2-3 Activation Functions

Zhonglei Wang

WISE and SOE, XMU, 2025

Contents

Why activation?

1. Neural networks mainly involve two steps for each neuron
   - Linear transformation
   - Activation (nonlinear transformation)
2. If there is no activation, neural networks are simple linear models!
3. Activation empowers neural networks to learn complex structures

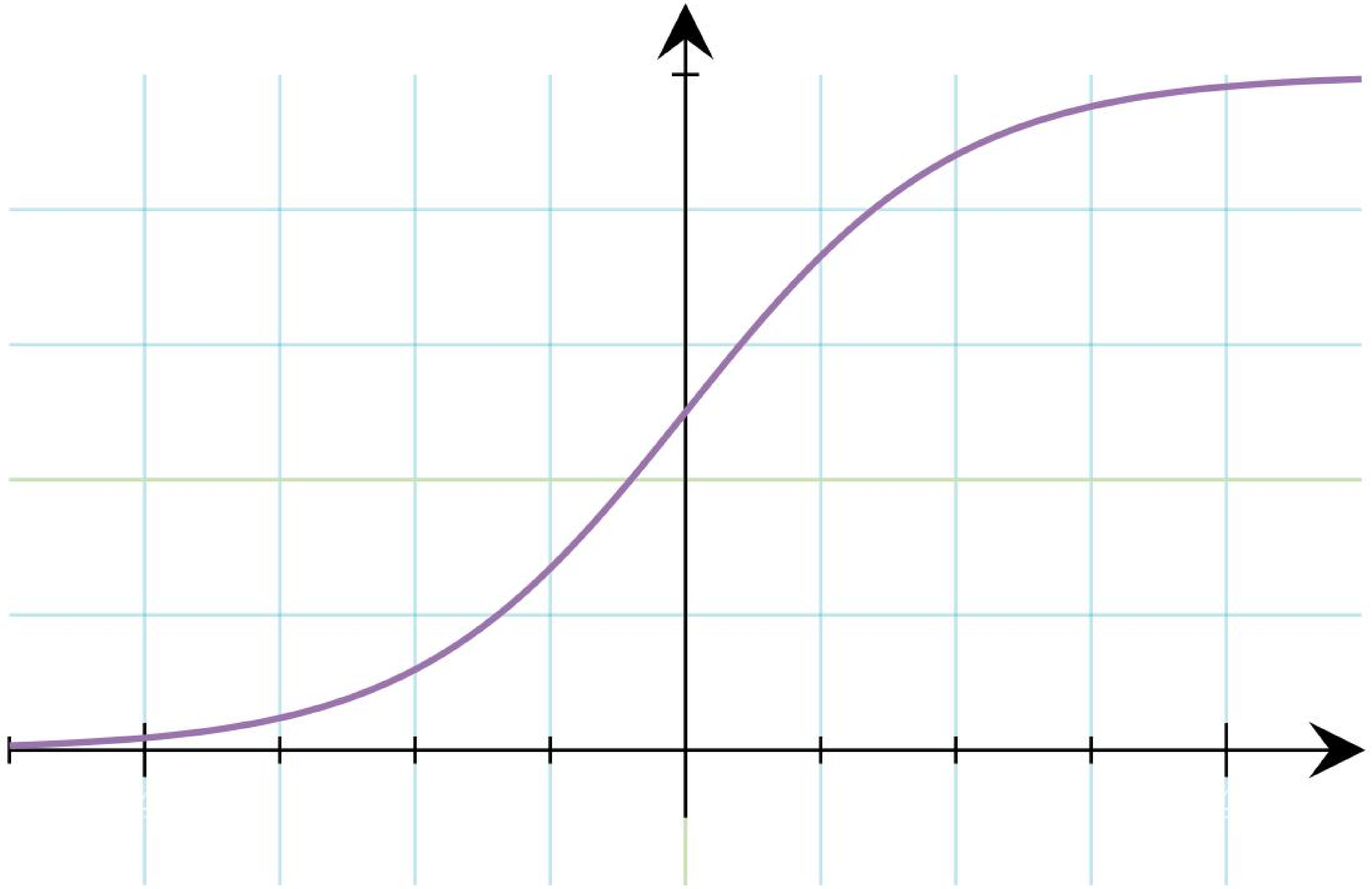
Sigmoid activation function

1. Form

$$\sigma(z) = \frac{1}{1 + \exp(-z)}$$

- Range: $(0, 1)$
- Derivative: $\sigma'(z) = \sigma(z)\{1 - \sigma(z)\}$

Sigmoid activation function



Sigmoid activation function

1. Advantages
   - Range is $(0, 1)$, and it is commonly used for binary classification (last layer).
   - Gradient vanishes (actually it is not good) if z is large, so it is robust to outliers
   - Easy to obtain its derivative for backpropagation.
2. Disadvantages
   - Gradient vanishing.
     - ▷ For a neural network with 5 layers, we may have $0.25^5 \approx 0.001$!
   - Output is not symmetric about 0.
   - Heavy computation for the derivative due to the exponential.

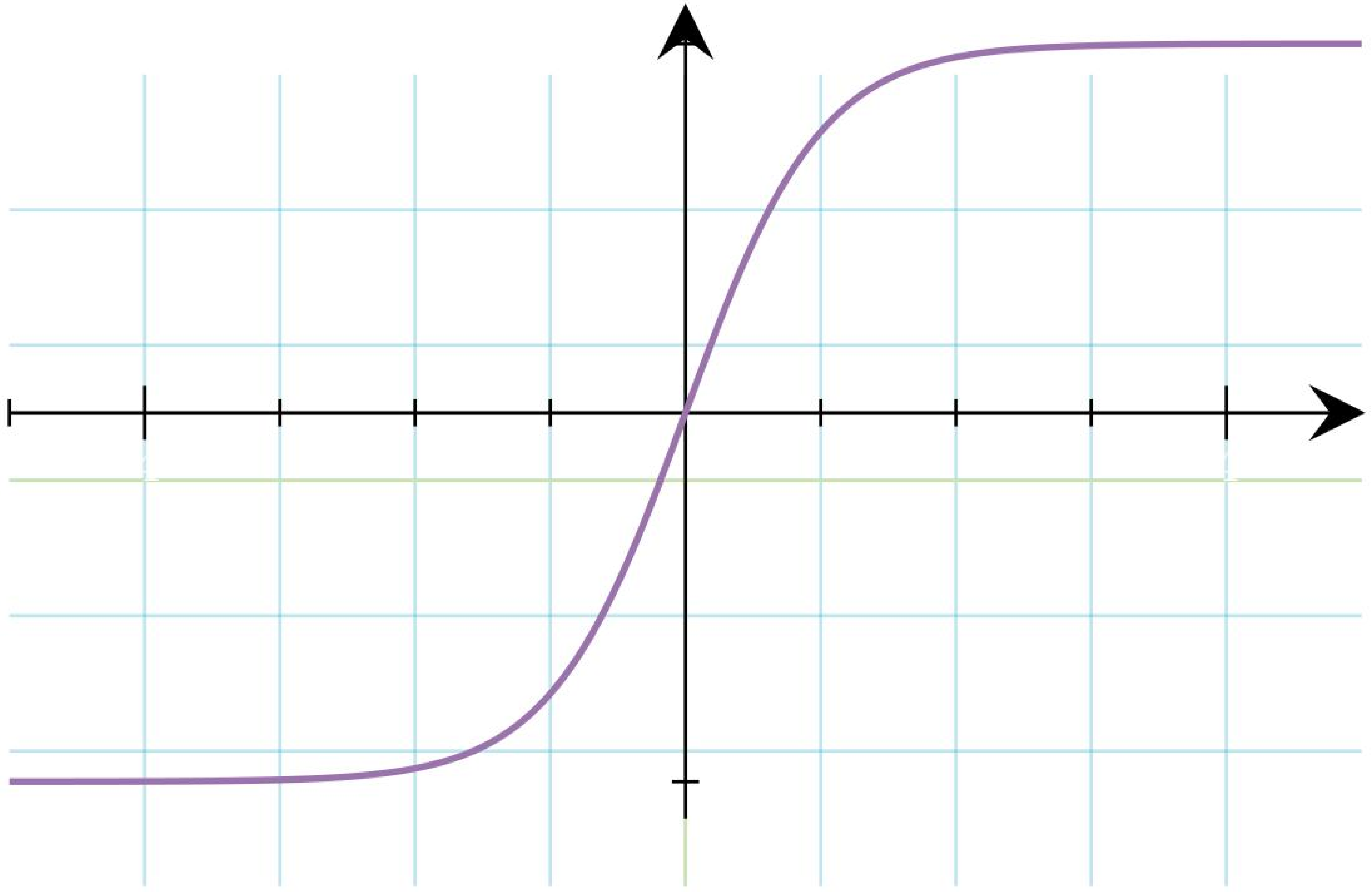
Tanh activation function

1. Form

$$\sigma(z) = \frac{2}{1 + \exp(-2z)} - 1$$

- Range: $(-1, 1)$
- Derivative: $\sigma'(z) = 1 - \sigma^2(z)$

Tanh activation function



Tanh activation function

1. Advantages
   - Mean zero for this activation function
   - Compared with sigmoid, the magnitude of gradient is larger near 0.
   - Function is symmetric about 0.
2. Disadvantages
   - Gradient vanishing.
   - Heavy computation for the derivative due to the exponential.
   - Do not induce sparsity

ReLU activation function

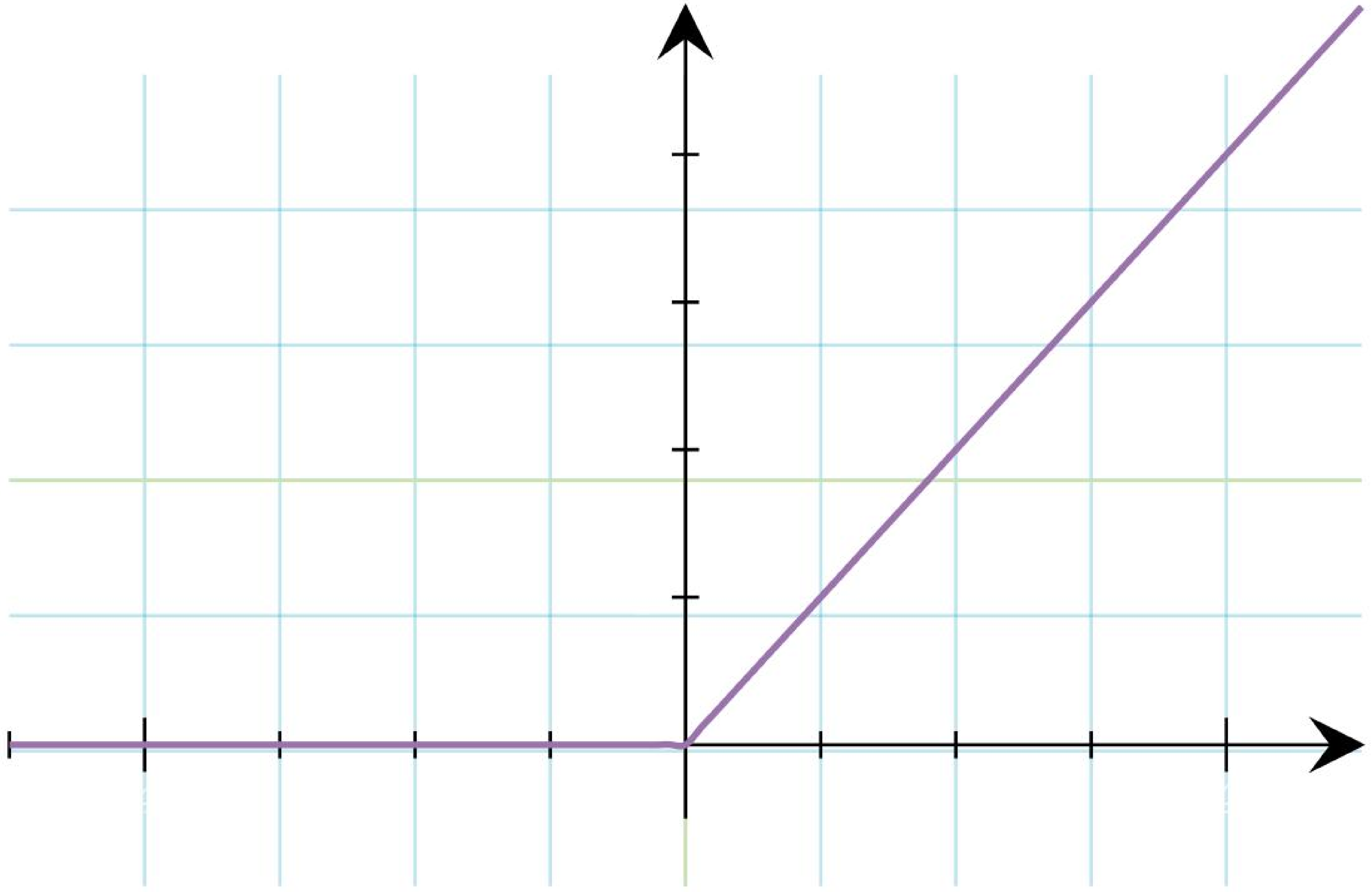
1. Form (Rectified Linear Unit)

$$\sigma(z) = \max\{0, 1\}$$

   - Range: $[0, \infty)$
   - Derivative:

$$\sigma'(z) = \begin{cases} 0 & z \leq 0 \\ 1 & z > 0 \end{cases}$$

ReLU activation function



ReLU activation function

1. Advantages
    - High computation efficiency
    - Alleviate the gradient vanishing problem in some sense
    - Sparse activation
    - Fast convergence rate
2. Disadvantages
    - Dying ReLU.
    - Output is not symmetric about 0
    - Not differentiable at $z = 0$

Leaky ReLU activation function

1. Form (Leaky Rectified Linear Unit)

$$\sigma(z) = \begin{cases} \alpha z & z \leq 0 \\ z & z > 0 \end{cases}$$

- α: a small number. For example, $\alpha = 0.01$
- Range: $(-\infty, \infty)$
- Derivative:

$$\sigma'(z) = \begin{cases} \alpha & z \leq 0 \\ 1 & z > 0 \end{cases}$$

Leaky ReLU activation function