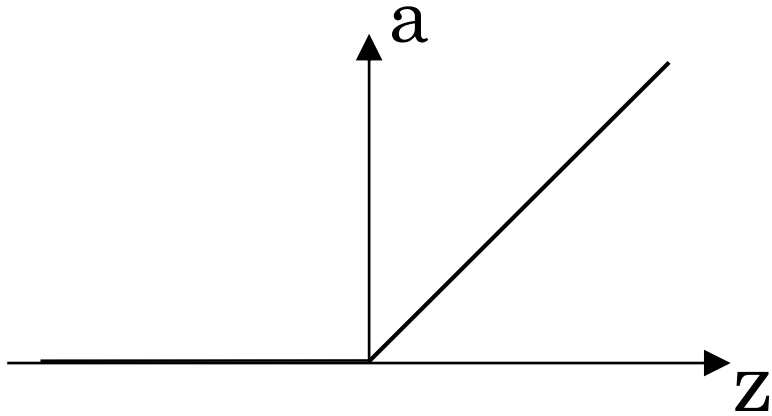
ReLU

$$g(z) = \max(0, z)$$

$$\rightarrow g'(z) = \begin{cases} 0 & \text{if } z < 0 \\ 1 & \text{if } z \geq 0 \\ \text{undef.} & \text{if } z = 0 \end{cases}$$

$z = 0.0000\ldots 0$

Leaky ReLU

$$g(z) = \max(0.01z, z)$$

$$g'(z) = \begin{cases} 0.01 & \text{if } z < 0 \\ 1 & \text{if } z \geq 0 \end{cases}$$

Andrew Ng