Activation Functions in Deeper Playground

1. ReLU (Rectified Linear Unit)

Formula: f(x) = max(0, x)

- Outputs the input directly if it's positive, otherwise zero.

- Very fast to compute and promotes sparse activations.

Use When:

- Deep feedforward or CNN models.

- Want fast convergence and reduced vanishing gradient.

Caution:

- Can suffer from "dying ReLU" problem.

2. Leaky ReLU

Formula: f(x) = x if x > 0 else alphax (alpha ~ 0.01)

- Allows a small negative slope to prevent dying neurons.

Use When:

- ReLU leads to inactive neurons.

- Need better gradient flow.

3. ELU (Exponential Linear Unit)

Formula: f(x) = x if x > 0 else alpha(e^x - 1)

- Smooth negative part, better zero-centered outputs.

Use When:

- Want faster learning and generalization.

Caution:

- Slightly slower due to exponentials.

4. Sigmoid

Formula: f(x) = 1 / (1 + e^(-x))

- Maps input to (0, 1). Great for binary output.

Use When:

- Binary classification output.

Caution:

- Vanishing gradients and not zero-centered.

5. Tanh (Hyperbolic Tangent)

Formula: f(x) = (e^x - e^(-x)) / (e^x + e^(-x))

- Maps input to (-1, 1). Better centered than sigmoid.

Use When:

- Faster learning in shallow nets or RNNs.

Caution:

- Still suffers from vanishing gradients.

6. Swish (SiLU - Sigmoid Linear Unit)

Formula: f(x) = x * sigmoid(x)

- Smooth, non-monotonic function. Better than ReLU in deep nets.

Use When:

- State-of-the-art deep networks.

Caution:

- Slightly more computationally expensive.

7. SoftPlus

Formula: f(x) = ln(1 + e^x)

- Smooth approximation of ReLU, always differentiable.

Use When:

- Probabilistic models (e.g., VAEs).

Caution:

- Less sparse and more expensive than ReLU.

Summary Table

Activation	Output Range	Zero-Centered?	Smooth?	Best Used In
ReLU	[0, inf)	No	No	CNNs, MLPs
Leaky ReLU	(-inf, inf)	Partial	No	Deep nets with dead ReLUs
ELU	(-alpha, inf)	Yes	Yes	BatchNorm + deep nets
Sigmoid	(0, 1)	No	Yes	Output layer (binary)
Tanh	(-1, 1)	Yes	Yes	RNNs, shallow nets
Swish	(-0.28, inf)	Approx.	Yes	Modern deep nets
SoftPlus	(0, inf)	No	Yes	VAEs, regression heads