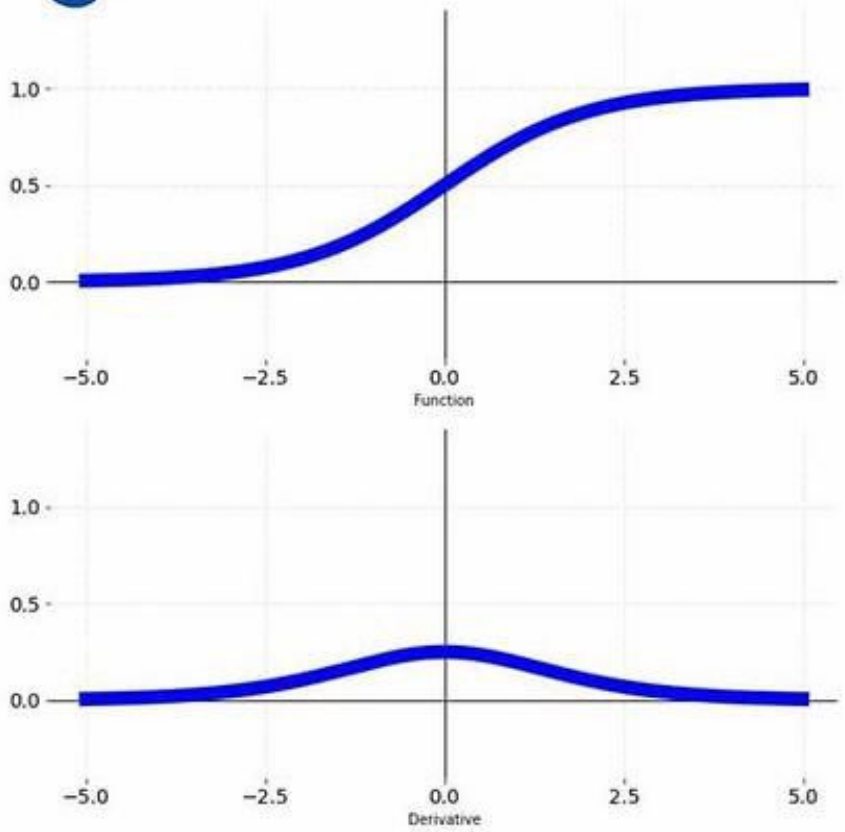
Sigmoid Function



$$f(x) = \frac{1}{1+e^{-x}} \qquad f'(x) = f(x)(1 - f(x))$$

It is a mathematical function having a characteristic that can take any real value and map it to between 0 to 1 shaped like the letter "S". The sigmoid function also called a logistic function.

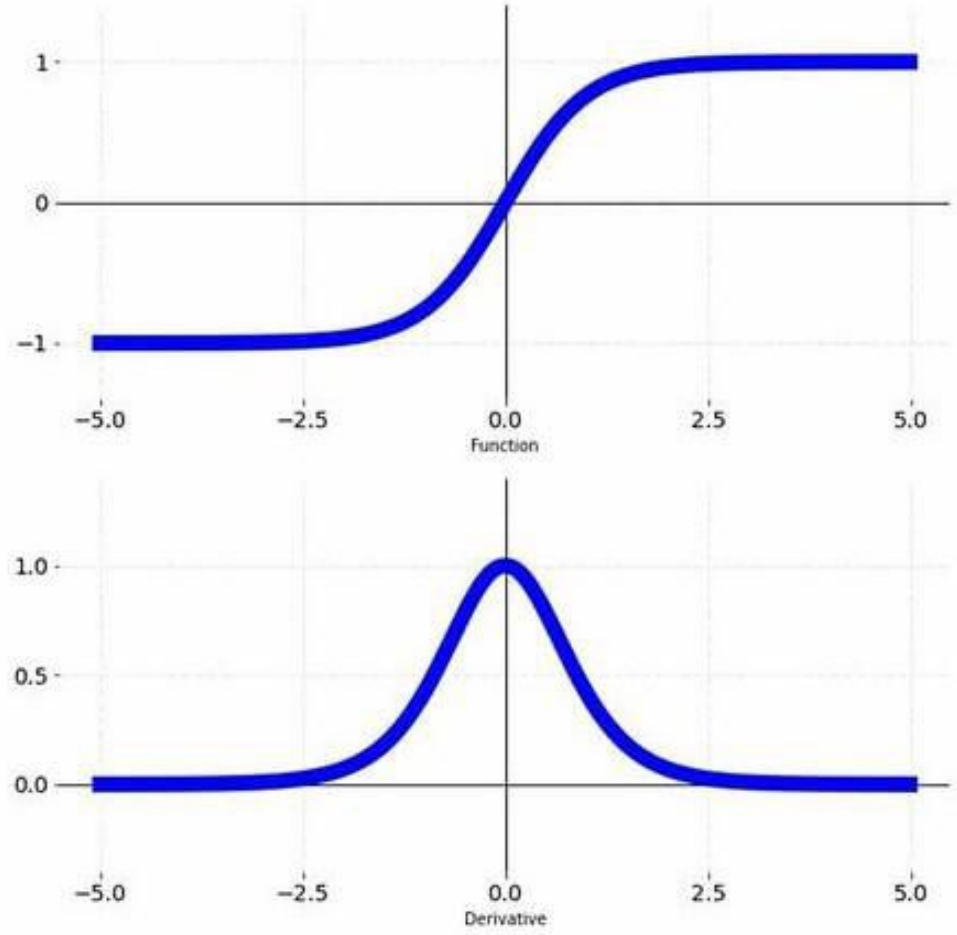
Pros

- Smooth gradient which doesn't give high jumps.
- Output is between 0 and 1.

Cons

- Prone to gradient vanishing
- Function output is not zero-centered
- Power operations are relatively time consuming

TanH Function



$$f(x) = \frac{1 - e^{-x}}{1 + e^{-x}} \qquad f'(x) = \frac{2xe^{x}}{(e^{x}+1)^{2}}$$

It's actually mathematically shifted version of the sigmoid function.the tanh function is used for the hidden layer and the sigmod function is used for the output layer.

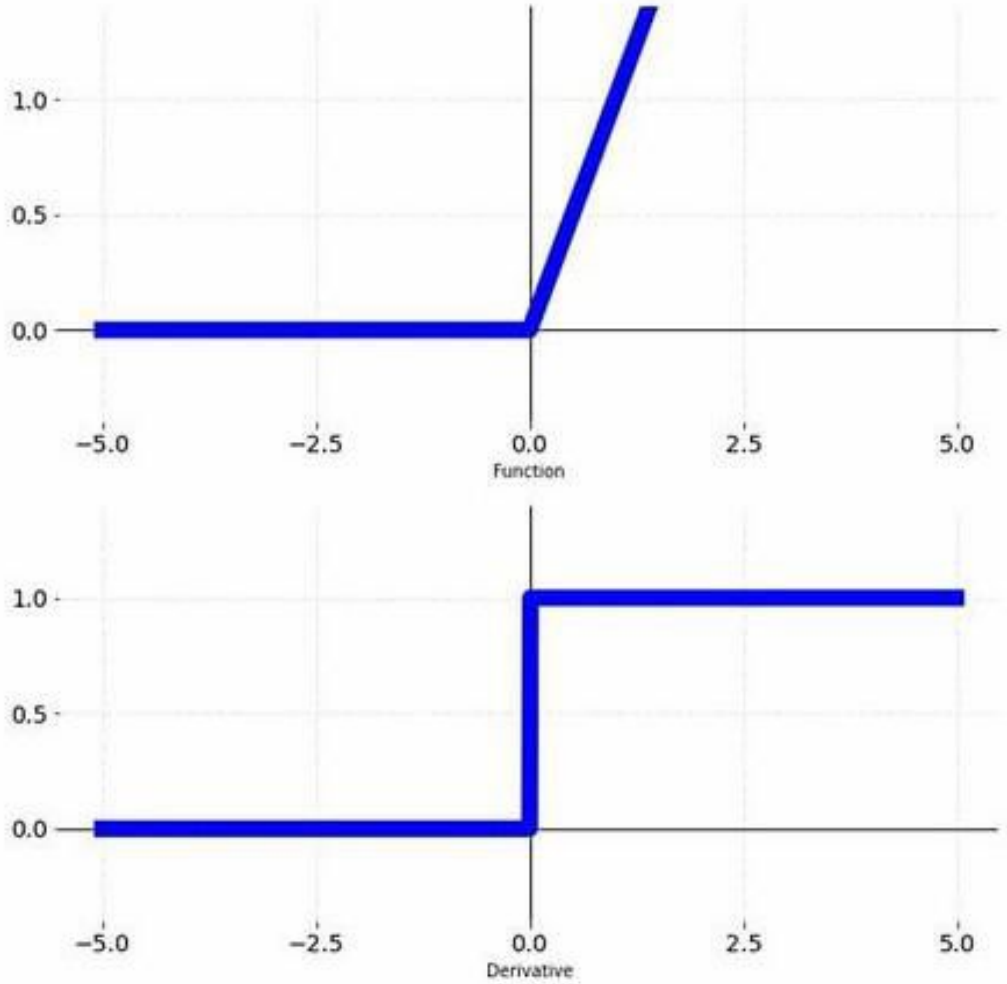
Pros

- Smooth gradient which doesn't give high jumps.
- Function is Zero Centered.
- The negative inputs will be mapped strongly negative and the zero inputs will be mapped near zero in the tanh graph.

Cons

- Prone to gradient vanishing
- Power operations are relatively time consuming.

ReLU Function



$$f(x) = \begin{cases} 0\,;x \leq 0 \\ x\,;\ x > 0 \end{cases} \quad f'(x) = \begin{cases} 0\,;x \leq 0 \\ 1\,;x > 0 \end{cases}$$

Rectified Linear Unit

Piecewise Linear Function which has made backpropagtion faster. ReLU performs better than Sigmoid and TanH.

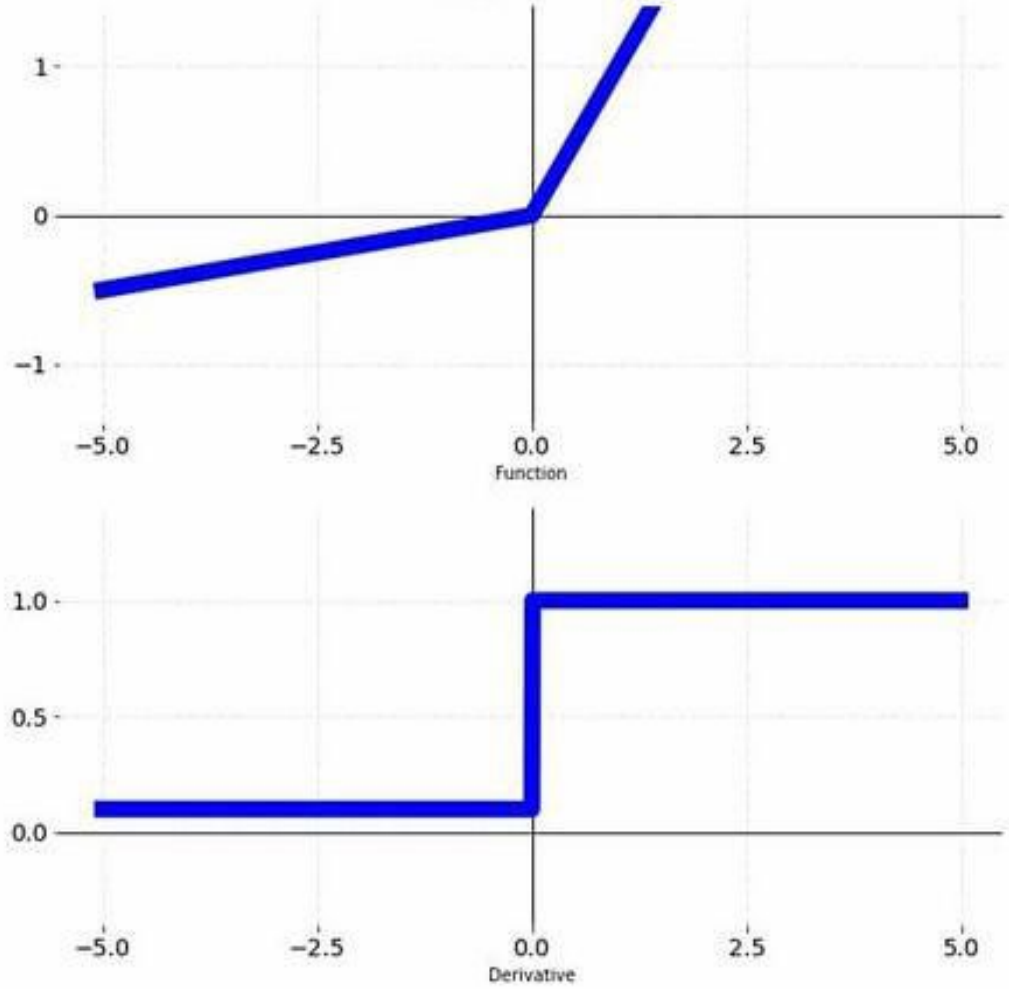
Pros

- Faster Computation
- Doesn't face problem of vanishing gradient.
- Linear Behaviour

Cons

- It is not Zero Centric
- For Negative values, output becomes 0 which gives rise to 'dying ReLU' problem.

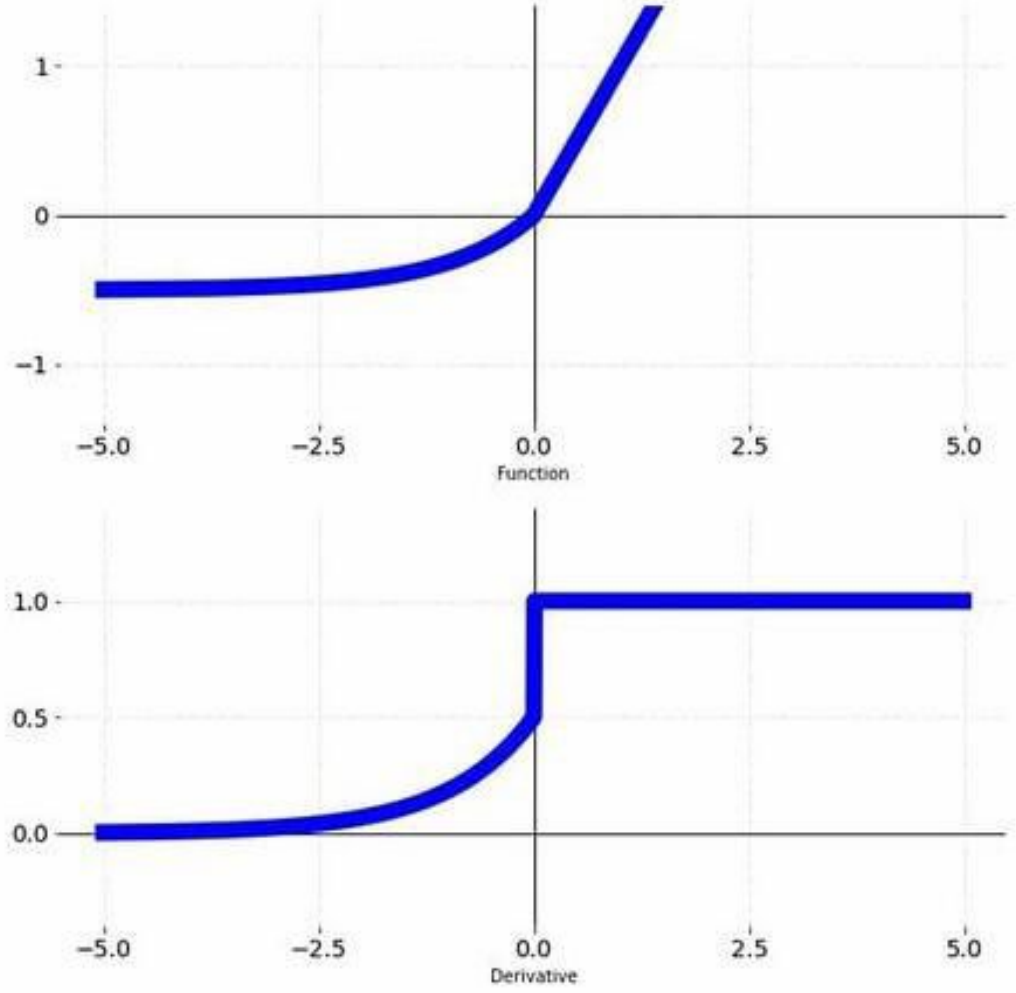
Leaky ReLU



$$f(x) = \begin{cases} \alpha x \; ; x \leq 0 \\ x \; ; \; x > 0 \end{cases} \qquad f'(x) = \begin{cases} \alpha \; ; x \leq 0 \\ 1 \; ; x > 0 \end{cases}$$

- To overcome dying ReLU, We allow some leaky-ness on the negative side. Leaky ReLU always performs better than ReLU.
- **Parametric Rectified Linear Unit(PReLU):**
    - We adaptively learn the parameter that controls the shape and leaky-ness (alpha).

Exponential Linear Unit



$$f(x) = \begin{cases} \alpha(e^x - 1)\,; x \leq 0 \\ x\,;\ x > 0 \end{cases} \quad f'(x) = \begin{cases} \alpha(e^x - 1) + \alpha\ ; x \leq 0 \\ 1\,; x > 0 \end{cases}$$

ELU

Exponential Linear Unit is a function that tend to converge cost to zero faster and produce more accurate results. ELU has a extra alpha constant which should be positive number.

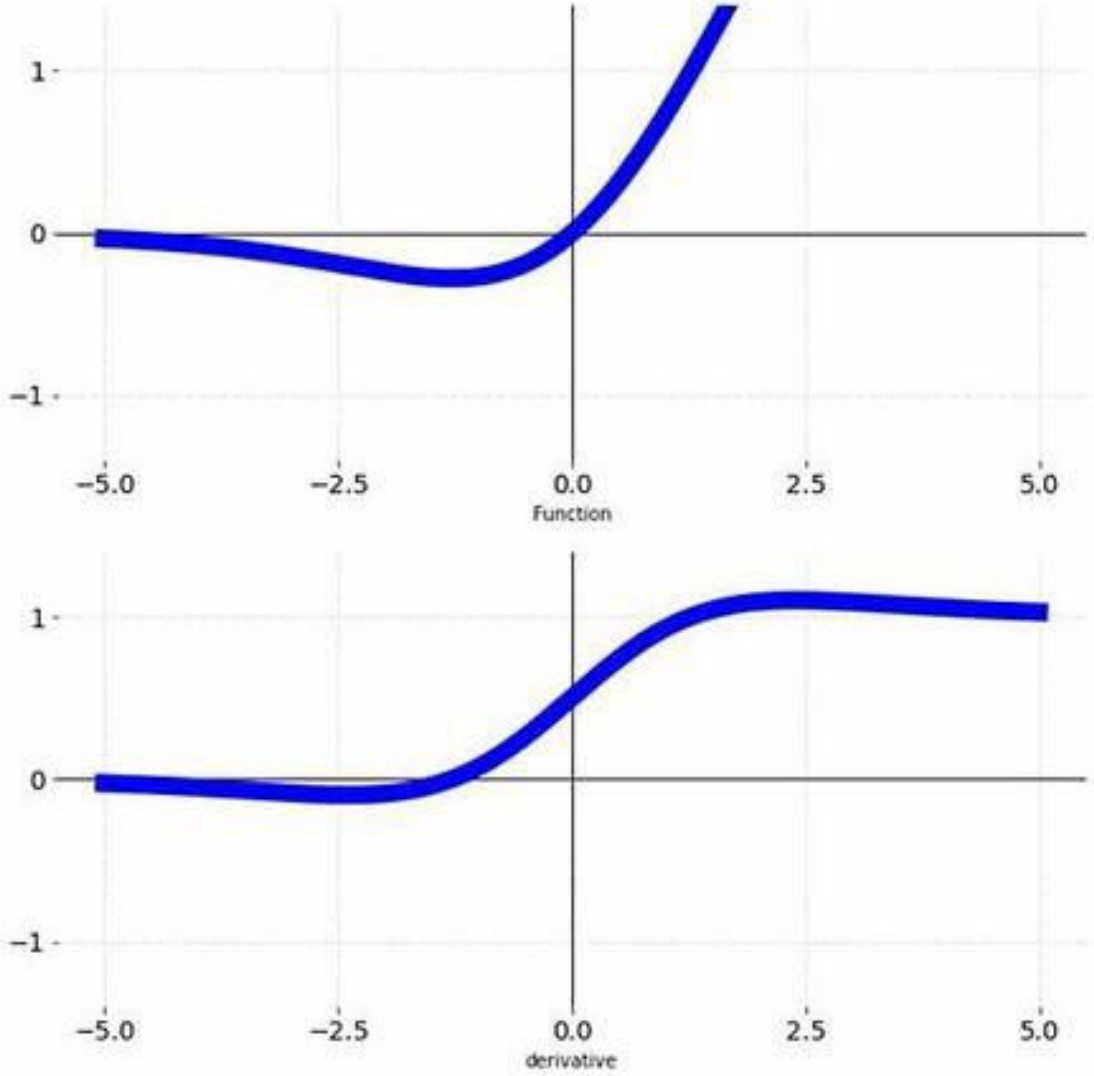
Pros

- ELU becomes smooth slowly until its output equal to -α whereas RELU sharply smoothes.
- ELU is a strong alternative to ReLU. ELU can produce negative outputs.

Cons

- For x > 0, it can blow up the activation with the output range of [0, inf].

Swish Function



$$f(x) = x * sigmoid(x) \quad f'(x) = \frac{e^{-x}(x+1)+1}{(1+e^{-x})^2}$$

Self Gating Function

- Self-Gating is the technique, an advantage of it is that it only requires a single input. Due to this Swish can easily replace ReLU as it also takes only a single scalar input.
- Smooth, non-monotonic function.
- Being unbounded on the x>0 side, it avoids saturation when training is slow due to near 0 gradients.
- Being bounded below induces a kind of regularization effect as large, negative inputs are forgotten.