Role of Activation Functions

✗ Without Activation

$$y = W_2(W_1x + b_1) + b_2$$

$$y = W'x + b'$$

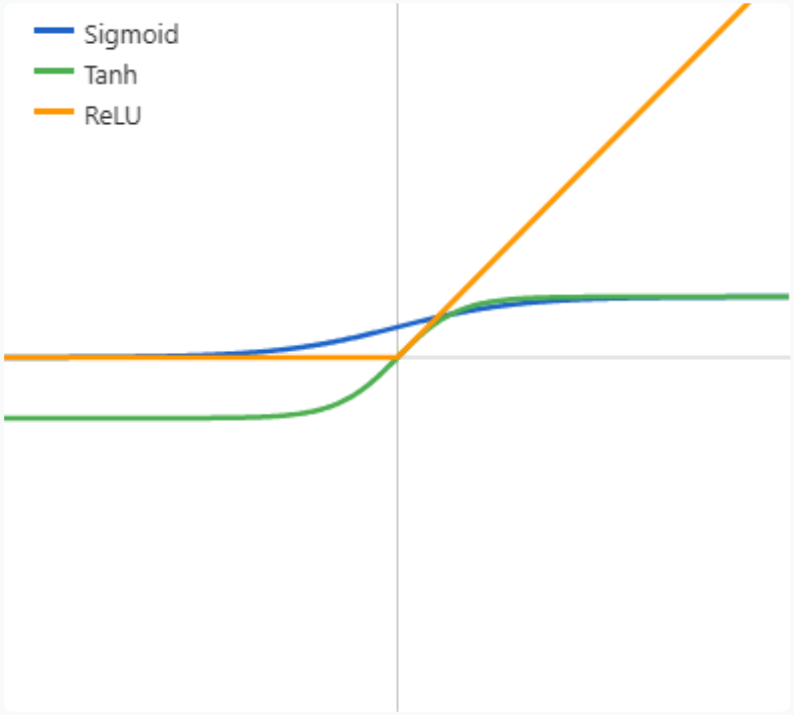
Multiple layers collapse into a single linear transformation. No advantage over logistic regression!

✓ With Activation

$$y = \sigma(W_2\sigma(W_1x + b_1) + b_2)$$

Non-linear transformations enable learning complex patterns and hierarchical representations.

Common Activation Functions



1 Non-linearity

Introduce non-linear transformations to learn complex patterns

2 Gradient Flow

Enable backpropagation by providing differentiable functions

3 Expressiveness

Increase model capacity to approximate complex functions