Computer Vision and Pattern Recognition
Report
Name: SAGAR CHANDRA MODOK
ID: 20-42749-1

Activation Functions

Step Activation Function:

The step activation function is a simple non-linear activation function. It takes an input value and returns a binary output of either 0 or 1, depending on whether the input is greater than or equal to a specified threshold. The equation for this function is-

$$f(x) = \{ \begin{matrix} 1 \; if \; x > 0 \\ 0 \; if \; x \leq 0 \end{matrix} \}$$

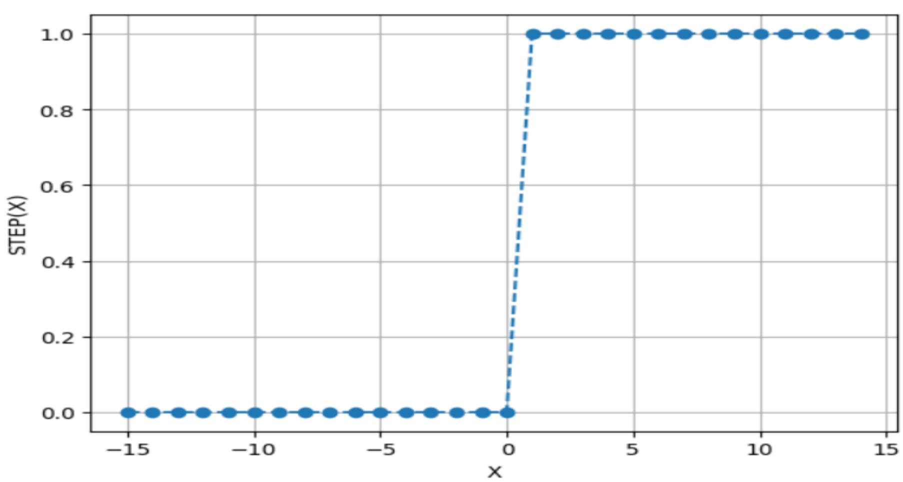


Fig-1: Step Function

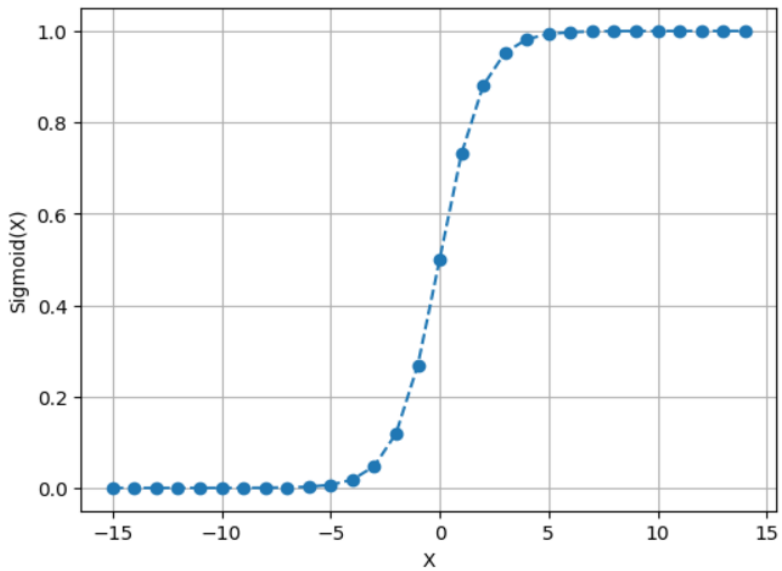
From the graph we can see that the output is a horizontal line at 0 for input values less than or equal to the threshold, and a horizontal line at 1 for input values greater than or equal to the threshold. It only returns 1(true) for input values that are greater than the threshold. For this reason, it is also known as threshold activation function. It is often used in binary classification problems, where the output needs to be either 0 or 1, such as in spam detection or sentiment analysis. However, its discontinuous nature can make it difficult to use in some situations, as it can cause problems with gradient-based optimization algorithms.

Sigmoid Activation Function:

The sigmoid function is a useful activation function that makes non-linear transformations of input values, especially in the context of probabilistic output that means the output of this function is always between 0 and 1. The graph looks like "S" shaped curve that gradually approaches an upper limit and a lower limit as the input value increases or decreases.

The equation for the sigmoid function is:

$$\sigma\,(x) = \frac{1}{1 + e^{(-x)}}$$



$$\sigma(x) = \; 1 + e^{(-x}$$

Fig-2: Sigmoi Fig-2: Sigmoid Function

The graph shows that for inputs close to zero, the output is close to 0.5, which is the midpoint of the function. As the input increases, the output approaches 1, and as the input decreases, the output approaches 0. The steepness of the curve is controlled by the slope parameter, which determines how quickly the output changes as the input changes. A steeper slope means that the output changes more quickly for small changes in the input, while a shallower slope means that the output changes more slowly for small changes in the input. But nowadays this activation function is rarely used for its vanishing gradient, computational expense and outputs not centered to 0 reasons.

Tanh Activation Function:

The tanh activation function is a type of sigmoid function that ranges from -1 to 1. Its graph has an "S" shape and is symmetric about the origin (0,0).

The formula of tanh function is:

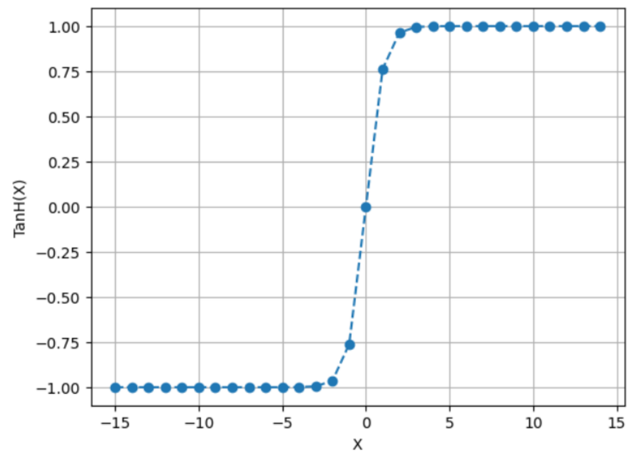


Fig-3: Tanh Function

In the graph we can see that at x = 0, the tanh function has a value of 0. As x approaches infinity, the tanh function approaches 1. As x approaches negative infinity, the tanh function approaches -1. This function is often used because it is differentiable and its output is bounded, which can help prevent exploding gradients during training. Additionally, it is similar to the sigmoid function but its output range is centered around 0, which can make it easier for the neural network to learn patterns in the data.

ReLu Activation Function:

The Rectified Linear Unit (ReLU) activation function generates a graph of a simple piecewise linear function and looks like a "hinge", with the output being zero for all negative inputs and increasing linearly for positive inputs.

The formula for this function is:

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

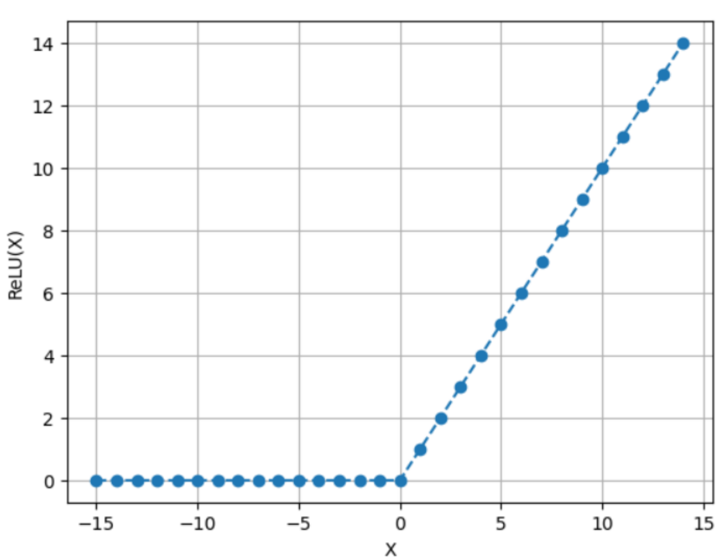


Fig-4: ReLu Function

The graph of the ReLU function is a straight line with a slope of 1 for all positive input values, and a flat line with an output of zero for all negative input values. The graph has a sharp bend at the origin where the function transitions from the flat line to the straight line. Because of its efficiency, sparsity and good generalization performance it is widely used. One potential downside of this function is, any negative input given to the ReLU activation function turns the value into zero immediately in the graph, which affects the resulting graph by not mapping the negative values appropriately.

SeLu Activation Function:

The Scaled Exponential Linear Unit (SELU) activation functions that induce self-normalizing properties.

The formula for this function is-

$$f(x) = \{ \begin{matrix} \lambda x \; if \; x \geq 0 \\ \lambda \alpha (e^x - 1) \; if \; x < 0 \end{matrix} \}$$

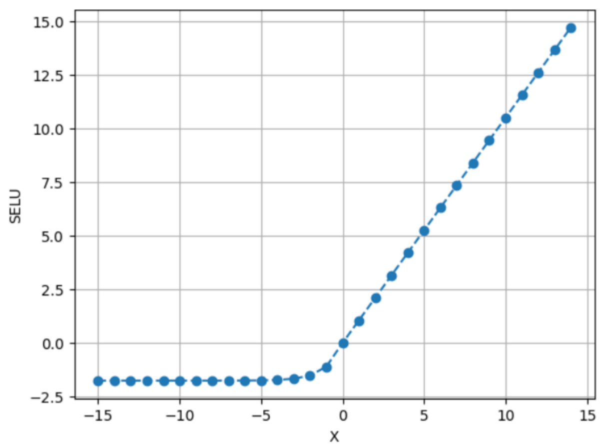


Fig-5: Selu Function

Here alpha(α) ≈ 1.67326324 and scale (λ) ≈ 1.05070098
If x is a positive value, then the output is x* λ but if x is equal to 0 or a negative value then we have a function in output that goes up to 0. One of its advantages is it can self-normalize the hidden units, which means that the output of each layer has zero mean and unit variance. This can lead to faster convergence and better performance, especially in deep neural networks. This function can also outperform ReLu. But because of its complexity and instability it is not suitable for all type of neural network and data.

ELU Activation Function:

ELU is an activation function that tend to converge cost to zero faster and produce more accurate results. The formula for this function is-

$$f(x) = \{ \begin{matrix} x \; if \; x > 0 \\ \alpha (e^x - 1) \; if \; x \leq 0 \end{matrix} \}$$

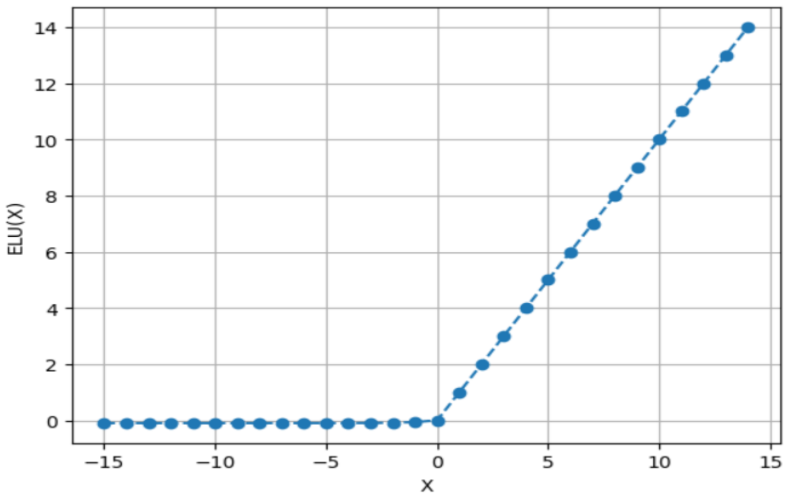


Fig-6: Elu Function

Here alpha is a hyperparameter that determines the value of the function when x is negative.

The graph of the ELU function is similar to the ReLU function, but it has a smooth curve for negative values of x, which can help address some of the drawbacks of the ReLU function, such as the "dying ReLU" problem, which occurs when a large number of neurons in a network are giving the output zero and unable to learn. When x is positive, the ELU function behaves like an identity function, returning x but it can alsoblow up the activation with the output range of [0, inf] which is a drawback of this function. When x is negative, the function returns a smoothed version of the exponential function, scaled by the value of alpha.This allows the function to have a non-zero gradient for negative values of x, which can help prevent the "dying ReLU" problem.