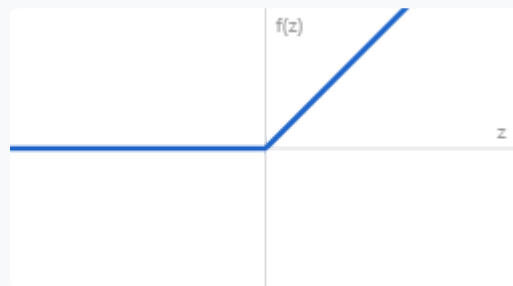


ReLU and Its Variants

ReLU



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

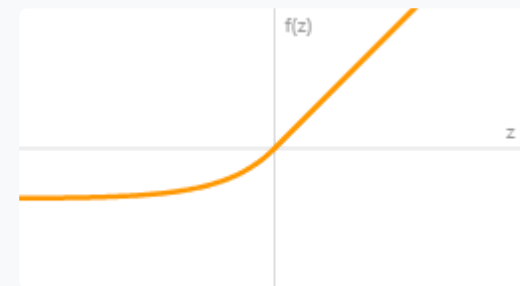
Leaky ReLU



$$f(z) = \max(\alpha z, z)$$

$$\alpha = 0.01$$

ELU



$$f(z) = z \text{ if } z > 0$$

$$\alpha(e^z - 1) \text{ if } z \leq 0$$

ReLU Advantages

- ✓ Simple computation
- ✓ No vanishing gradient
- ✓ Sparse activation
- ✗ Dead neurons ($z < 0$)

Leaky ReLU

- ✓ Fixes dying ReLU
- ✓ Small negative slope
- ✓ All neurons active
- ✓ Fast convergence

ELU Advantages

- ✓ Smooth everywhere
- ✓ Negative values push mean to zero
- ✓ Robust to noise
- ✗ Slower (exponential)

Feature	ReLU	Leaky ReLU	ELU
Computation Speed	Fast	Fast	Moderate
Dying Neurons	Yes	No	No

Feature	ReLU	Leaky ReLU	ELU
Zero-Centered	No	No	Nearly
Most Common Use	Default	Alternative	Special Cases