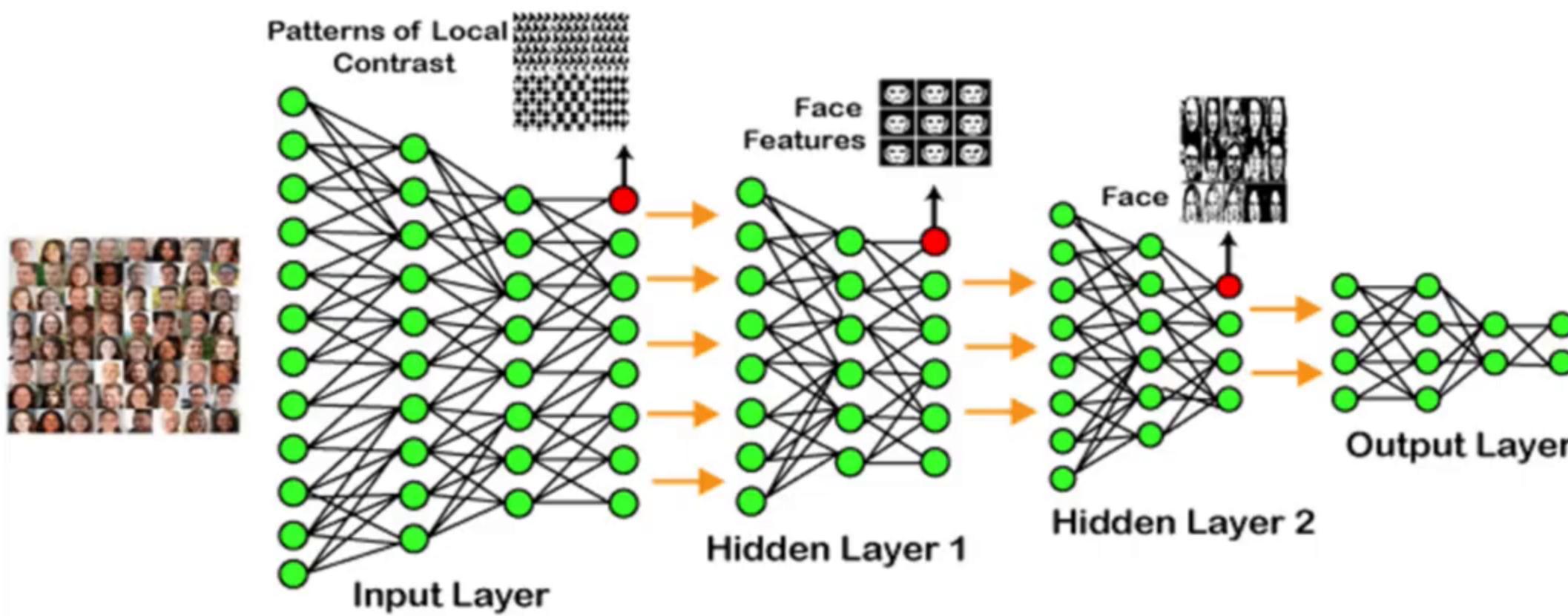


Variants of ReLU

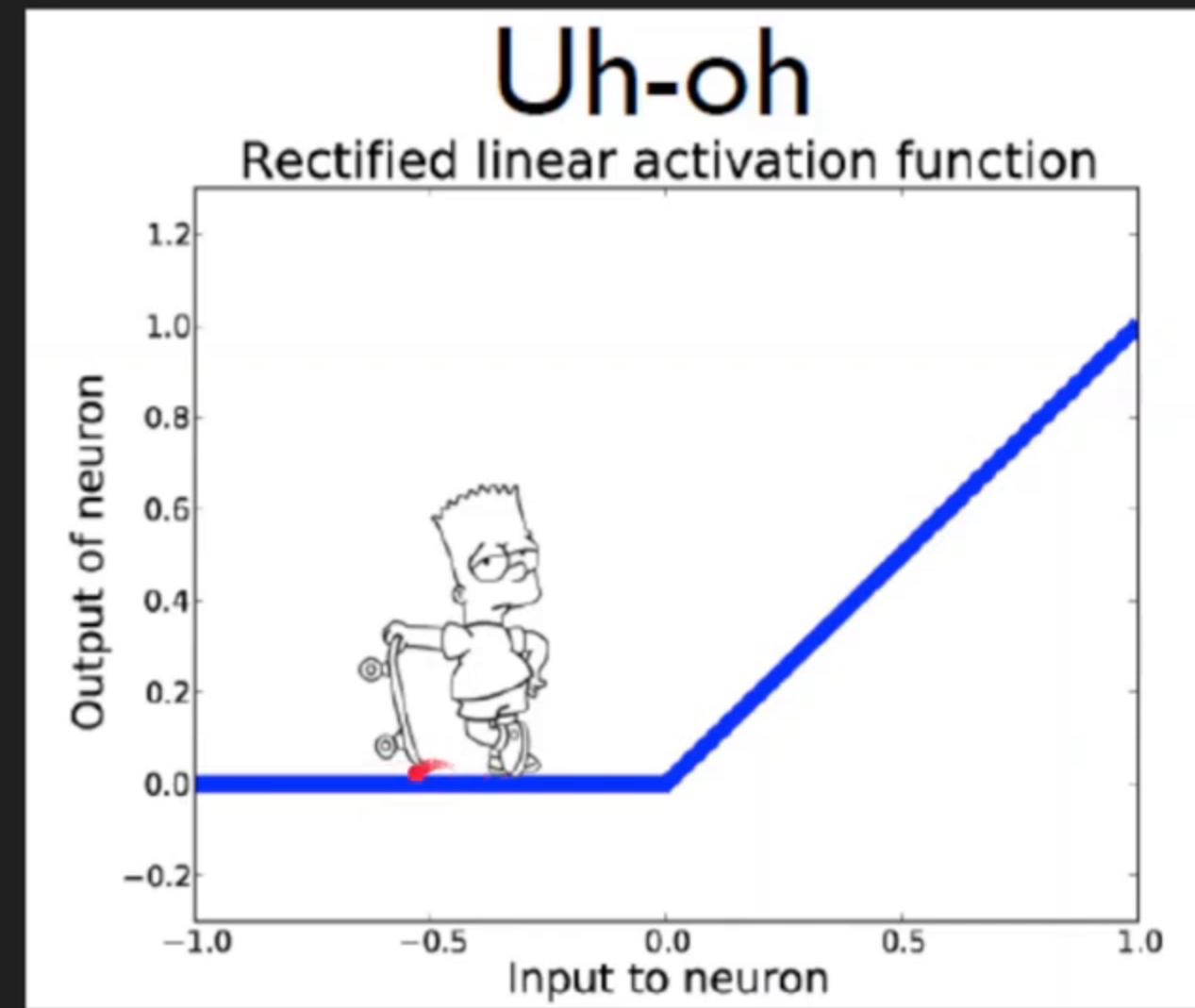
- Leaky ReLU
- Randomized Leaky ReLU
- Parametric ReLU
- Exponential Linear Unit
- Scaled Exponential Linear Unit

Deep Learning



Dead Neurons

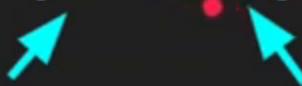
$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$$



Dead Neurons

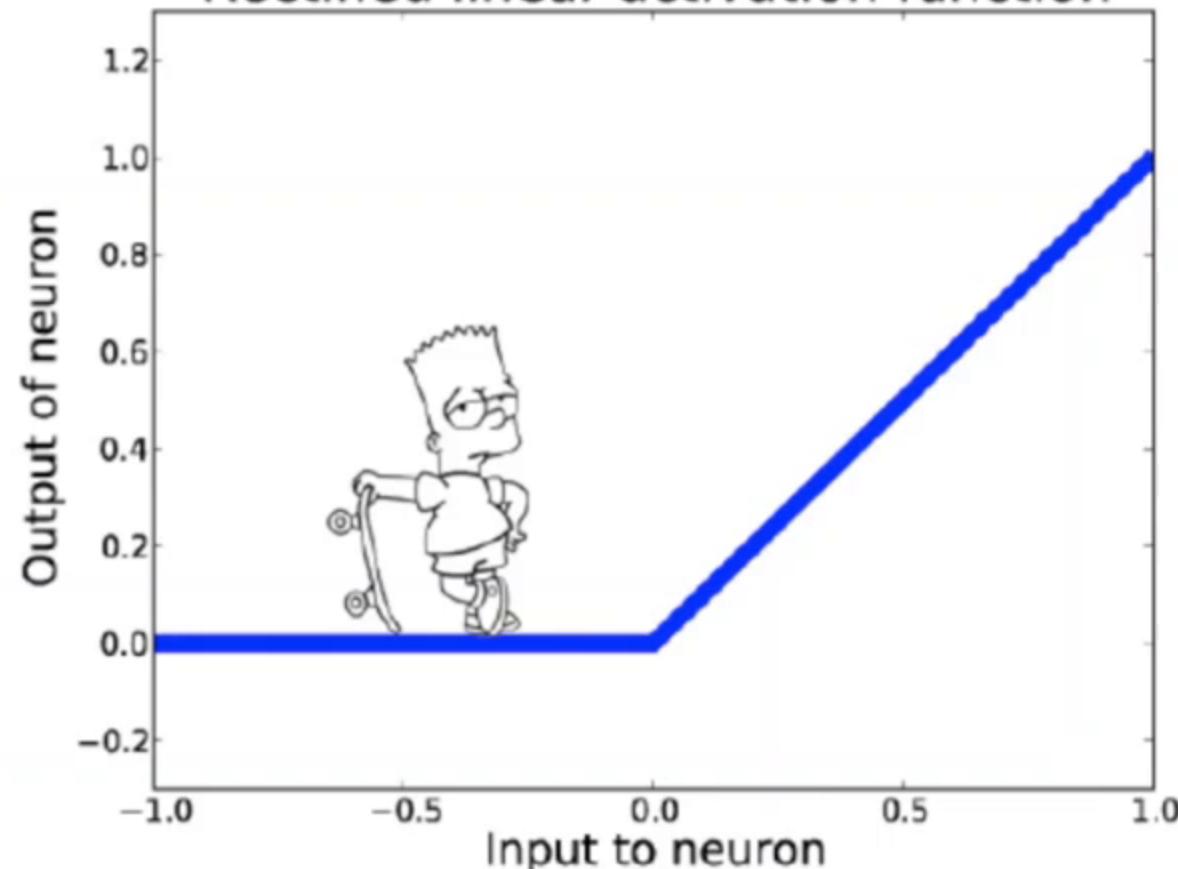
$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$$

$$Y = \text{ReLU}(W^*X + b)$$



Uh-oh

Rectified linear activation function



Dead Neurons

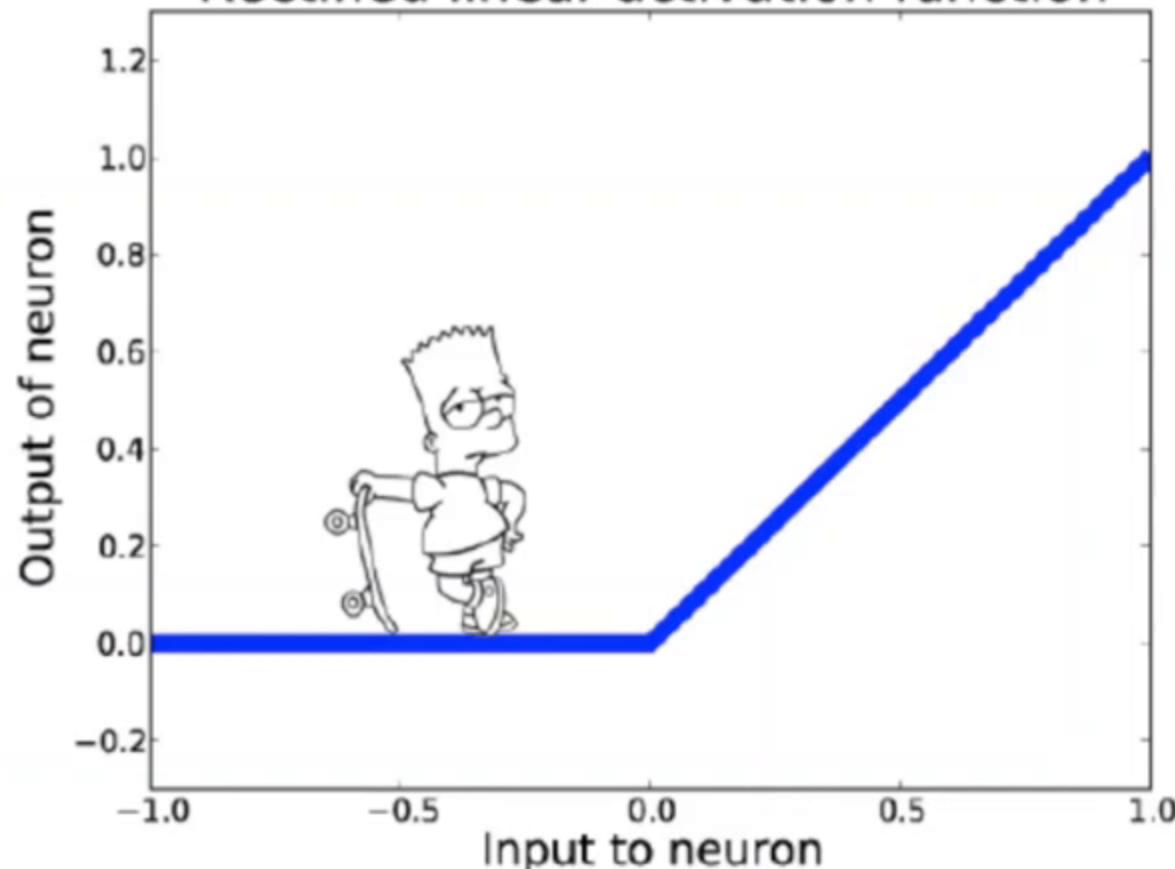
$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$$

$$Y = \text{ReLU}(W^*X + b)$$



Uh-oh

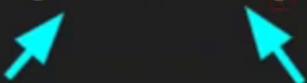
Rectified linear activation function



Dead Neurons

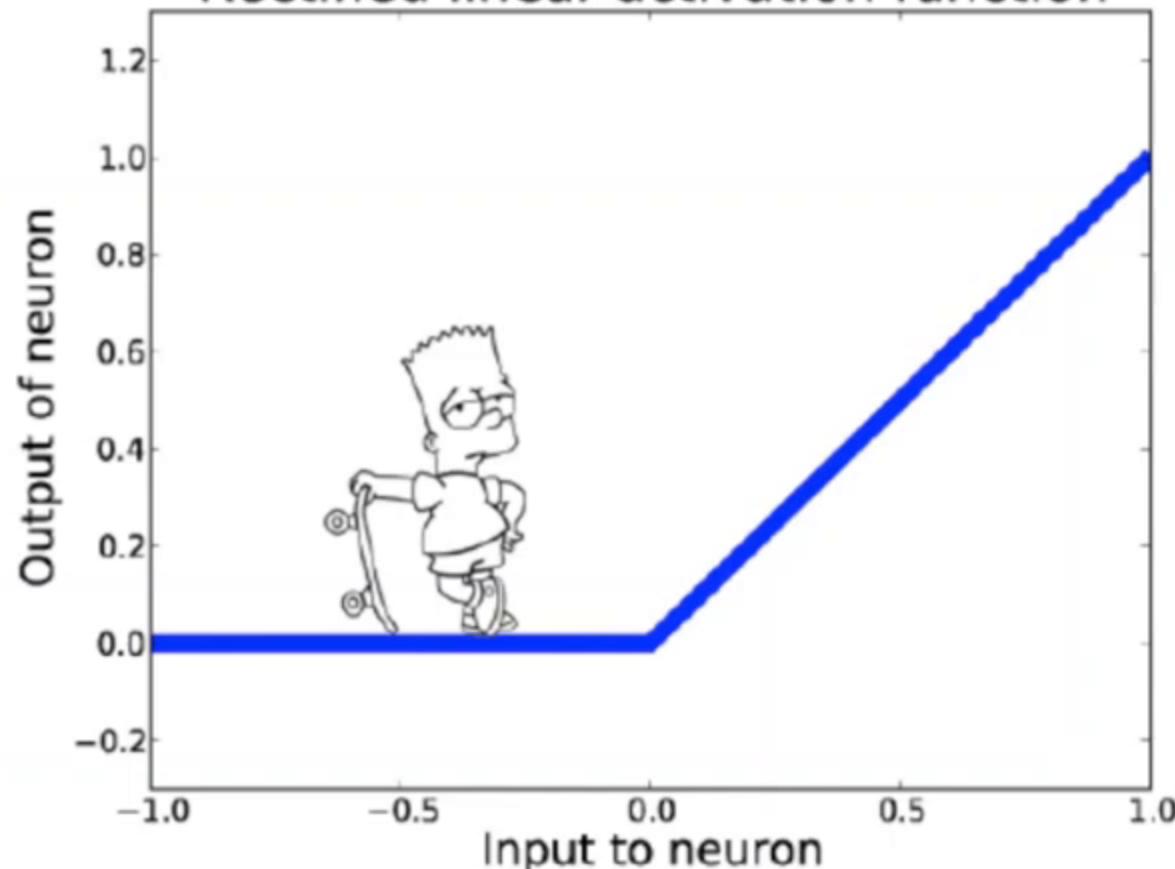
$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$$

$$Y = \text{ReLU}(W^*X + b)$$



Uh-oh

Rectified linear activation function



Solution

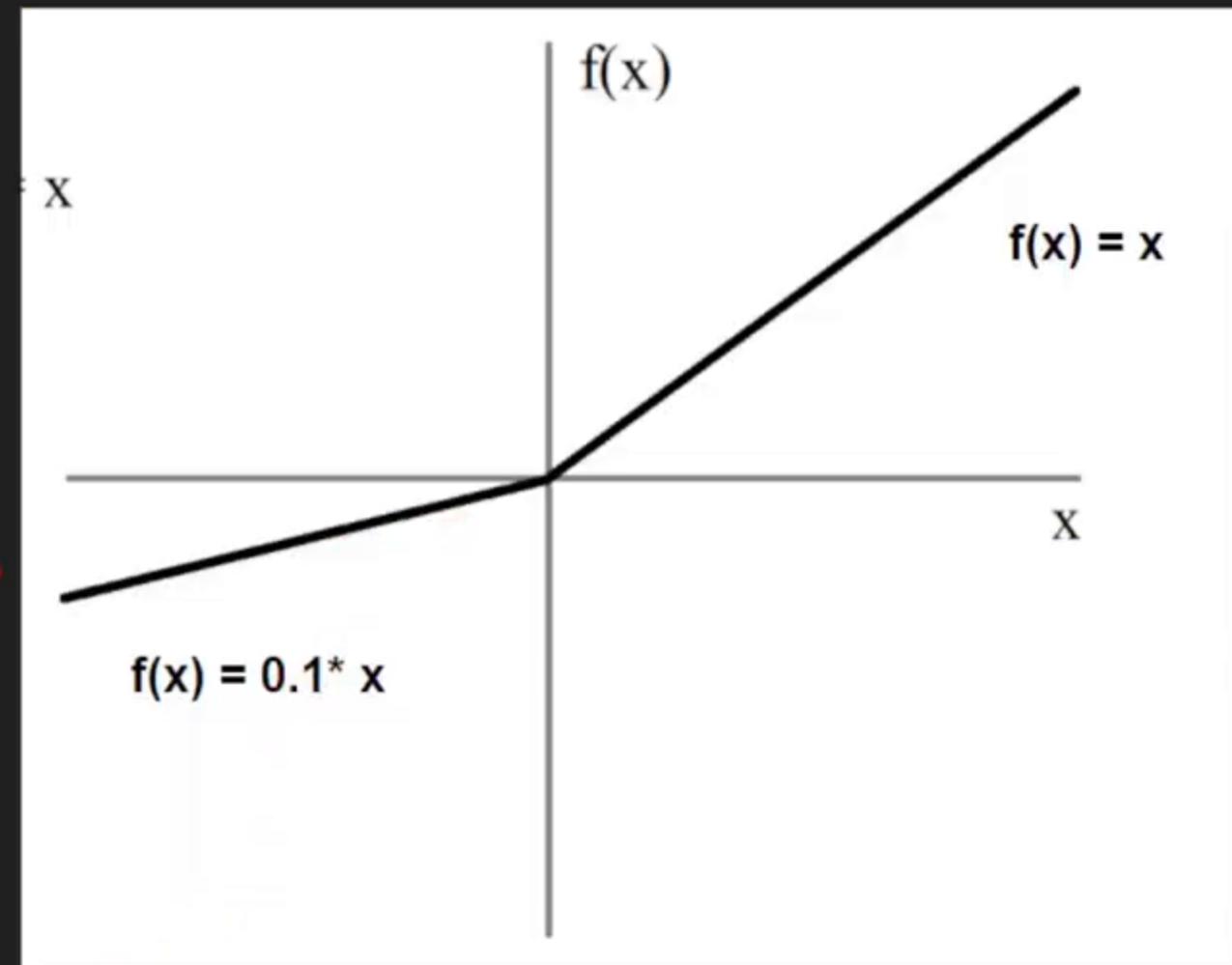
$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$$

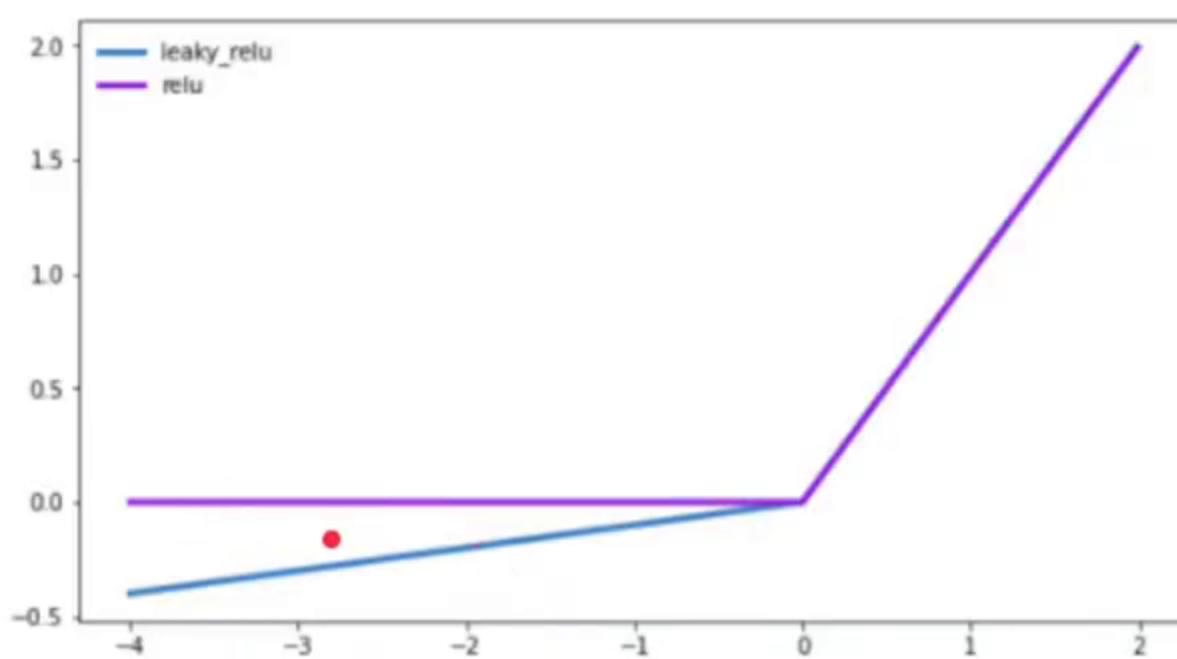
Solution

$$f(x) = \begin{cases} x & \text{for } x < 0 \\ \cancel{0} & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$$



Leaky ReLU

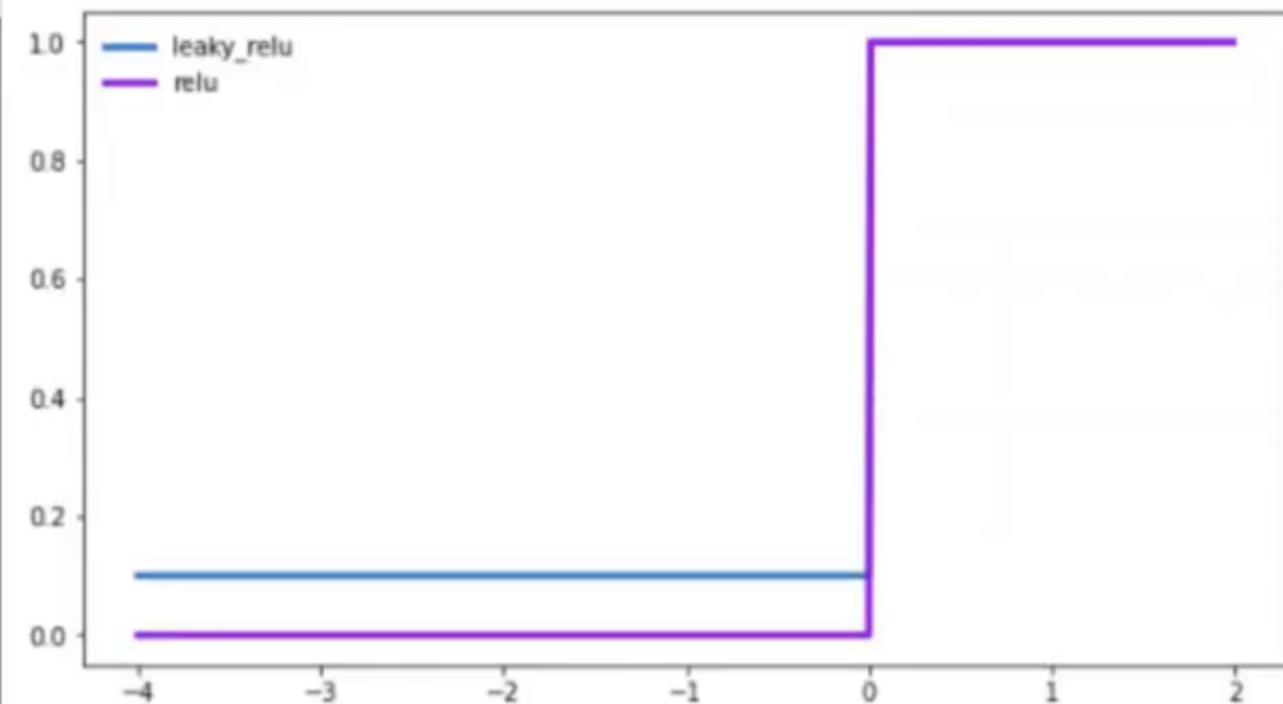


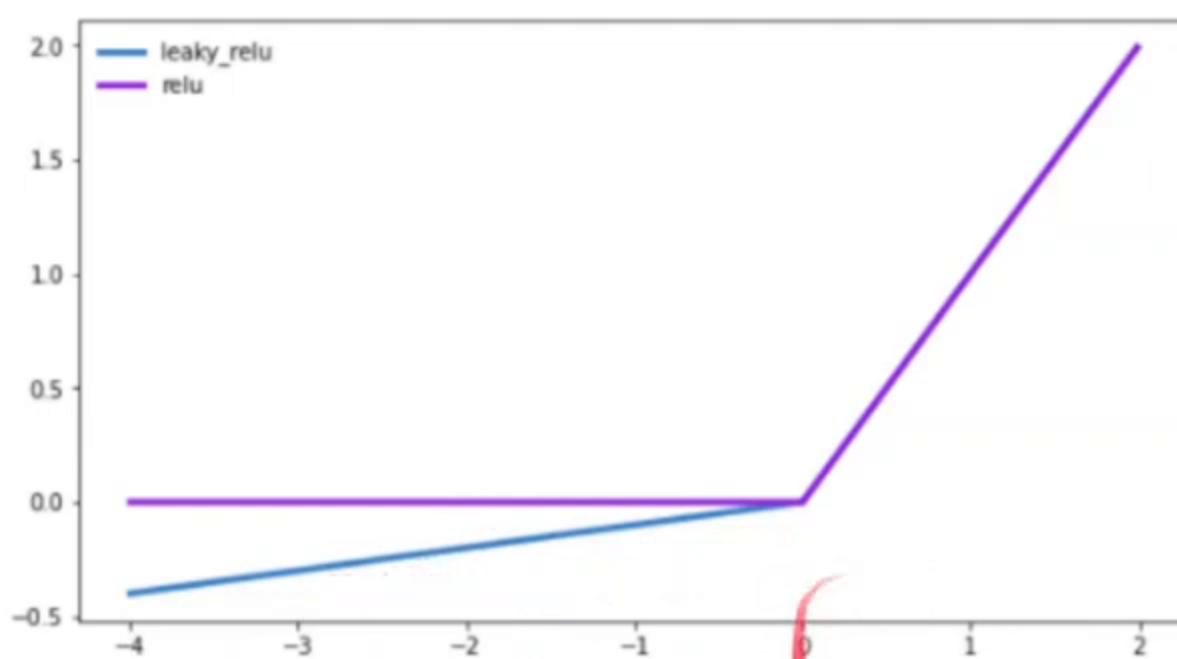


Leaky ReLU

$$f(x) = \max(0.1x, x)$$

Derivatives

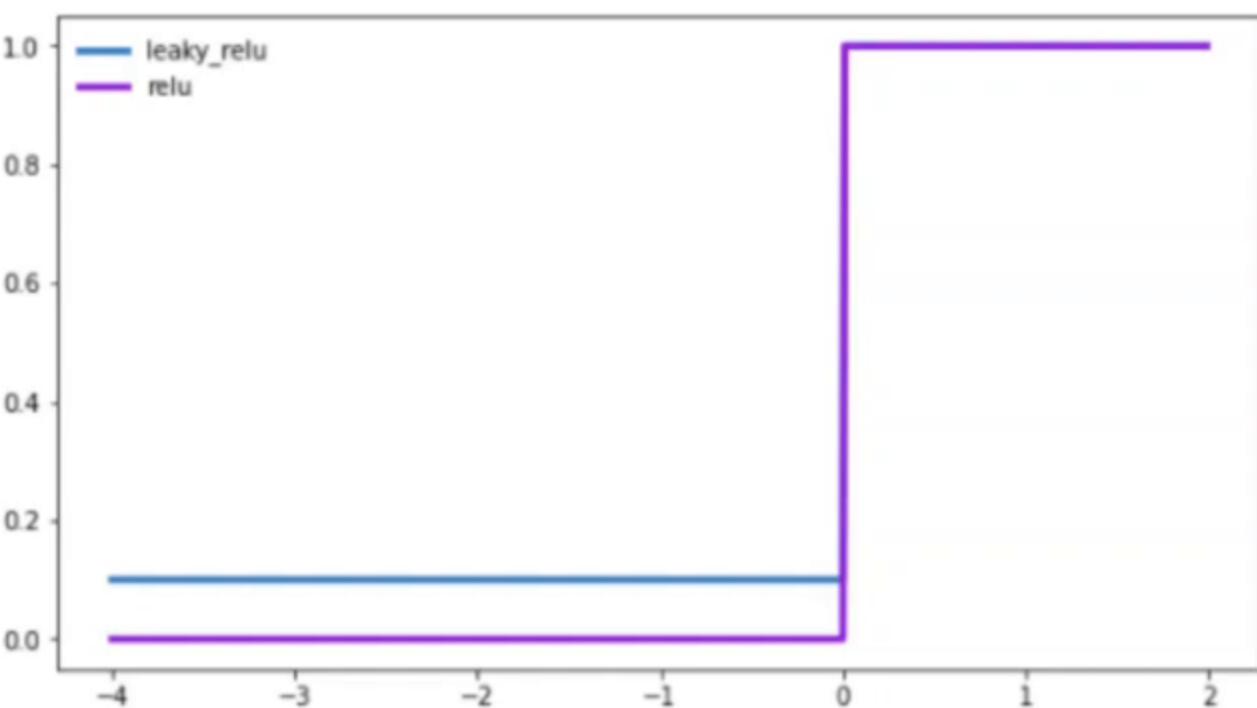


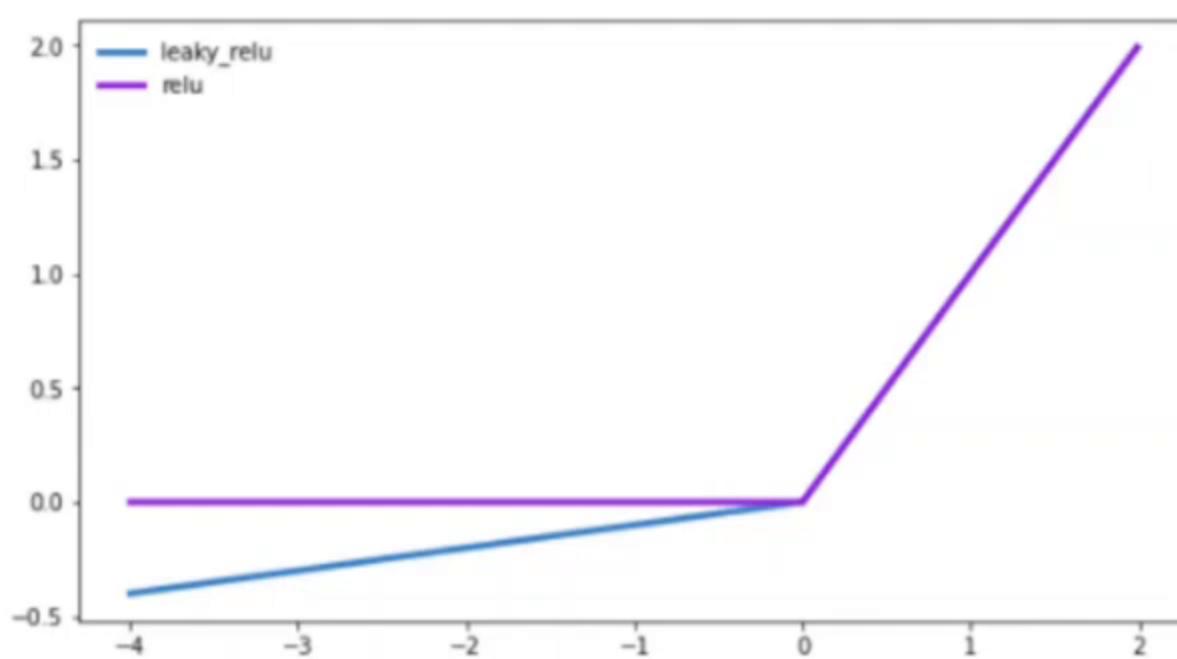


Leaky ReLU

$$f(x) = \max(0.1x, x)$$

Derivatives





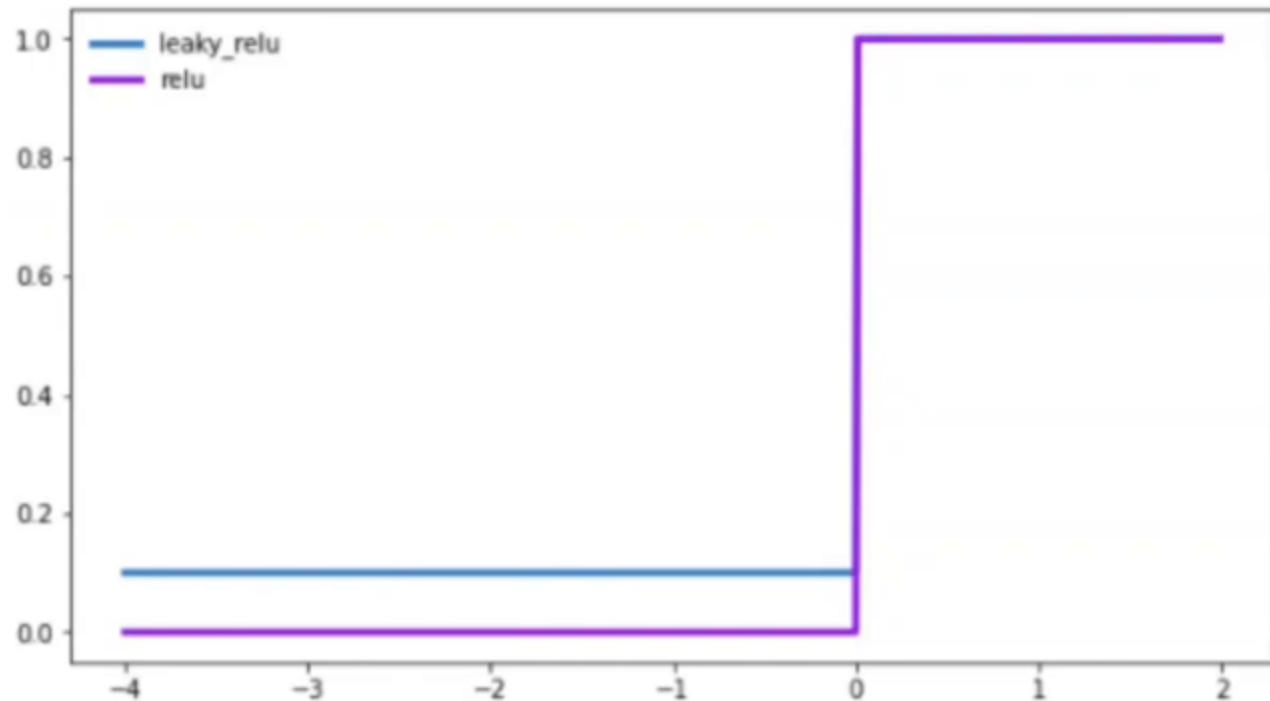
ReLU vs Leaky ReLU

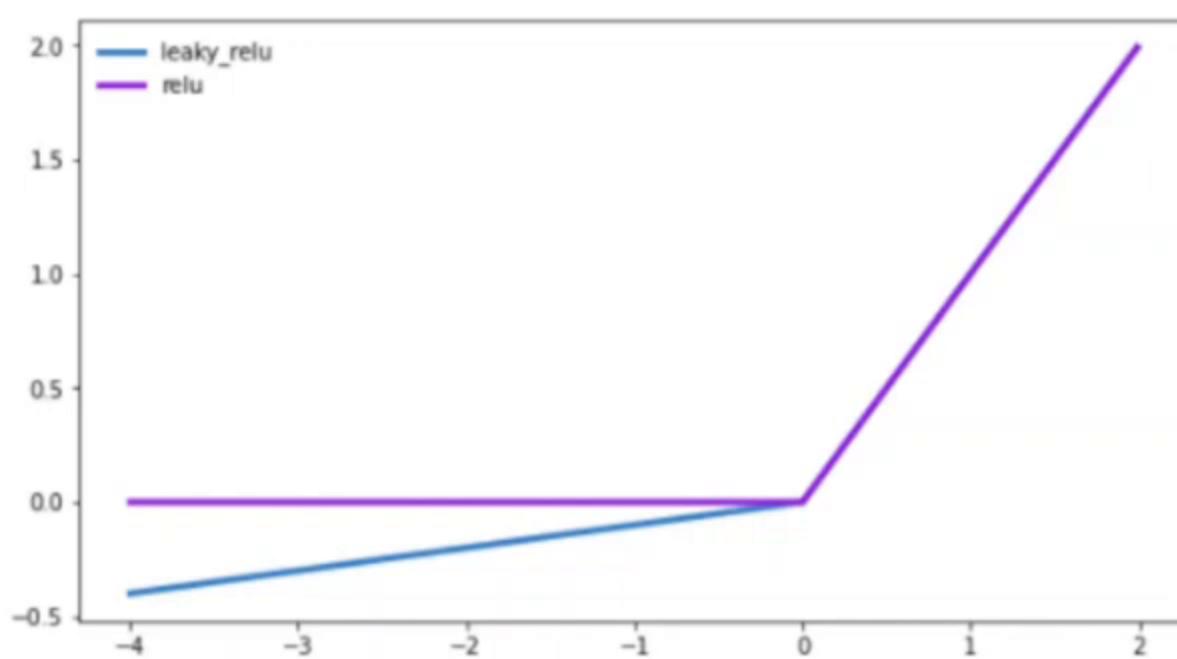
Why 0.1?

Leaky ReLU

$$f(x) = \max(0.1x, x)$$

Derivatives

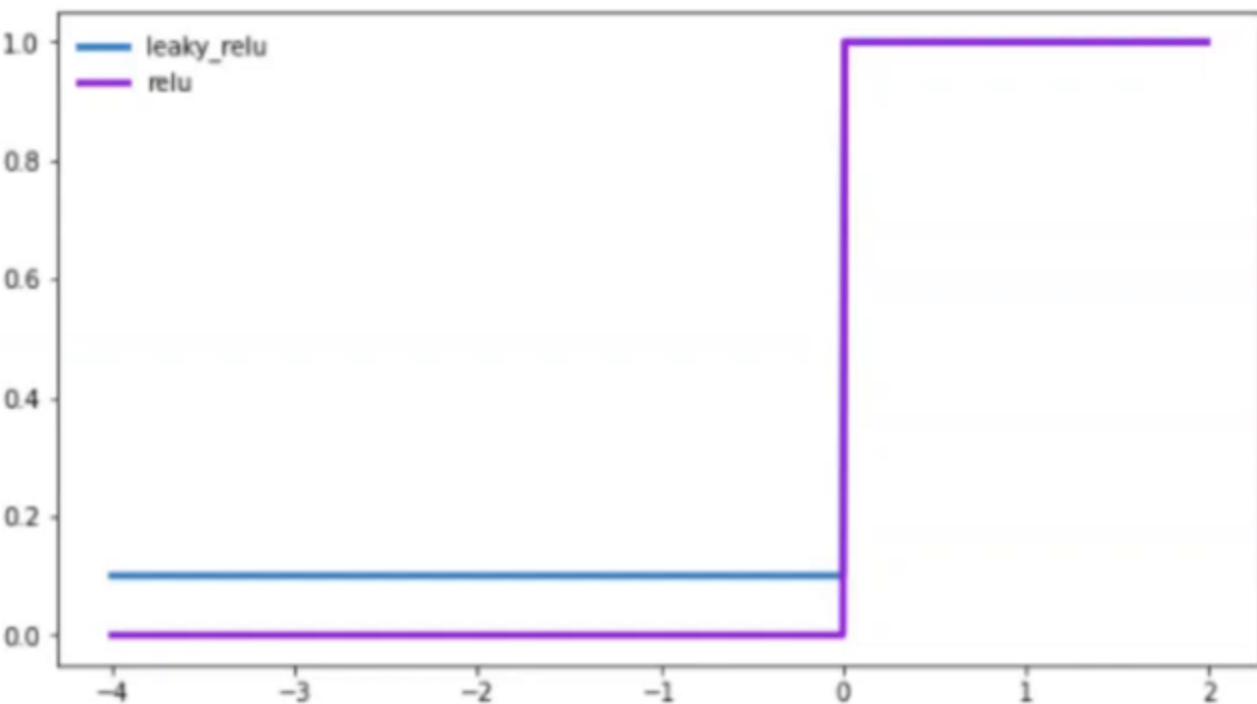


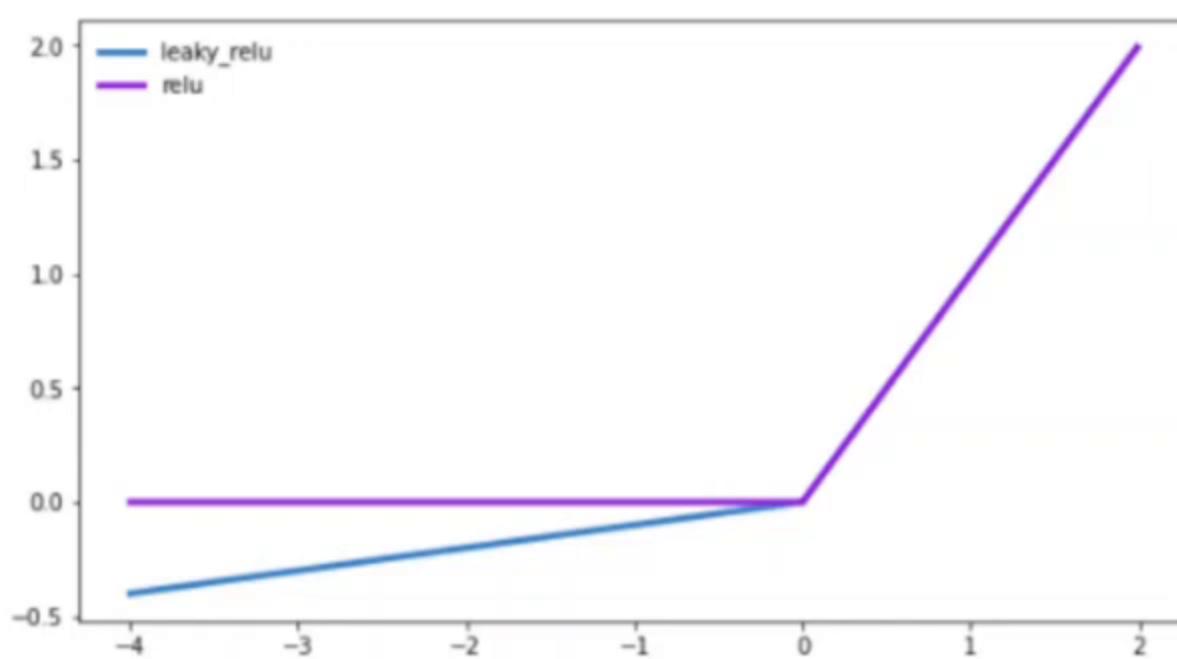


Leaky ReLU

$$f(x) = \max(0.1x, x)$$

Derivatives





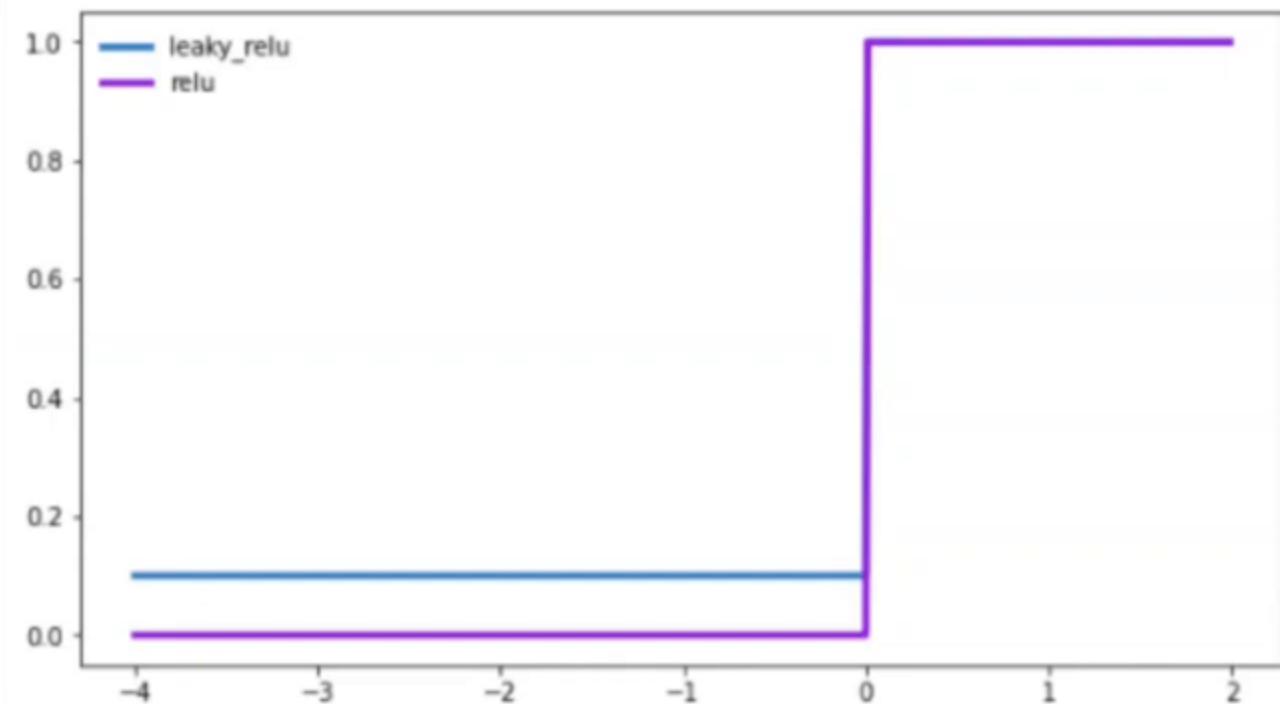
ReLU vs Leaky ReLU

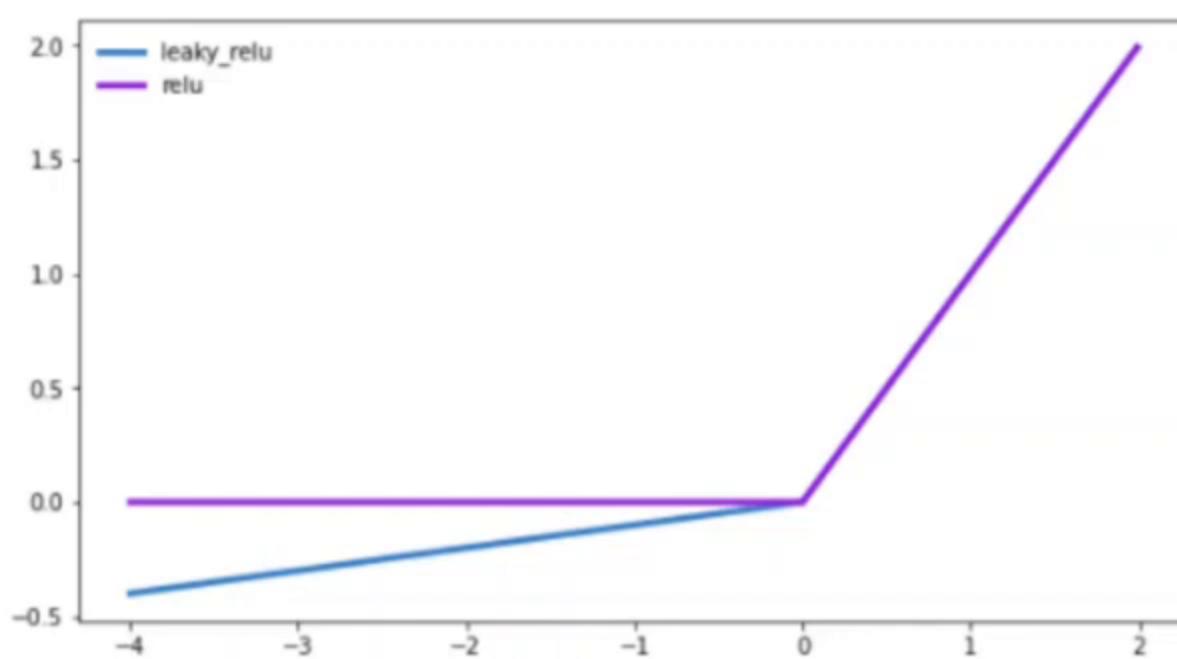
~~Why 0.1?~~ 0.5

Leaky ReLU

$$f(x) = \max(0.1x, x)$$

Derivatives





ReLU vs Leaky ReLU

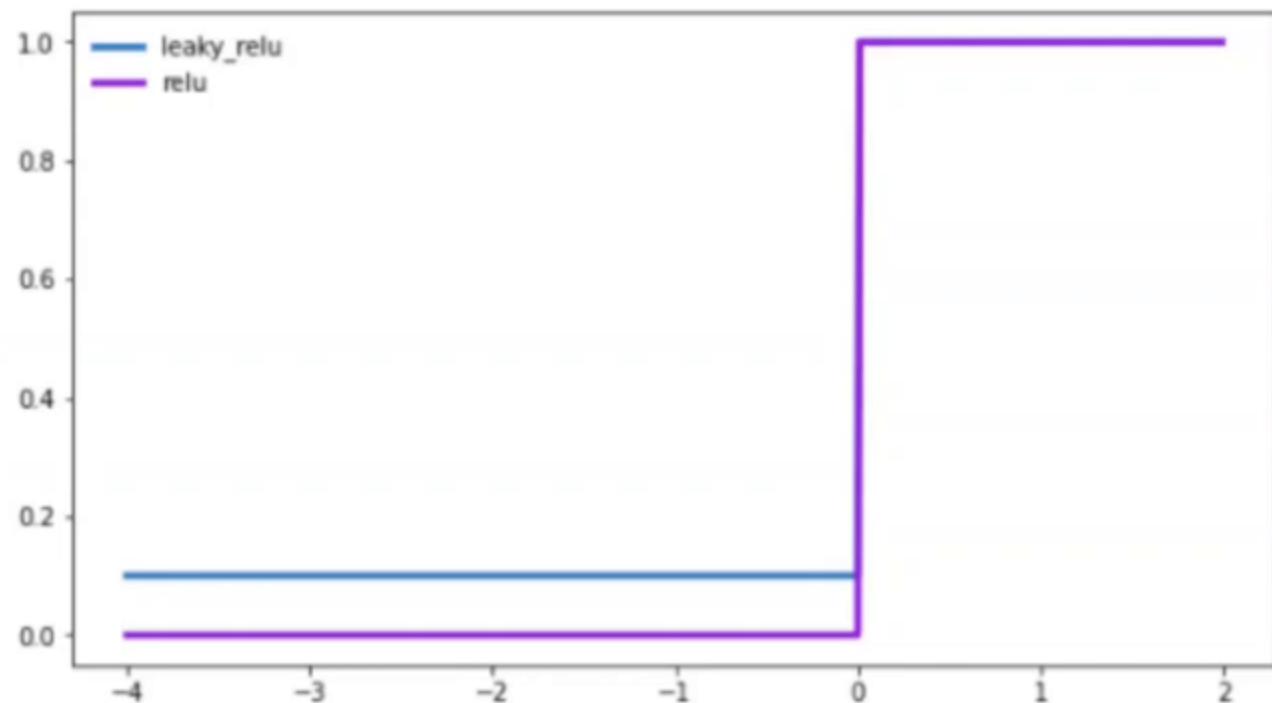
~~Why 0.1?~~ 0.5

Random Value - [0,1]

Leaky ReLU

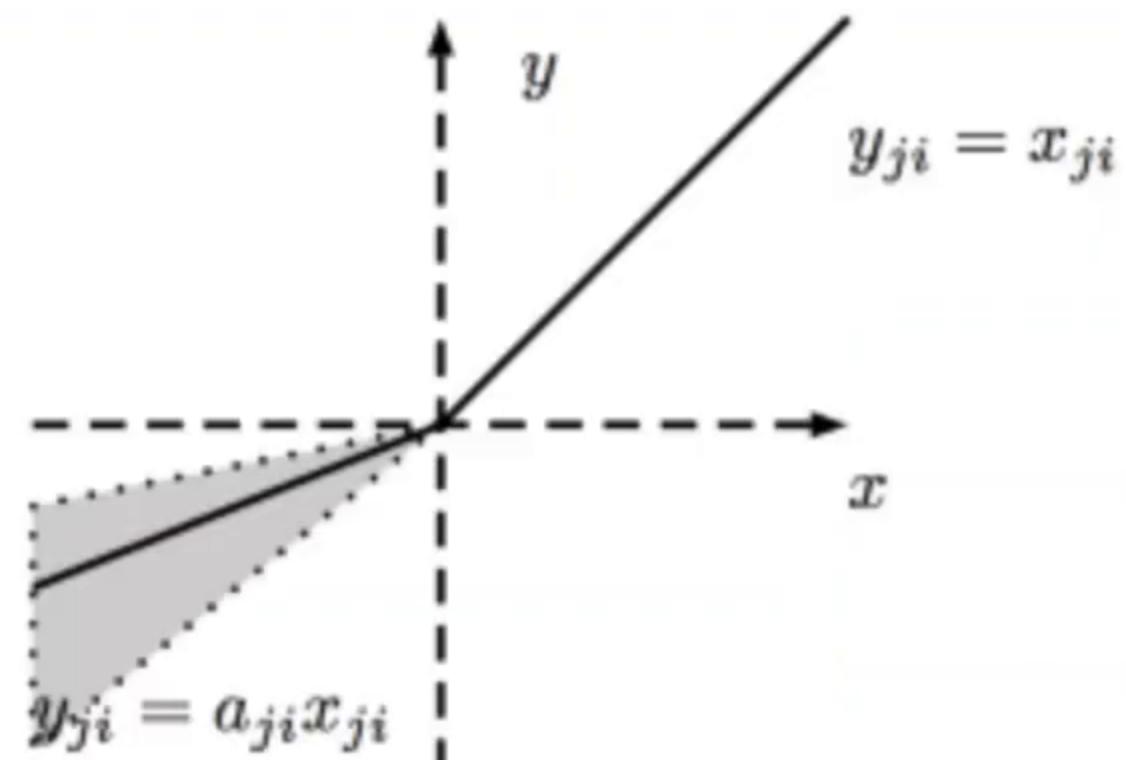
$$f(x) = \max(0.1x, x)$$

Derivatives



Randomized Leaky ReLU

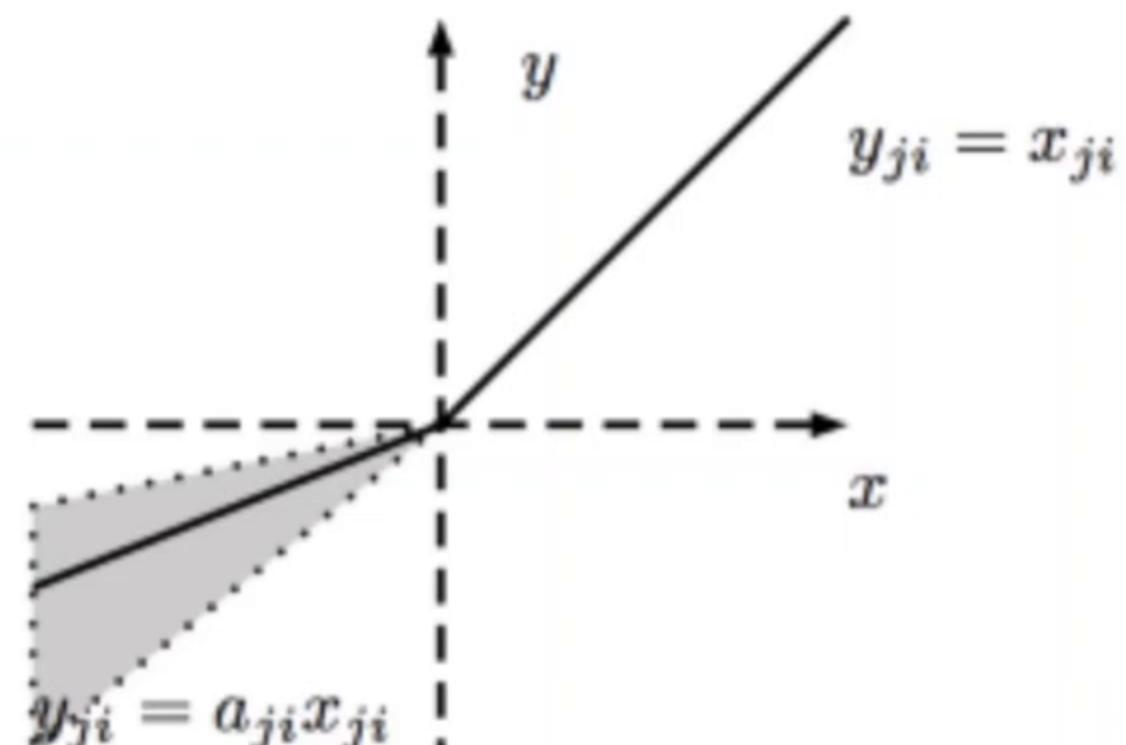
'a' - Random sample from
Uniform Distribution [0,1)



Randomized Leaky ReLU

Randomized Leaky ReLU

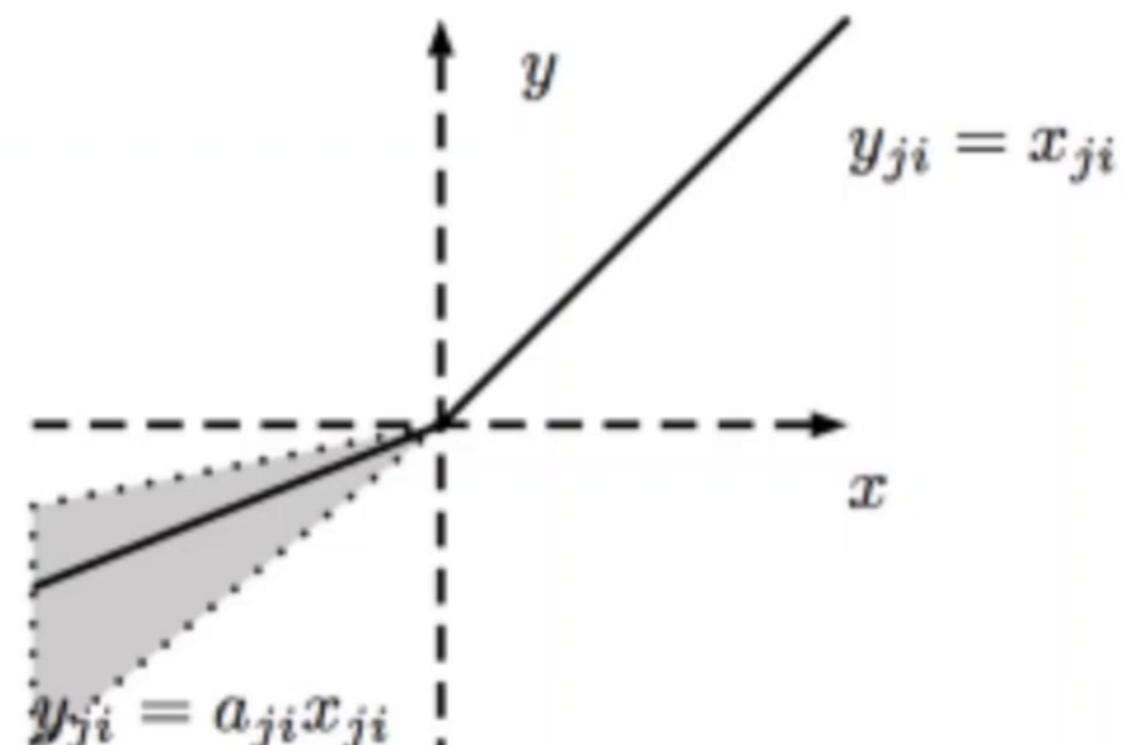
'a' - Random sample from
Uniform Distribution [0,1)



Randomized Leaky ReLU

Randomized Leaky ReLU

'a' - Random sample from
Uniform Distribution [0,1)



Randomized Leaky ReLU

Randomized Leaky ReLU

'a' - Random sample from
Uniform Distribution [0,1)

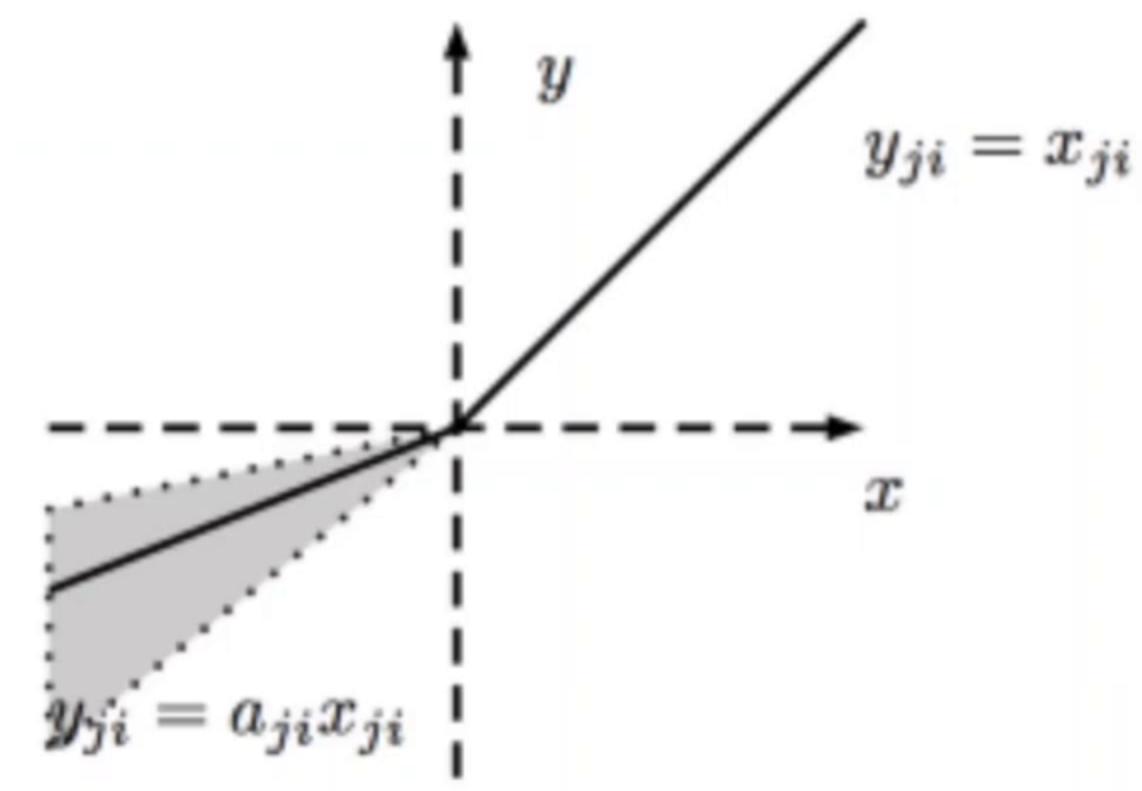
NDSB kaggle competition:

(3,8)

u

$$y_{ji} = \frac{x_{ji}}{\frac{l+u}{2}}$$

Test time



Randomized Leaky ReLU

Randomized Leaky ReLU

'a' - Random sample from
Uniform Distribution [0,1)

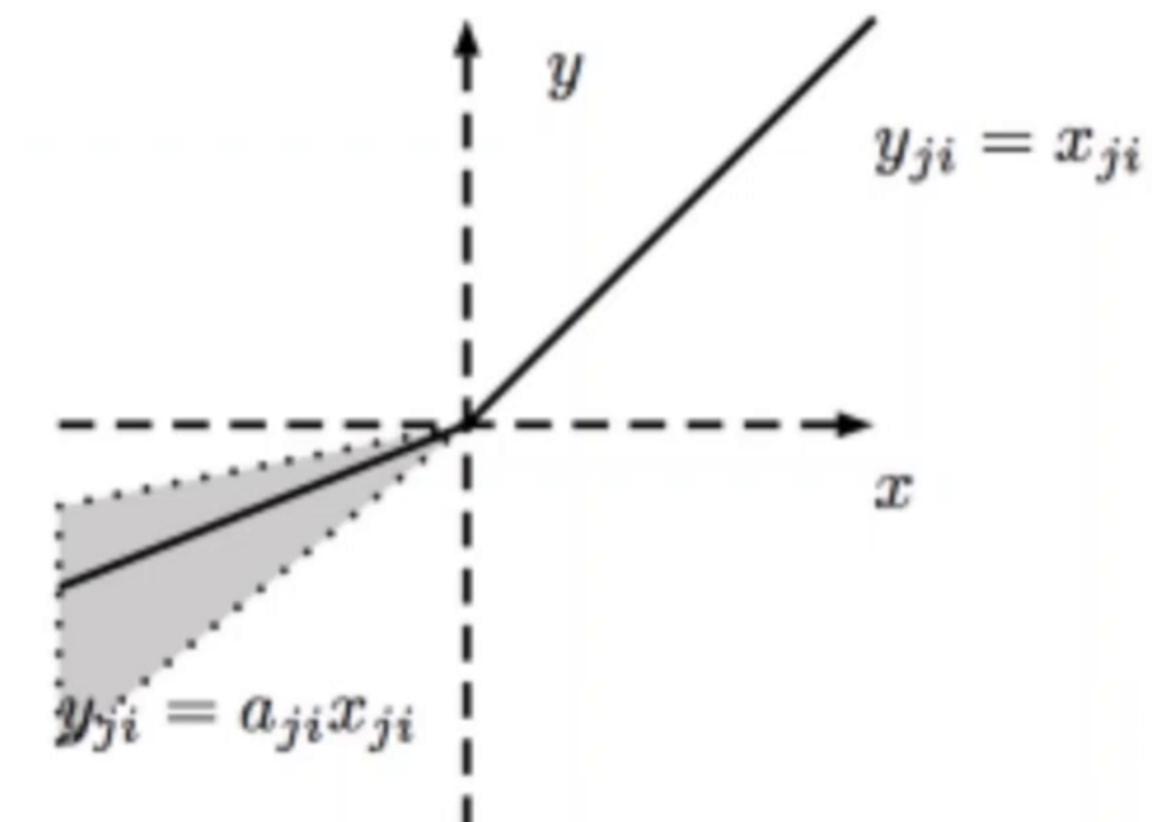
NDSB kaggle competition:

(3,8)

| u

$$y_{ji} = \frac{x_{ji}}{\frac{l+u}{2}}$$

Test time



Randomized Leaky ReLU

Randomized Leaky ReLU

'a' - Random sample from
Uniform Distribution [0,1)

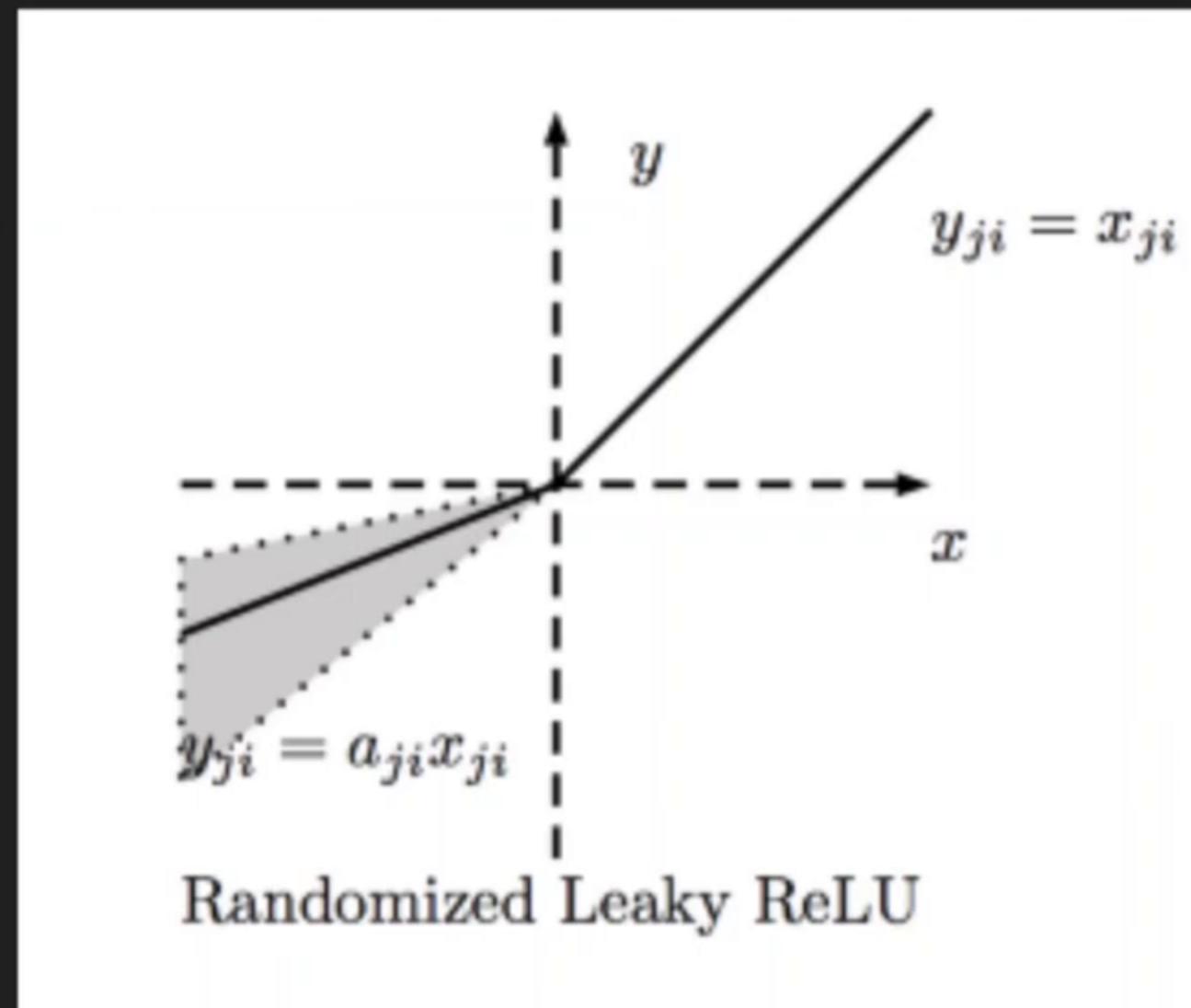
NDSB kaggle competition:

(3,8)

| u

$$y_{ji} = \frac{x_{ji}}{\frac{l+u}{2}}$$

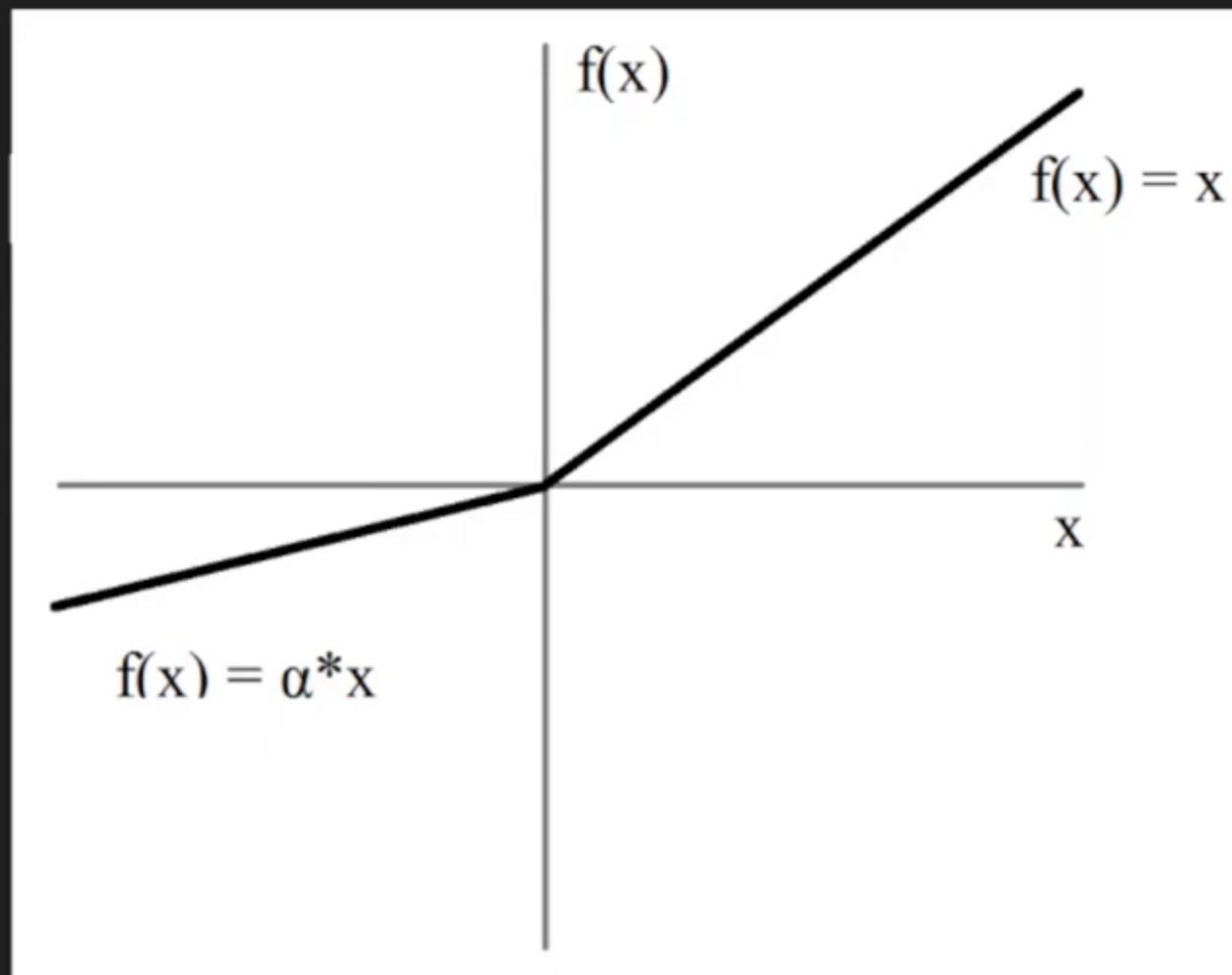
Test time



Parametric ReLU

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ ax & \text{otherwise} \end{cases}$$

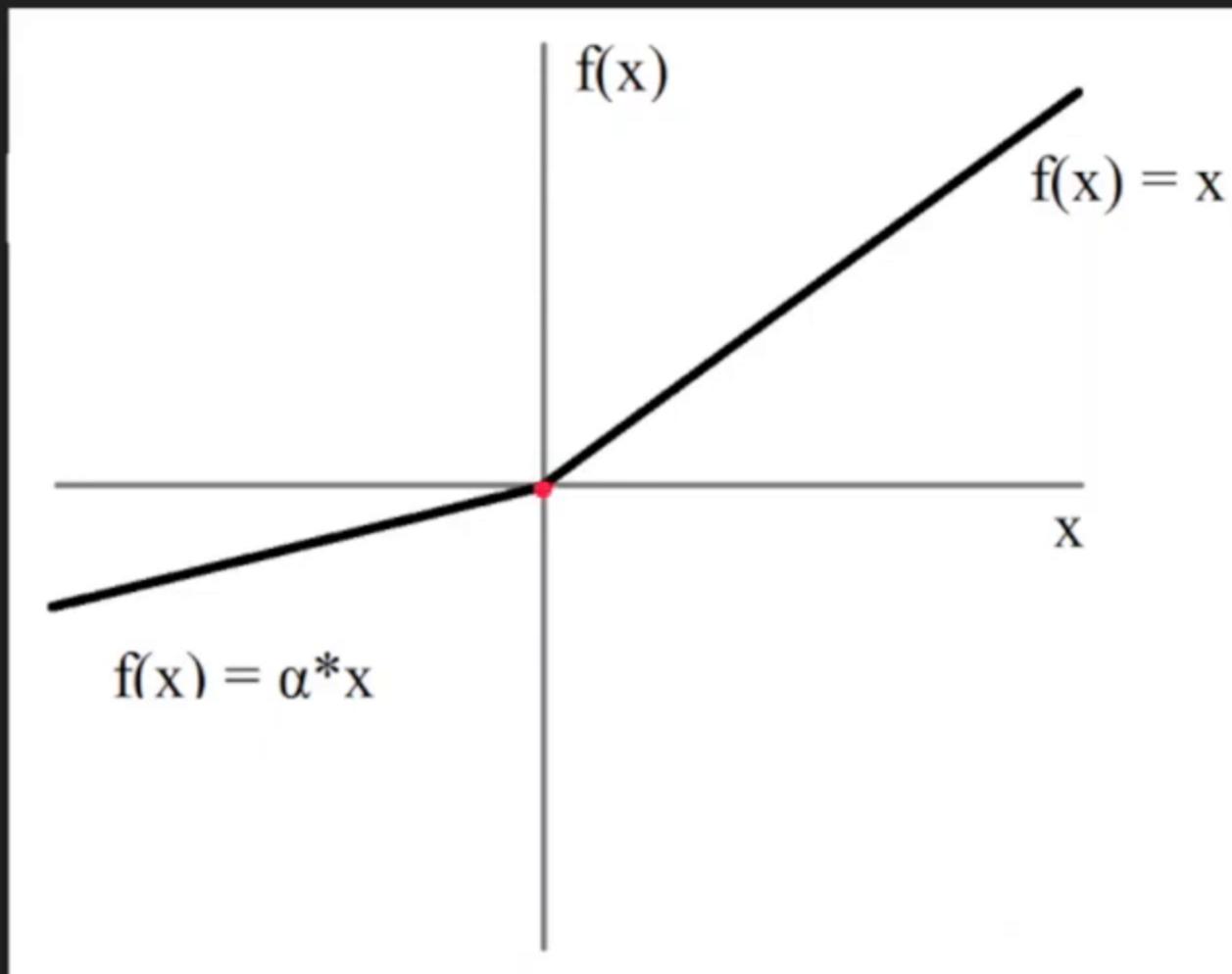
a - learned parameter



Parametric ReLU

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ ax & \text{otherwise} \end{cases}$$

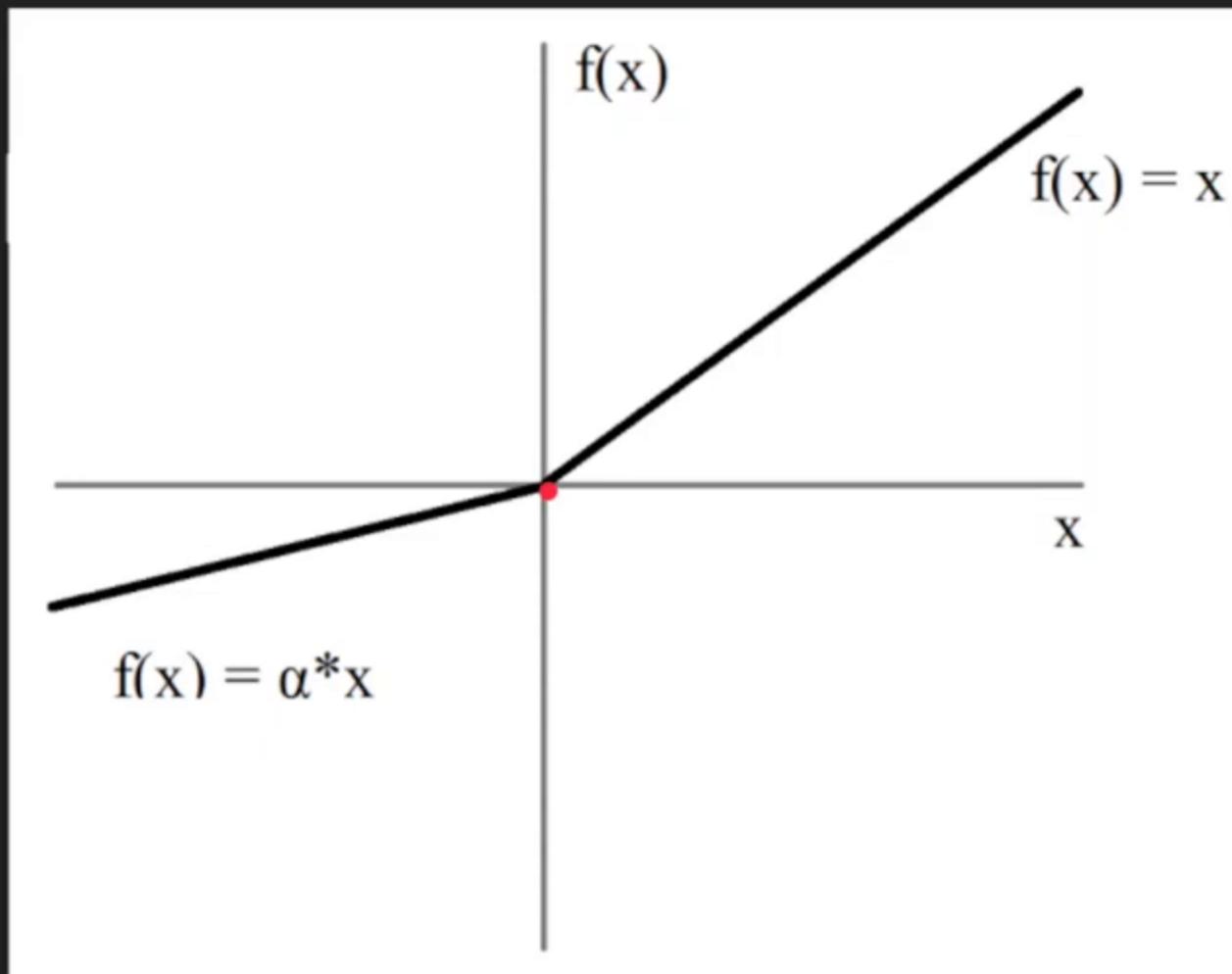
a - learned parameter



Parametric ReLU

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ ax & \text{otherwise} \end{cases}$$

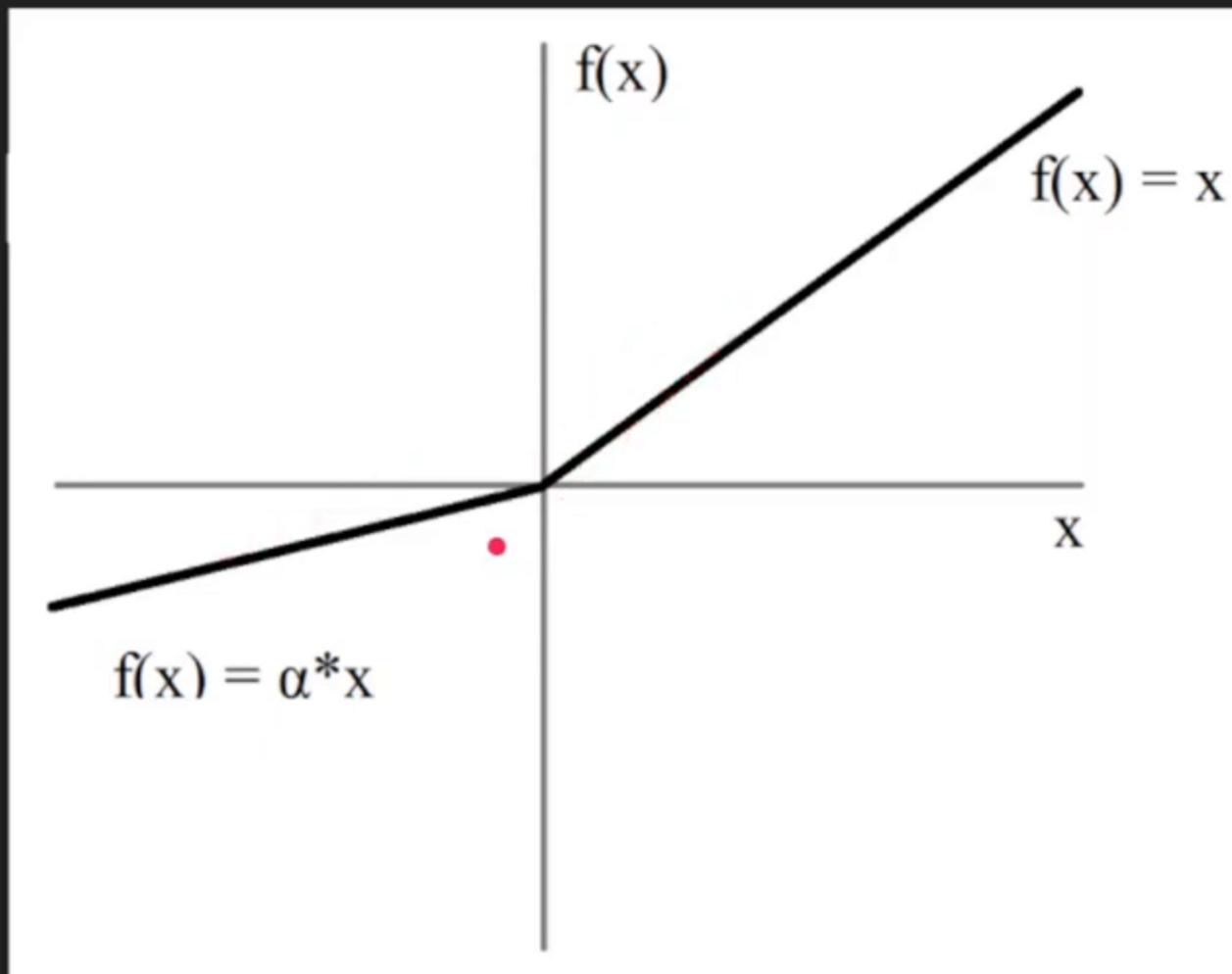
a - learned parameter



Parametric ReLU

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ ax & \text{otherwise} \end{cases}$$

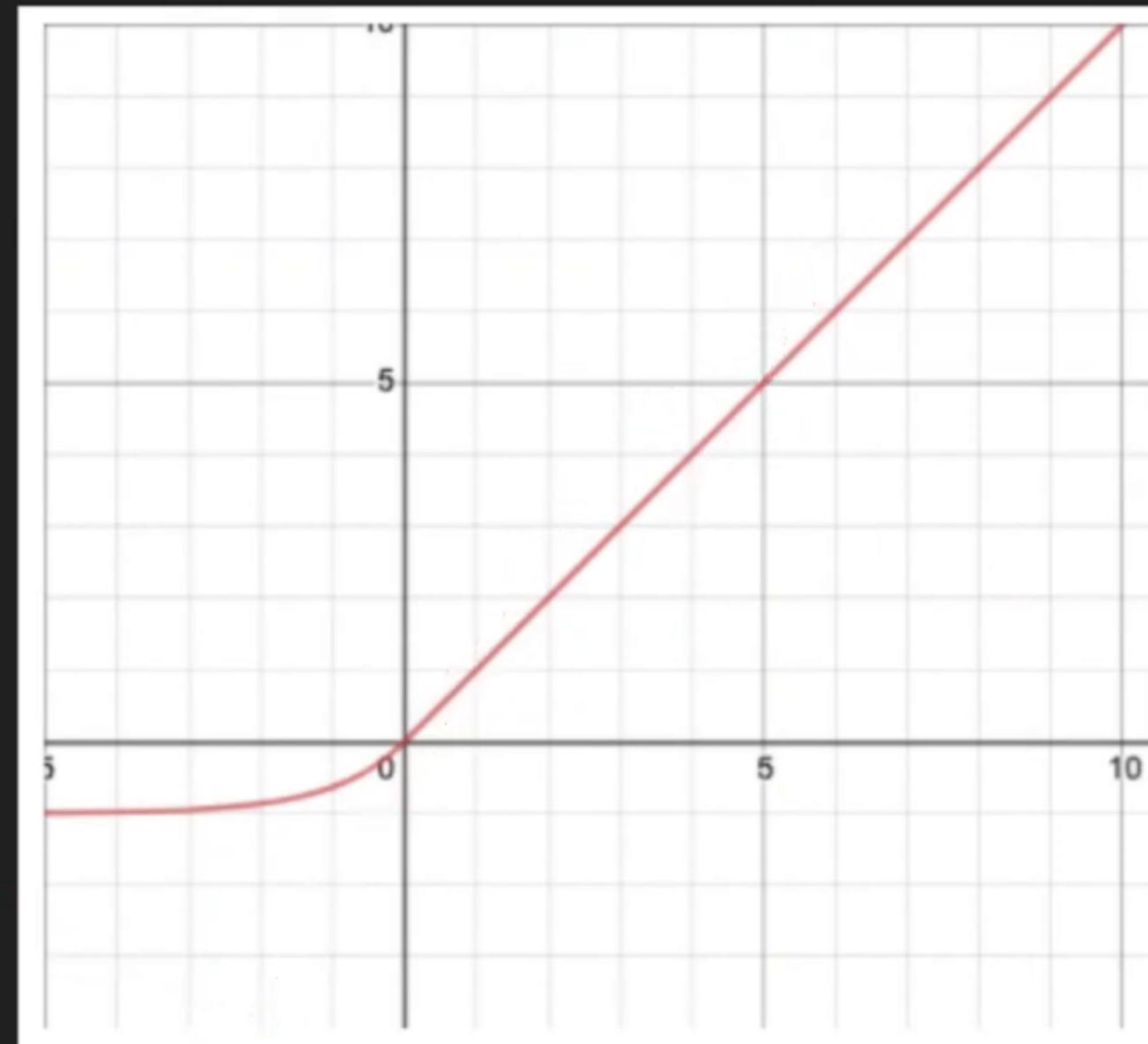
a - learned parameter



Exponential Linear Unit

$$y = \text{ELU}(x) = \exp(x) - 1 ; \text{ if } x < 0$$

$$y = \text{ELU}(x) = x ; \text{ if } x \geq 0$$

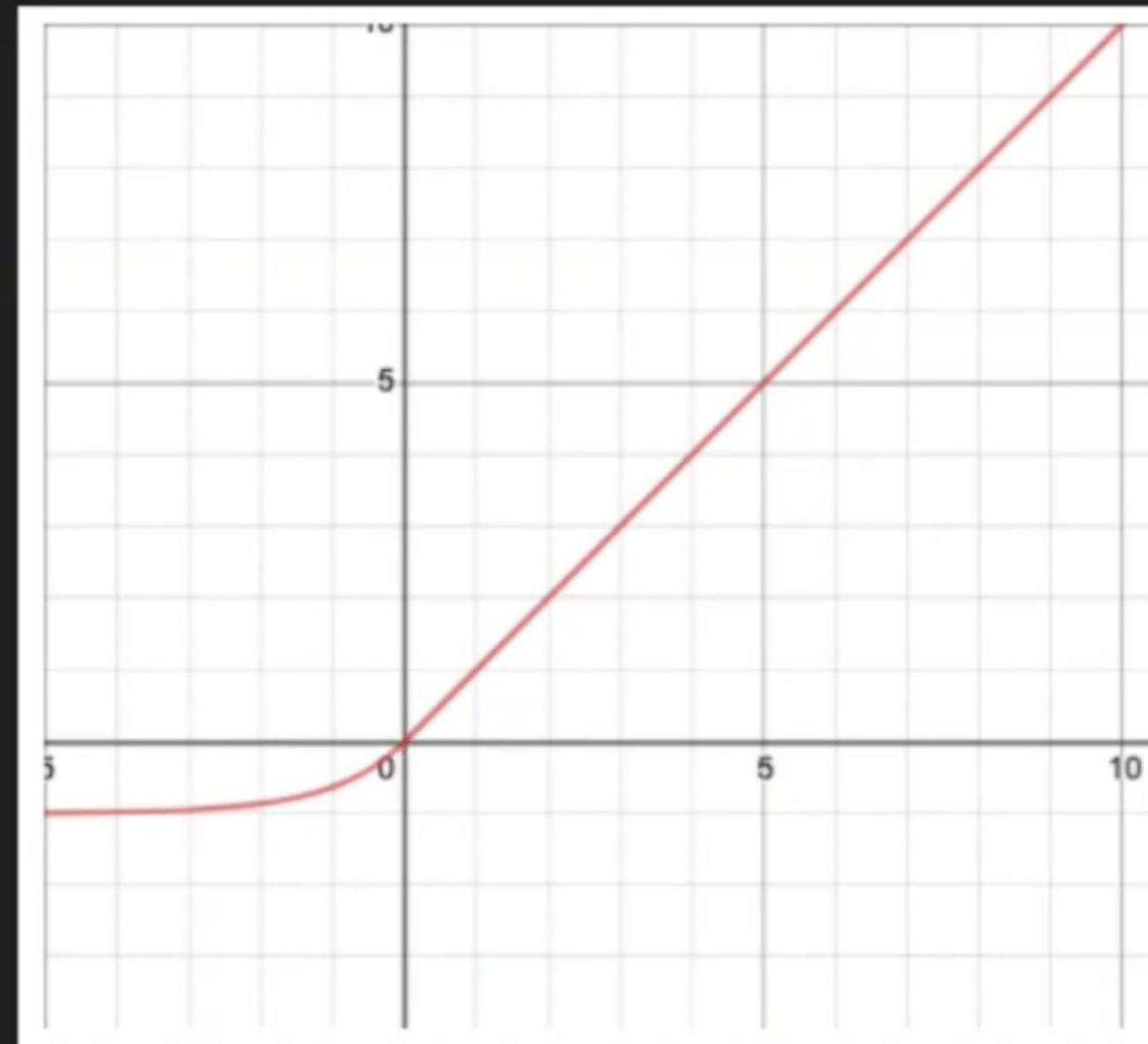


Exponential Linear Unit

$$y = \text{ELU}(x) = \exp(x) - 1 ; \text{ if } x < 0$$

$$\bullet$$

$$y = \text{ELU}(x) = x ; \text{ if } x \geq 0$$



Scaled ELU

$$\text{selu}_\bullet(x) = \lambda \begin{cases} x & \text{if } x > 0 \\ \alpha e^x - \alpha & \text{if } x \leq 0 \end{cases}$$

α and λ are fixed

Scaled ELU

$$\text{selu}(x) = \lambda \begin{cases} x & \text{if } x > 0 \\ \alpha e^x - \alpha & \text{if } x \leq 0 \end{cases}$$

α and λ are fixed

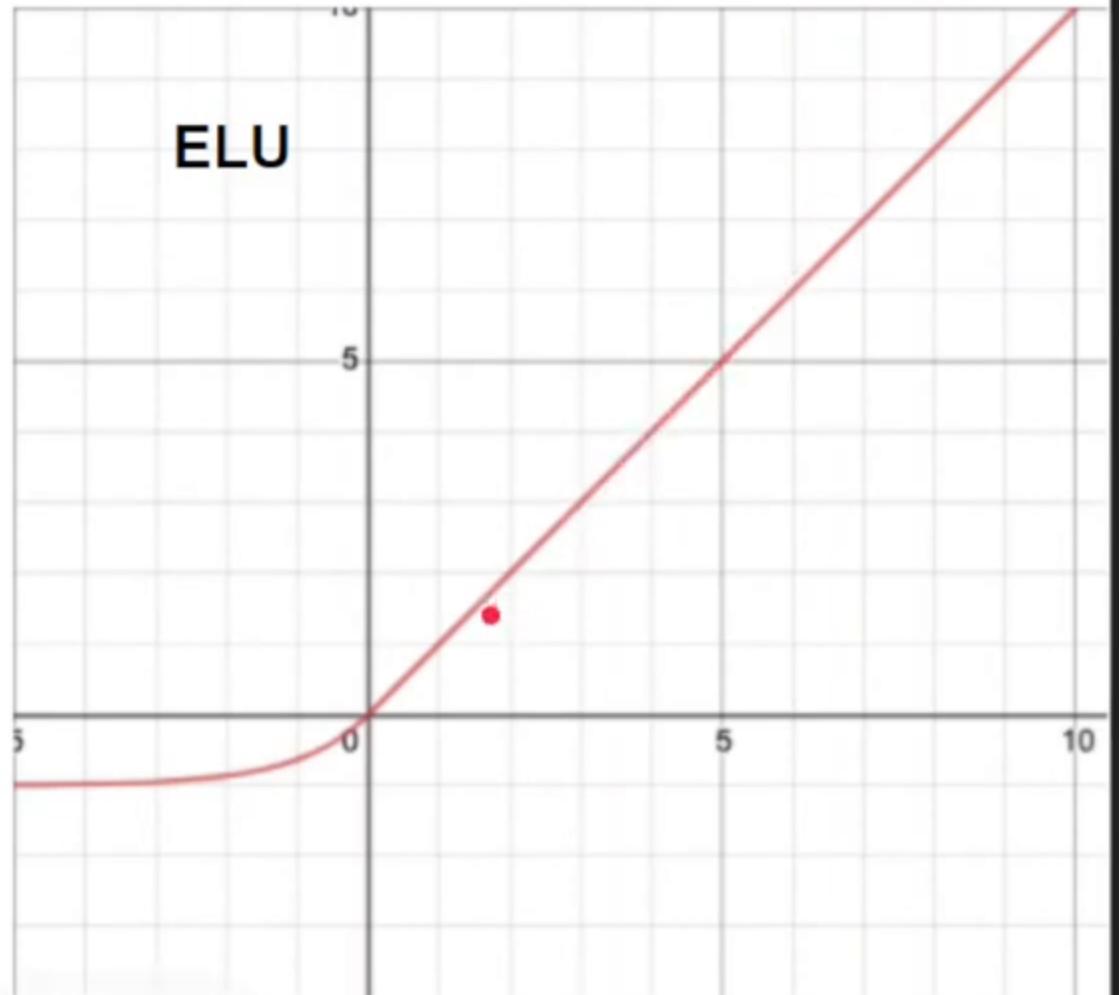


$$y = \text{ELU}(x) = \exp(x) - 1 ; \text{ if } x < 0$$

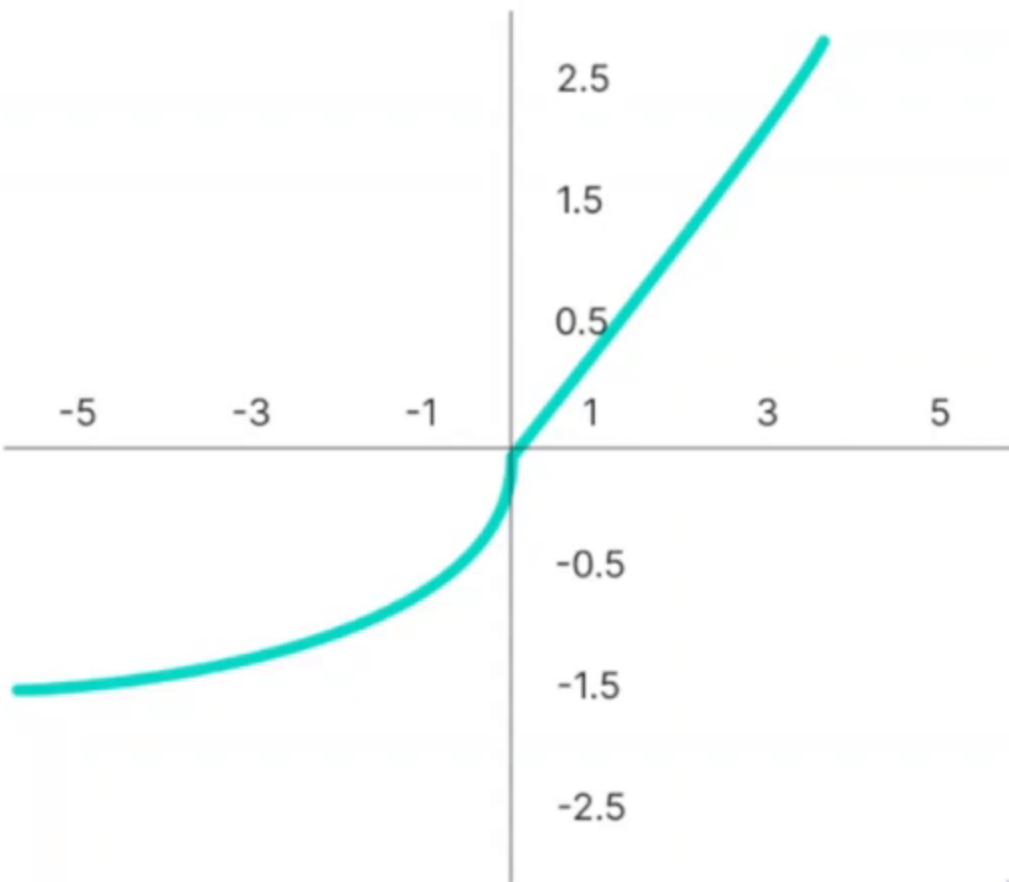
$$y = \text{ELU}(x) = x ; \text{ if } x \geq 0$$

$$\text{selu}(x) = \lambda \begin{cases} x & \text{if } x > 0 \\ \alpha e^x - \alpha & \text{if } x \leq 0 \end{cases}$$

ELU



SELU

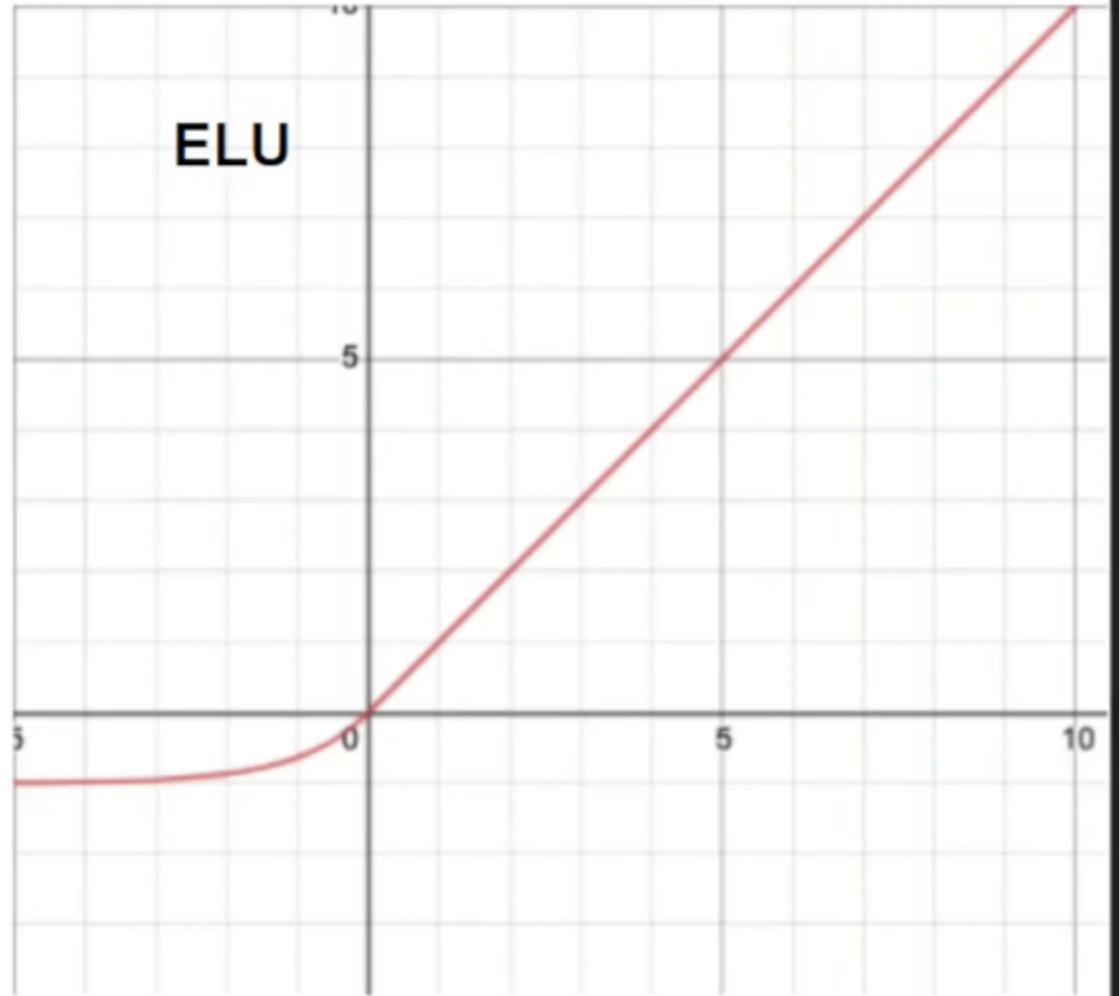


$$y = \text{ELU}(x) = \exp(x) - 1 ; \text{ if } x < 0$$

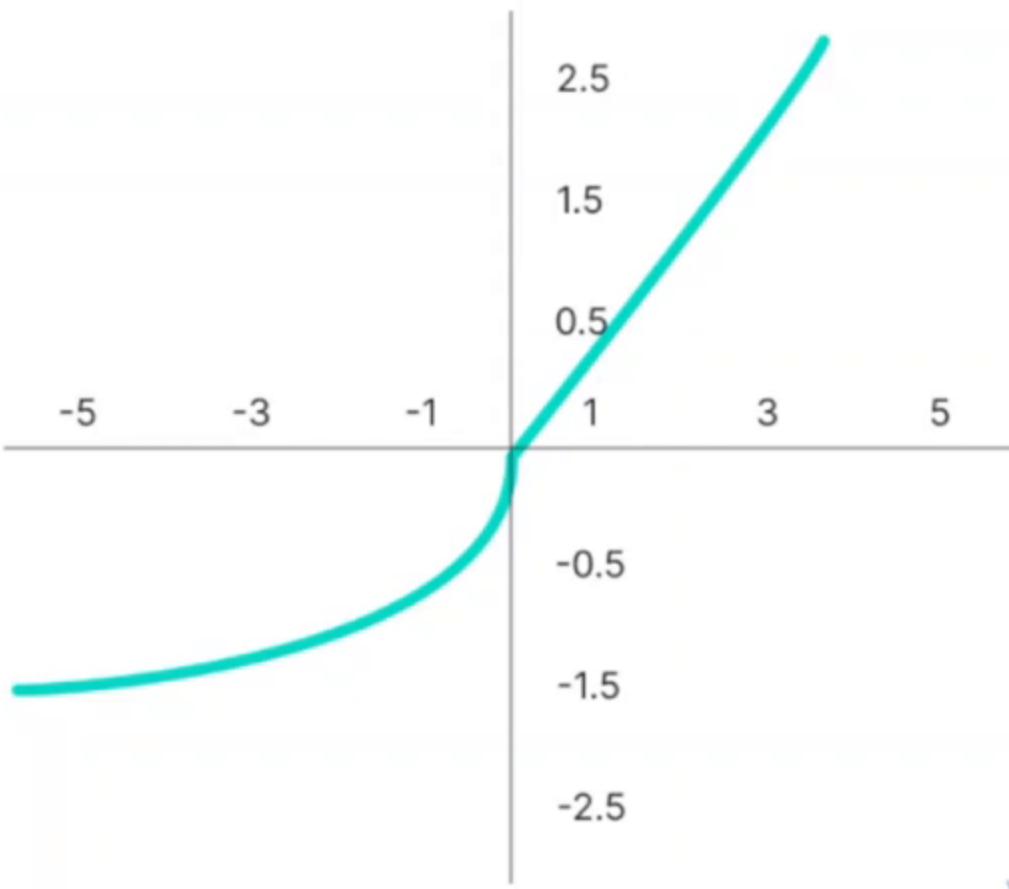
$$y = \text{ELU}(x) = x ; \text{ if } x \geq 0$$

$$\text{selu}(x) = \lambda \cdot \begin{cases} x & \text{if } x > 0 \\ \alpha e^x - \alpha & \text{if } x \leq 0 \end{cases}$$

ELU



SELU

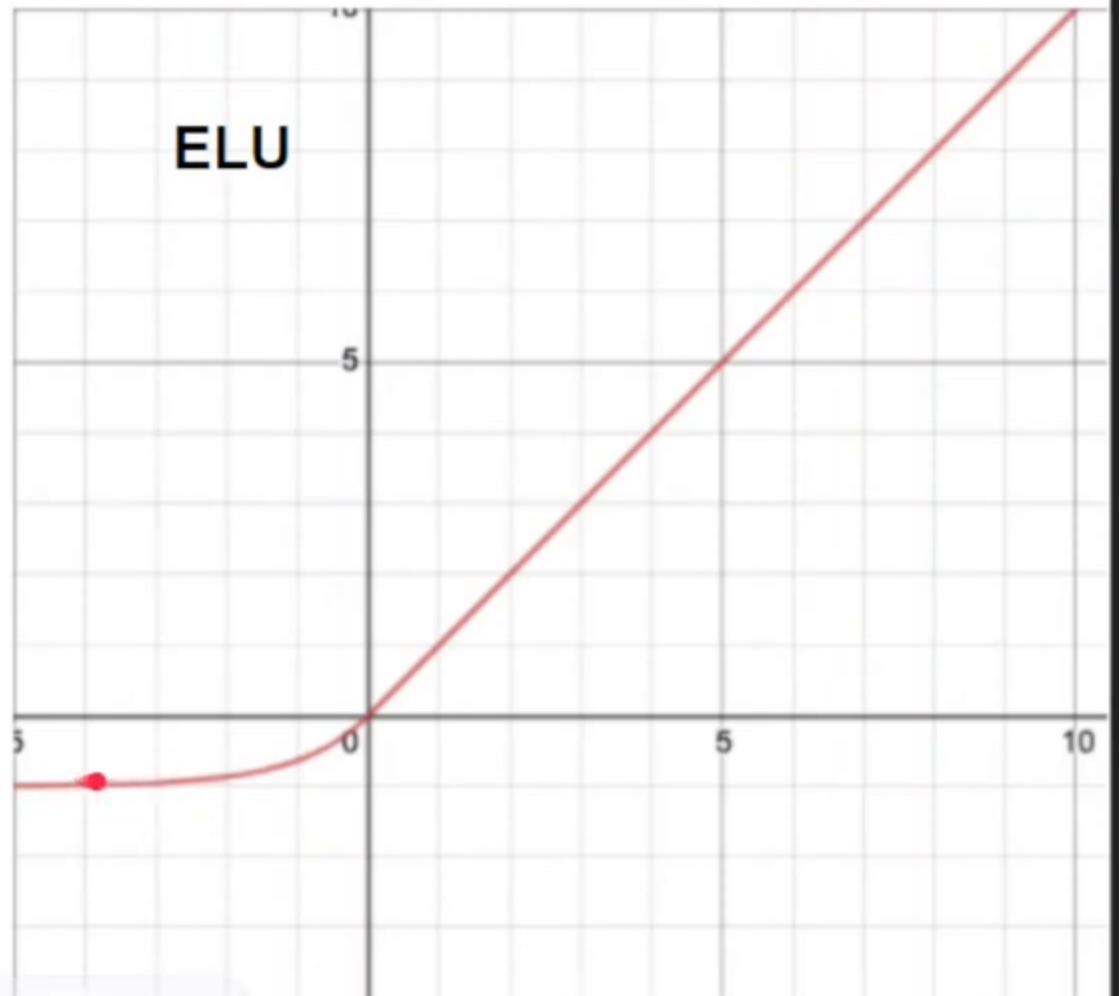


$$y = \text{ELU}(x) = \exp(x) - 1 ; \text{ if } x < 0$$

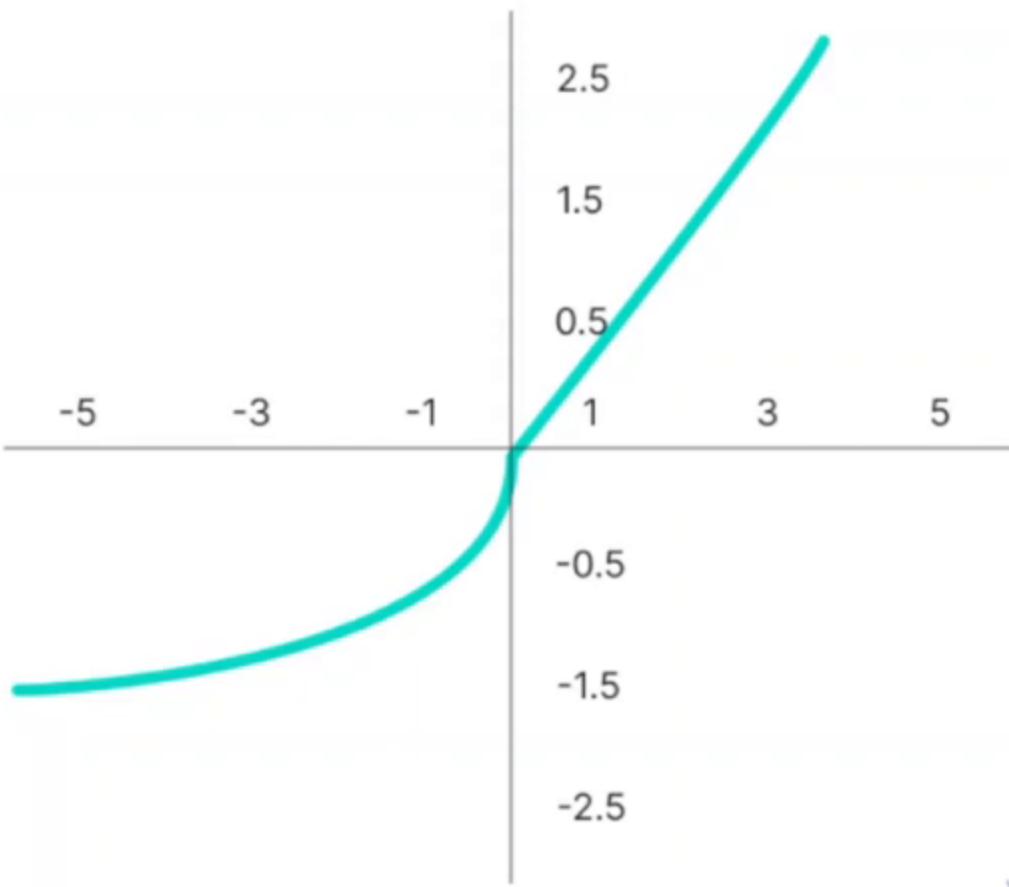
$$y = \text{ELU}(x) = x ; \text{ if } x \geq 0$$

$$\text{selu}(x) = \lambda \begin{cases} x & \text{if } x > 0 \\ \alpha e^x - \alpha & \text{if } x \leq 0 \end{cases}$$

ELU



SELU



Summary

ReLU

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$$

Leaky ReLU

$$f(x) = \begin{cases} 0.01x & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$$

Parametric ReLU

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ ax & \text{otherwise} \end{cases}$$

ELU

$$y = \text{ELU}(x) = \exp(x) - 1 ; \text{ if } x < 0$$
$$y = \text{ELU}(x) = x ; \text{ if } x \geq 0$$

Scaled ELU

$$\text{selu}(x) = \lambda \begin{cases} x & \text{if } x > 0 \\ \alpha e^x - \alpha & \text{if } x \leq 0 \end{cases}$$

