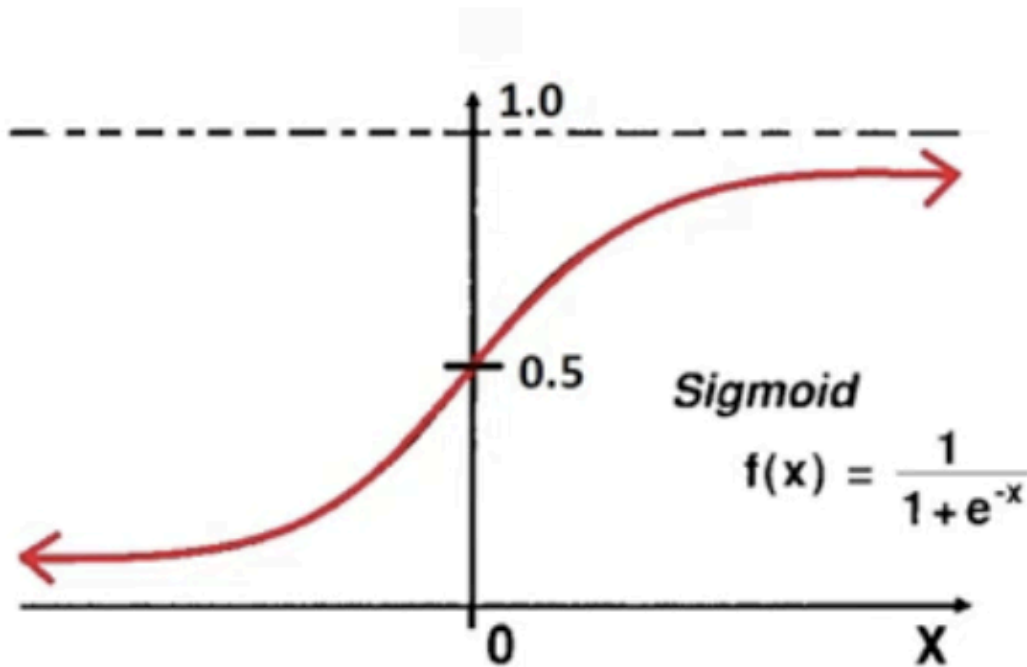


Activation Functions for Deep Learning

🔑 1. Sigmoid Activation Function

📊 Formula:



📈 Output Range: (0, 1)

✅ Advantages:

- Smooth and differentiable.
- Useful for **binary classification** (as output layer).
- Maps any input to a probability-like output.

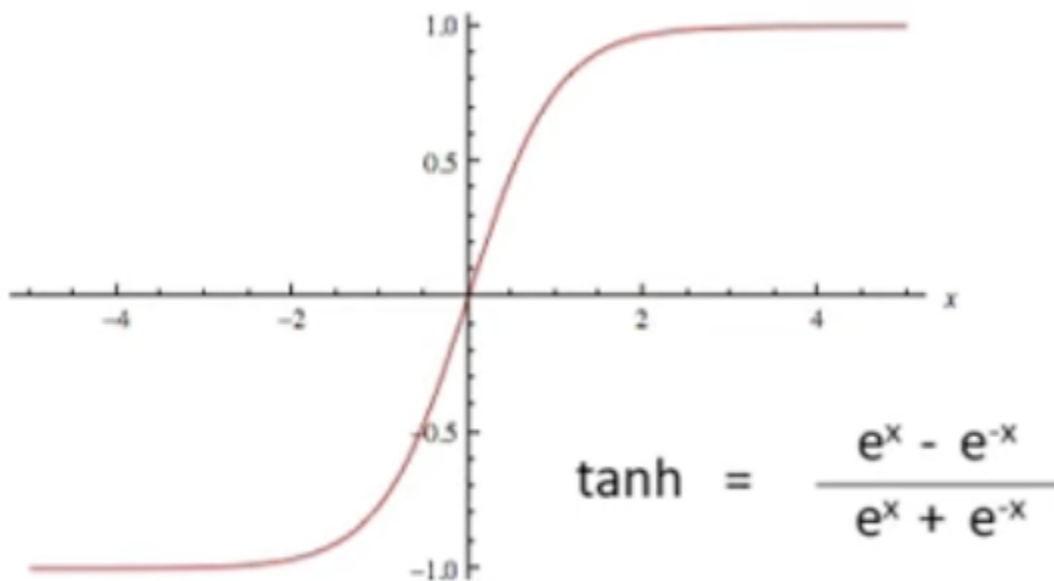
❌ Disadvantages:

- **Vanishing gradient problem:** Gradients become very small for large positive or negative inputs → slows learning.
 - **Not zero-centered:** Can cause zig-zag updates in gradient descent.
-

🔑 2. Tanh (Hyperbolic Tangent) Function



Formula:



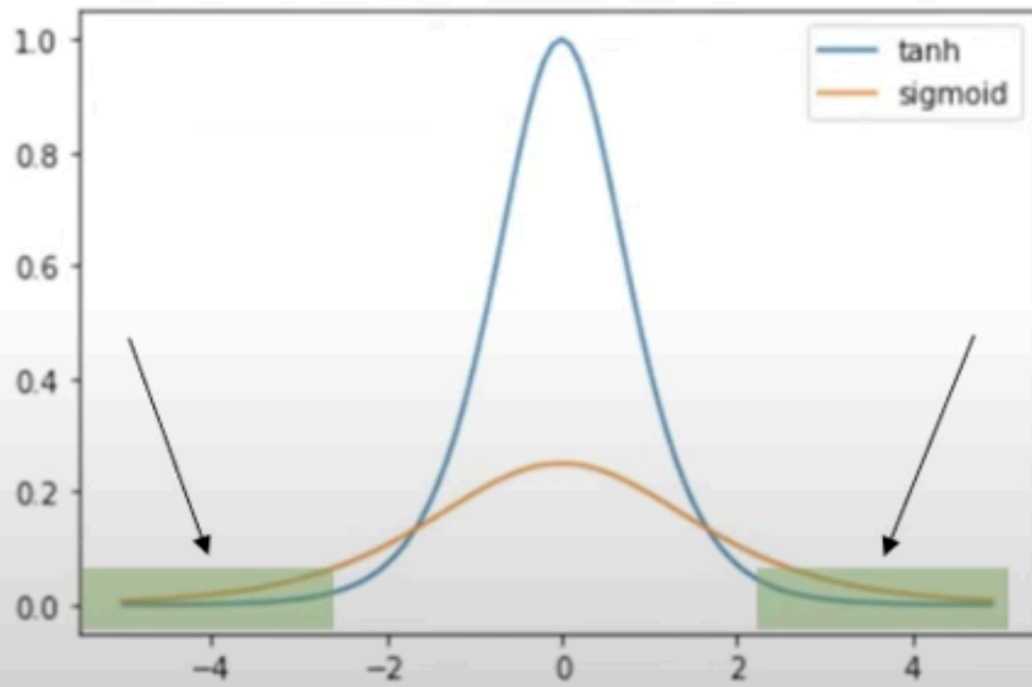
Output Range: (-1, 1)

✅ Advantages:

- Zero-centered → better weight updates than sigmoid.
- Stronger gradients than sigmoid in some regions.

✗ Disadvantages:

- Still suffers from **vanishing gradient** problem.



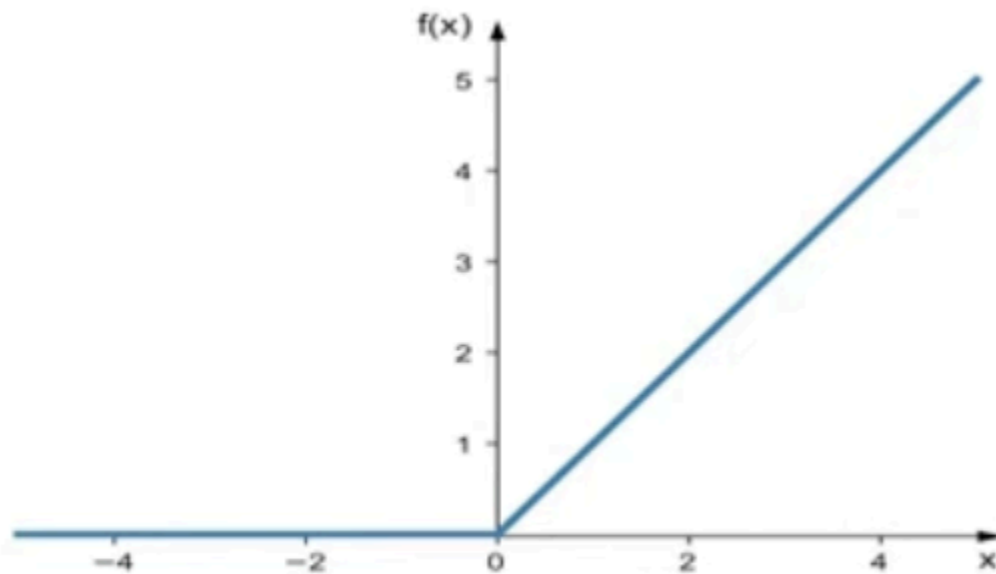
Derivatives

Vanishing Gradient Problem

🔑 3. ReLU (Rectified Linear Unit)



Formula:



$$F(x) = \max(0, x)$$



Output Range: $[0, \infty)$



Advantages:

- **Computationally efficient.**
- Avoids vanishing gradients in positive domain.

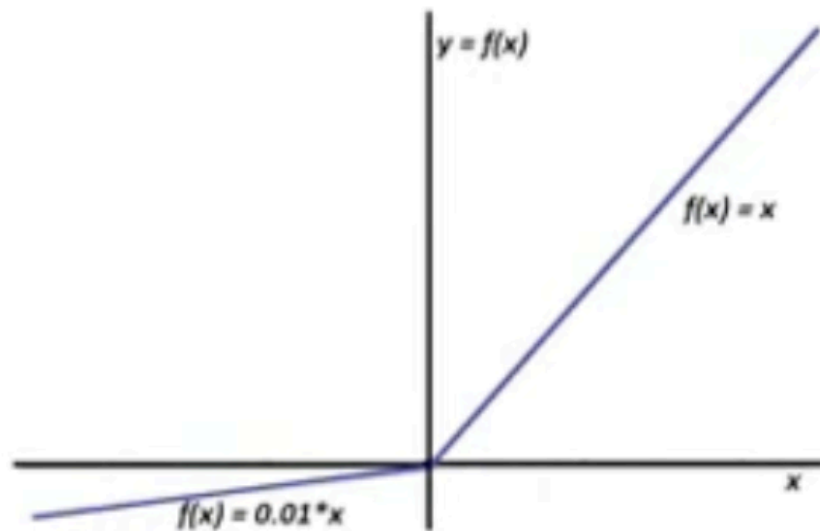


Disadvantages:

- **Dying ReLU problem:** Some neurons may always output 0 (if input < 0), and never learn (gradient = 0).
-

🔑 4. Leaky ReLU

📊 Formula:



✅ Advantages:

- Fixes Dying ReLU problem by allowing a small, non-zero gradient when $x < 0$.
- Still fast and simple like ReLU.

❌ Disadvantages:

- α is a hyperparameter and must be tuned.

- May still not solve the issue completely in some models.
-

6. ELU (Exponential Linear Unit)



Formula:



Advantages:

- Negative values bring mean activations closer to zero (helps convergence).
- Smooth curve → better gradient flow than ReLU.



Disadvantages:

- Slower to compute due to exponential function.
 - α must be tuned.
-



7. Softmax



Formula (for output layer in multi-class classification):



Output Range: (0, 1), and sum of all outputs = 1



Advantages:

- Converts raw scores into probabilities across multiple classes.
- Essential for **multi-class classification** problems.

✗ Disadvantages:

- Not used in hidden layers.
 - Sensitive to outliers and large input values → numerical instability (handled using **log-softmax** in practice).
-