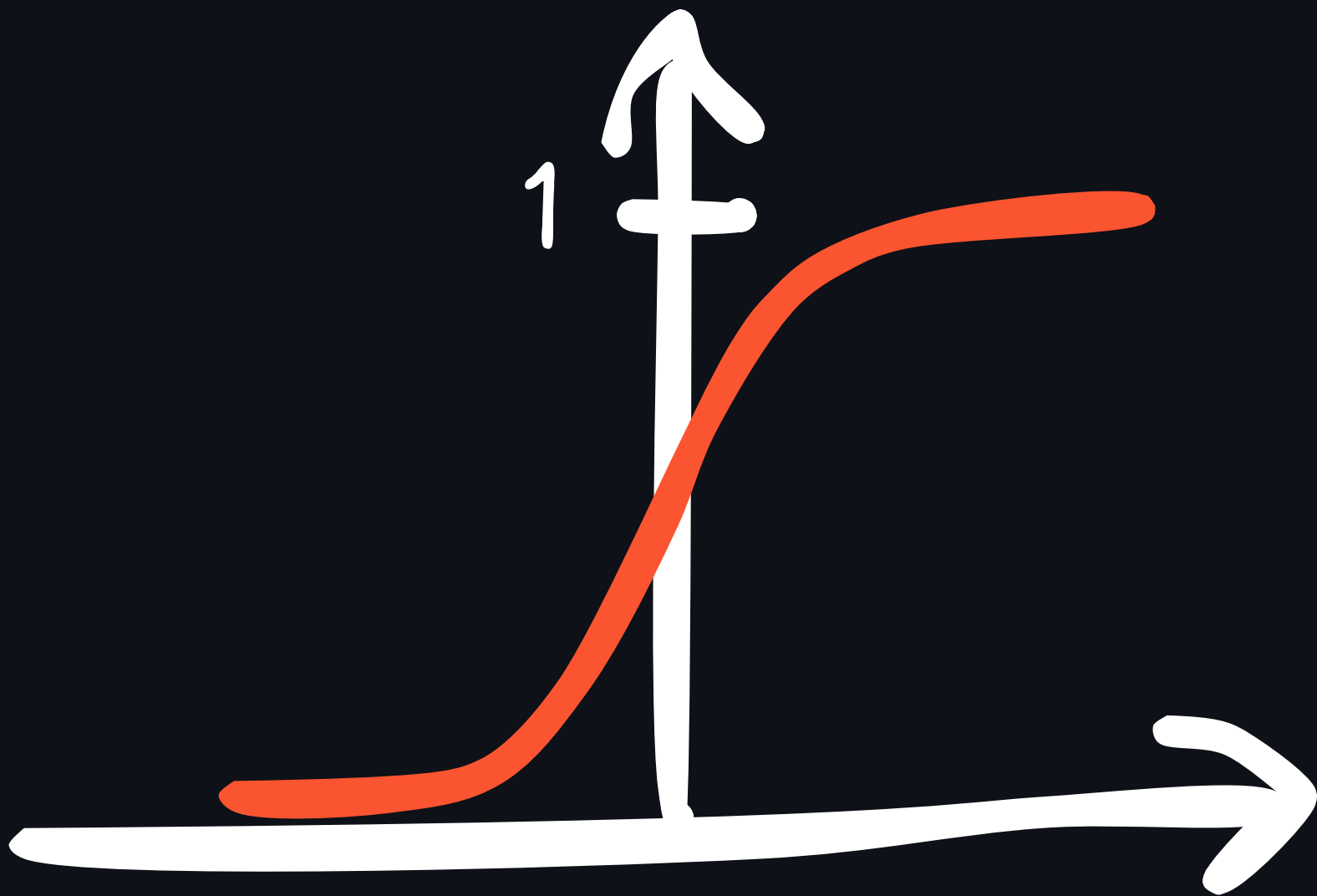


Activation Functions

The **Sigmoid** activation function.



$$a(z) = \frac{1}{1 + e^{-z}}$$

- ✓ Perfect for classical **binary classification** where the output is either 0 or 1.
- ✓ Output the probability of belonging to a certain class.

Activation Functions

The **Softmax** activation function.

Output
Layer

Probabilities

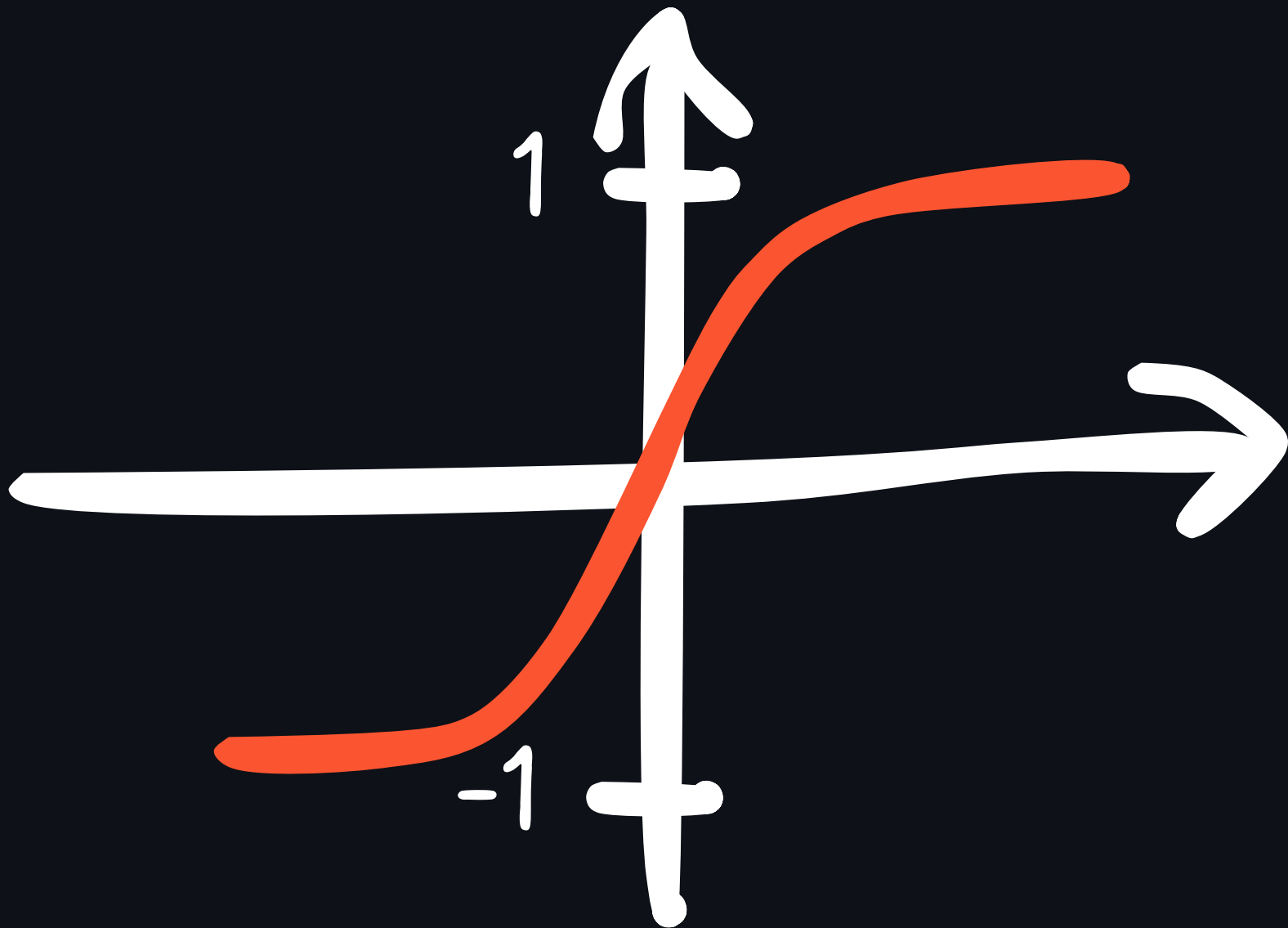
$$\begin{pmatrix} 2.1 \\ 3 \\ 4.8 \\ 3.2 \\ 0.1 \end{pmatrix} \longrightarrow a(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \longrightarrow \begin{pmatrix} 0.8 \\ 0.11 \\ 0.03 \\ 0.04 \\ 0.02 \end{pmatrix}$$

$$a(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \text{ for } j = 1, \dots, K \text{ and } \mathbf{z} = (z_1, \dots, z_K)$$

- ✓ Generalization of sigmoid for **multi-class classification** tasks like image classification.
- ✓ Output **probabilities for each class** from 1 to K.

Activation Functions

The TanH (Hyperbolic Tangent) activation function.

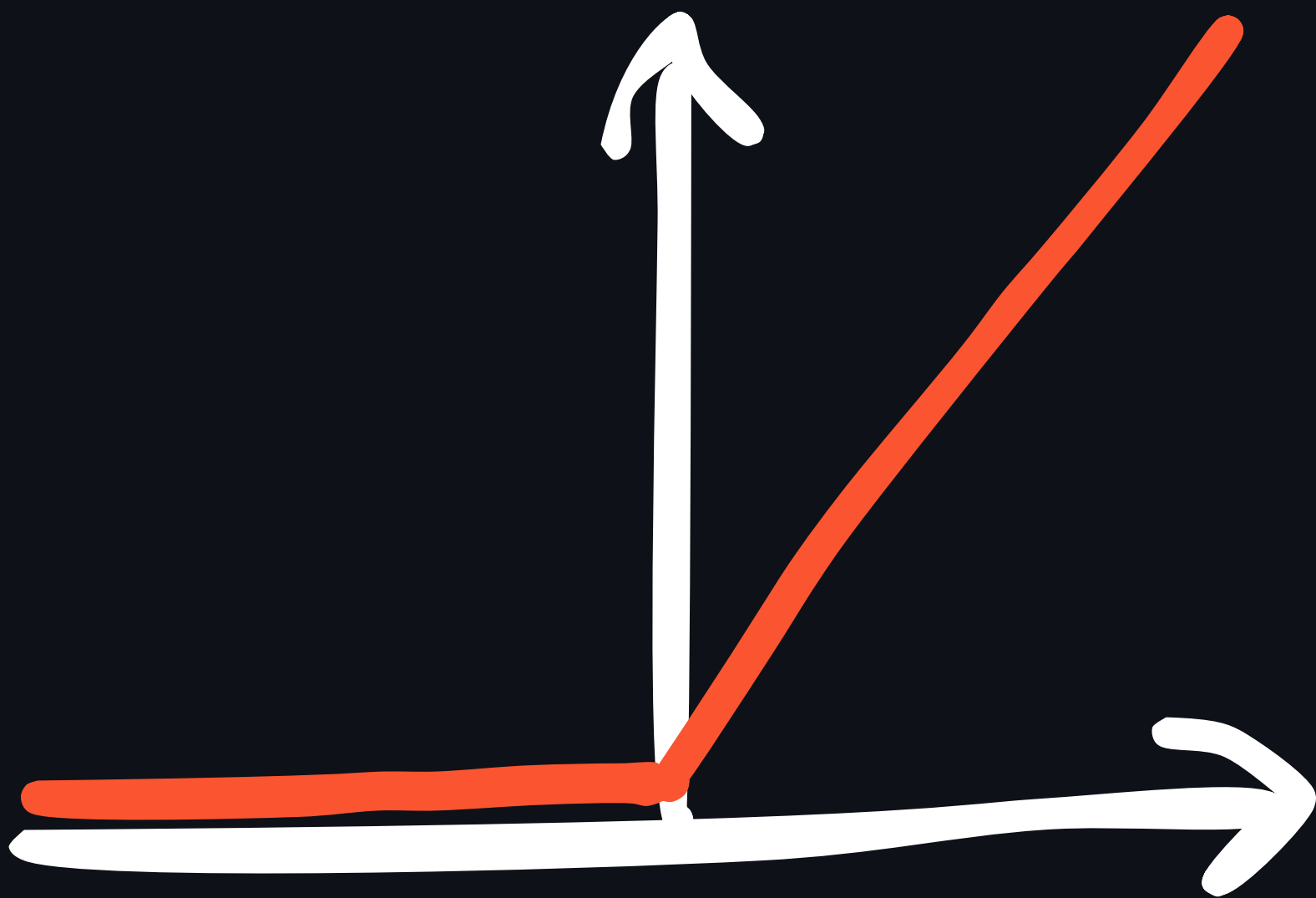


$$a(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

- ✓ Often used in hidden layers of a neural network. It centers the data, improving the learning for subsequent layers.

Activation Functions

The **ReLU** (Rectified Linear Unit) activation function.

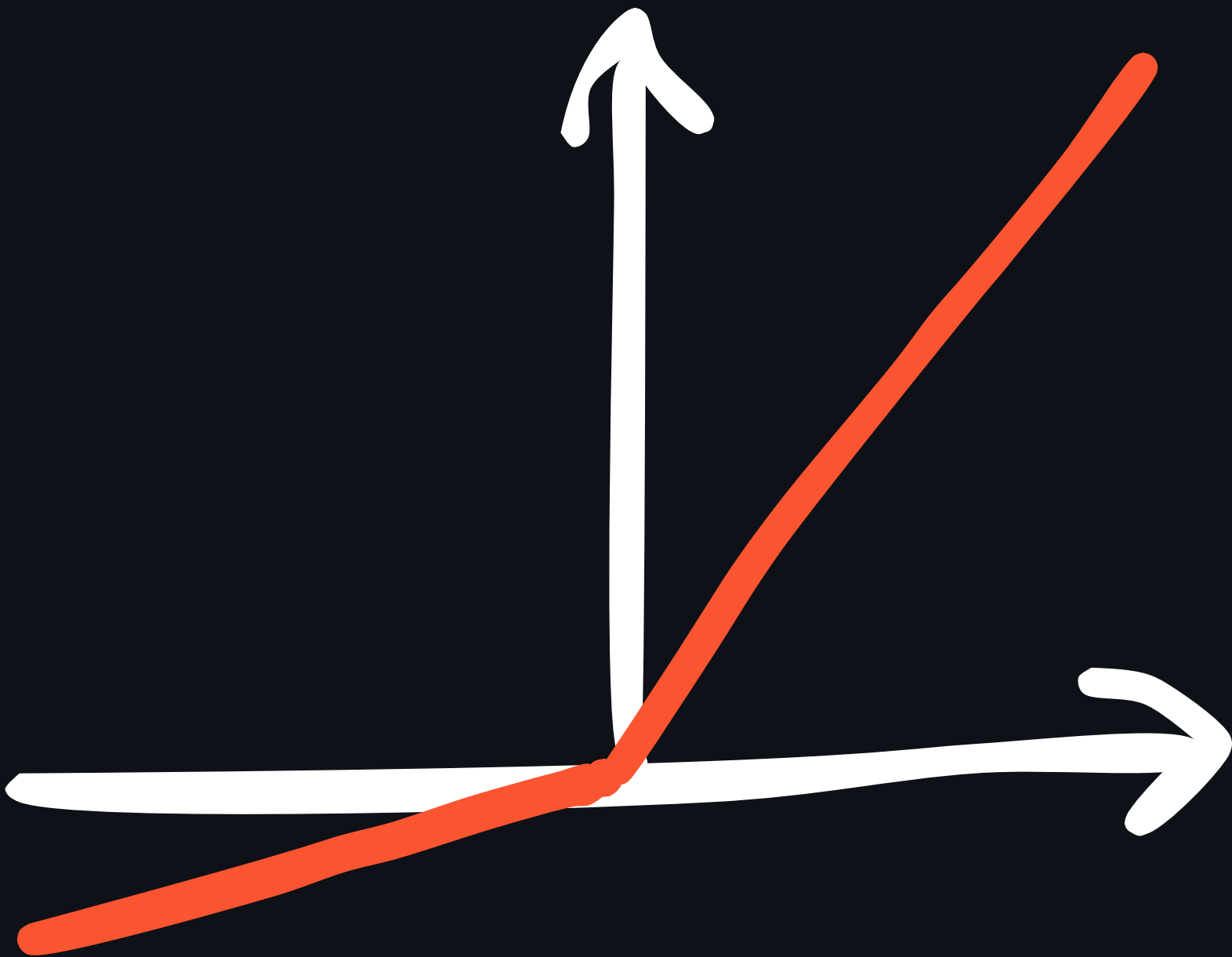


$$a(z) = \max(0, x)$$

- ✓ Widely used in CNNs for tasks like image recognition due to its computational efficiency.

Activation Functions

The **leaky ReLU** activation function.

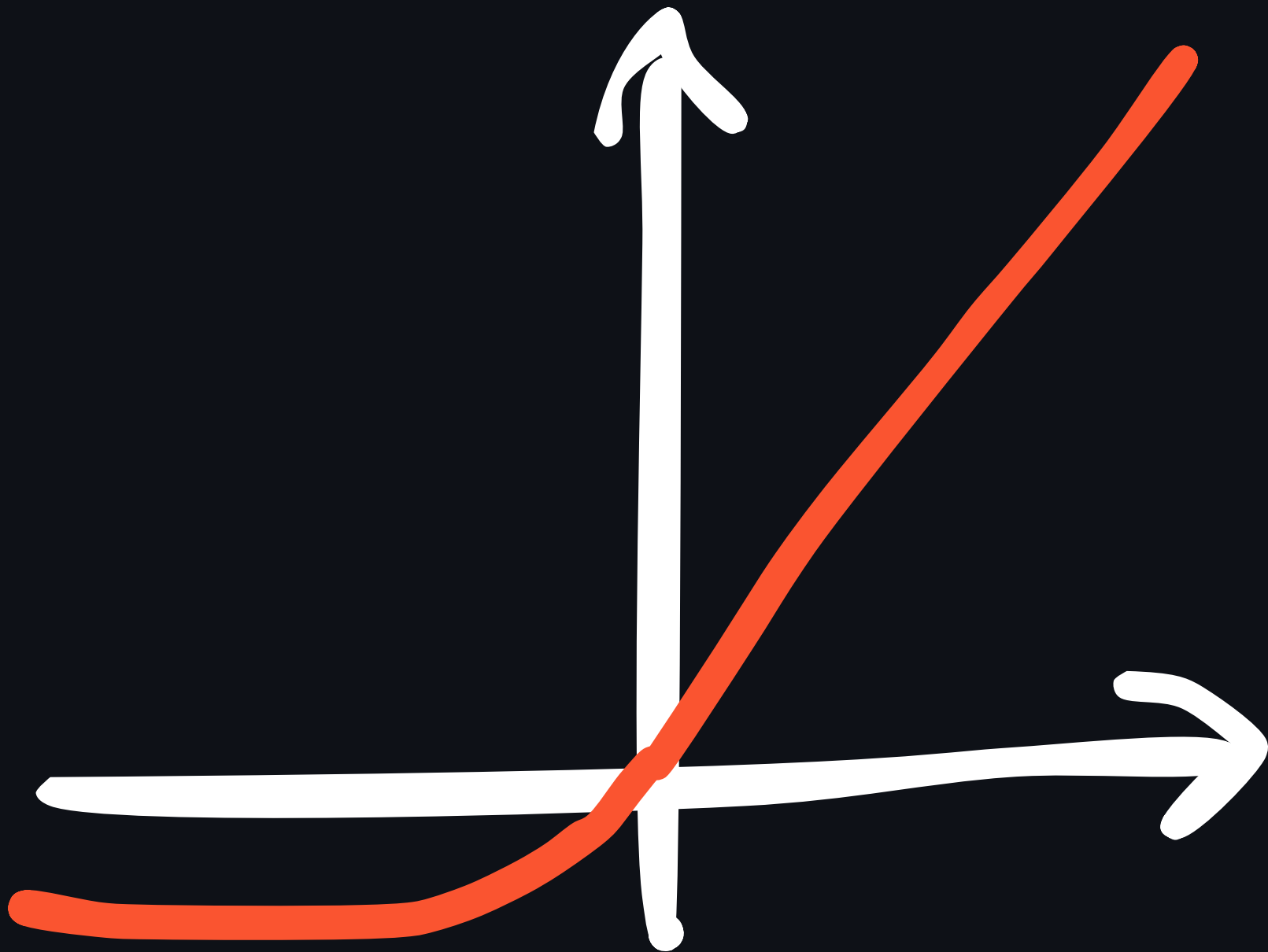


$$a(z) = \begin{cases} z & \text{if } z > 0, \\ \alpha z & \text{otherwise.} \end{cases}$$

- ✓ Helps prevent **dead neurons in ReLU**. Useful in deep learning models where avoiding the dying ReLU problem is critical.

Activation Functions

The ELU (Exponential Linear Unit) activation function.

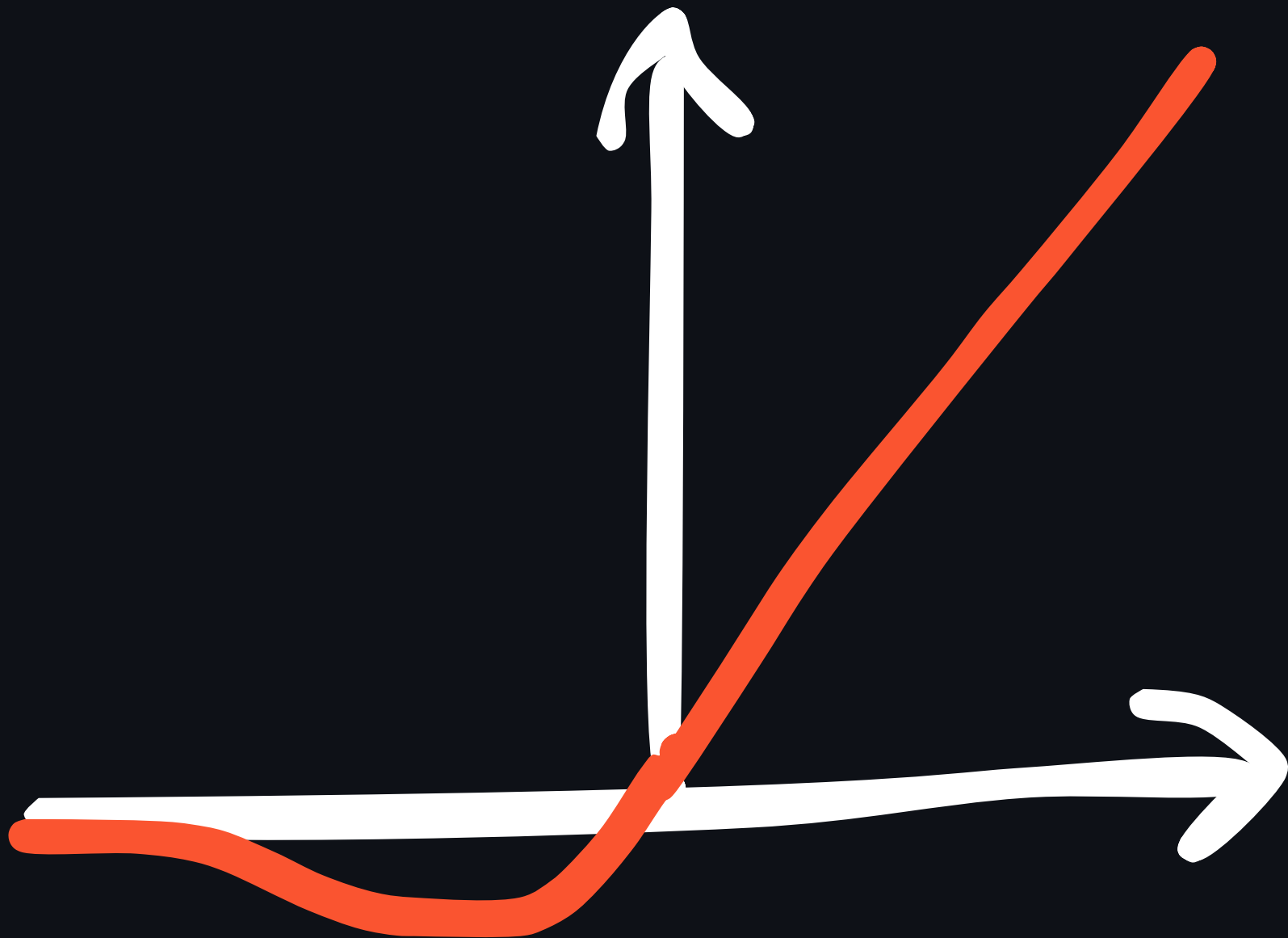


$$a(z) = \begin{cases} z & \text{if } z > 0, \\ \alpha(e^z - 1) & \text{if } z \leq 0. \end{cases}$$

- ✓ ELU improves accuracy for deep neural networks by allowing negative values for activations.

Activation Functions

The **GeLU** (Gaussian Error Linear Unit) activation function.



$$a(z) = zP(Z \leq z)$$

- ✓ Non-linear activation function that approximates the expectation of a Gaussian distribution.

Thanks for Reading

I hope this helps!