Function	Output Range	Common Use	Pros	Cons
Sigmoid	(0, 1)	Binary output	Probabilistic output	Vanishing gradient
Tanh	(-1, 1)	Hidden layers (early NN)	Zero-centered	Still vanishes
ReLU	(0, ∞)	Hidden layers	Fast & sparse	Dying neurons
Leaky ReLU	(-∞, ∞)	Hidden layers	Solves dying ReLU	Adds extra parameter
PReLU	(-∞, ∞)	Hidden layers	Learns negative slope	Slightly more compute
ELU	(-α, ∞)	Hidden layers	Smooth learning	Slower than ReLU
Softmax	(0, 1)	Output layer (multi-class)	Probabilities	Saturates for large inputs