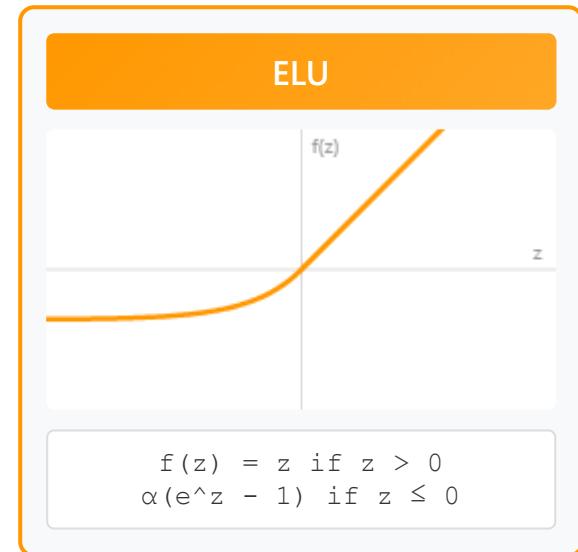
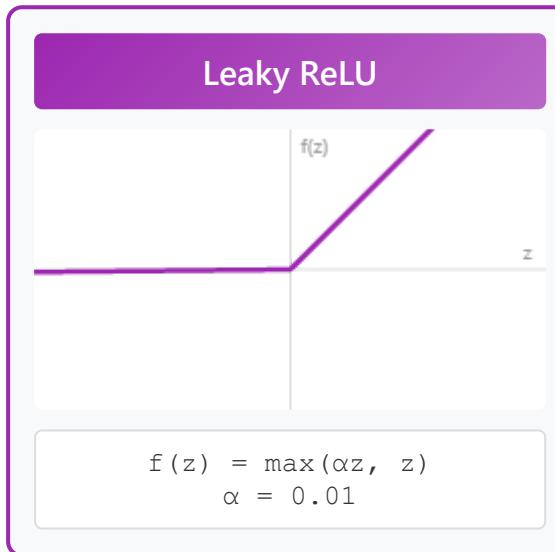
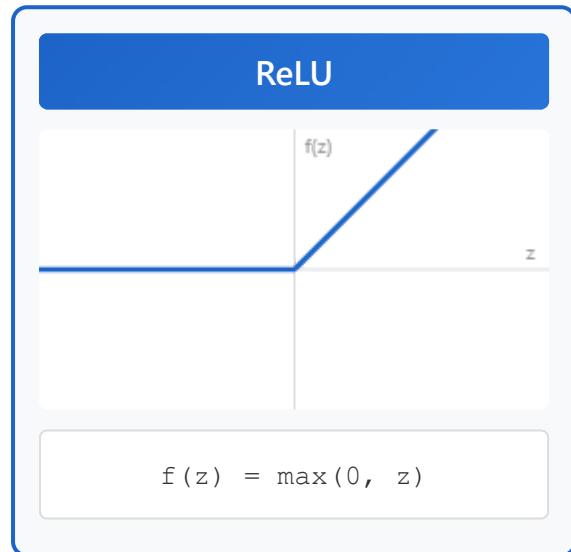


## ReLU and Its Variants



#### 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