ACTIVATION FUNCTIONS

A neural network's activation function is an essential element. It is used to predict a neuron's or group of neurons' output from a collection of inputs. Moreover, activation functions are in charge of introducing non-linearity into the neural network, enabling more intricate choice limits and more precise forecasts. The activation functions come in a variety of forms, such as Step, Sigmoid, ReLU, and Tanh. The properties of six of the most popular activation functions—Step, Sigmoid, Tanh, ReLU, ELU, and SELU—will be compared in this research. Each type has particular advantages and disadvantages, and a neural network's performance can be strongly impacted by the activation function that is selected.

A binary function, the step activation function can either return 0 or 1, depending on the input. The binary classification issues use it the most since it is the simplest activation function. The step activation function's key benefit is that it is computationally effective and generates understandable results. It is ineffective for more complex jobs and can only be utilized for binary classification issues. The step function's gradient is zero, which is the major issue. As a result, during back-propagation, the step function is not very useful. Any gradient information from the input values is lost. The threshold value and a value 10000000 times that amount have equal effects. Moreover, disregards any input value.

The non-linear Sigmoid function produces values between 0 and 1. Given that it offers a smooth transition from 0 to 1, it is appropriate for both classification and regression applications. It provides the probability output for each class in multiclass classification. Another benefit of this function is that it returns a value between when used with (-infinity, +infinity) as in the linear function (0, 1). The vanishing gradient issue is one of the flaws with this function. Incase This function is not zero-centered; a zero-centered function has 0 in the middle of its function range. The majority of machine learning algorithms perform better with zero centers. Also, the cost of computing is high (exponential in nature).

In addition, the output values of Tanh activation functions vary from -1 to 1, and they are a sort of continuous activation function like Sigmoid but are symmetric across the origin. Problems involving classification are where this kind of activation function is most frequently applied. The outputs from Tanh will be zero centered. In addition, it resolves the issue of values with identical signs. As this function's nature is non-linear, errors can be easily back-propagated. The sigmoid function still suffers from the issue of vanishing gradients, which is a comparable flaw. Costly in terms of computation (exponential in nature). ReLU (rectified linear unit) activation functions are a sort of continuous activation function that, depending on whether the input is larger than or less than 0, either returns 0 or the input value. Deep learning networks are the most typical application of this kind of activation function. In the ReLU function is non-linear, easy to backpropagate errors, and can be used to trigger neurons in several layers. In addition, compared to the sigmoid and tanh activations, this function speeds up the convergence of stochastic gradient descent. It is computationally effective, which enables the network to converge relatively quickly. It has the ability to emit a real zero value. ReLU output is not zero centered, which reduces the neural network's effectiveness. The gradients of the weights will either be uniformly positive or uniformly negative during backpropagation. When the neuron becomes stuck on the negative side and consistently outputs zero, there is a dying ReLU problem. This occurs either when there is a substantial negative bias or when learning rate is high.

Similar to the ReLU function, the ELU (Exponential Linear Unit) function produces a negative value for negative inputs. because it guarantees that neurons are active despite negative inputs. The issue of disappearing gradients and expanding gradients is not a concern at ELU. ELU does not experience the issues associated with decaying neurons like ReLU does. In comparison to ReLU and its derivatives, using ELU results in a shorter training time and greater accuracy in neural networks. At any location, the ELU activation function is continuous and differentiable. The possibility of negative ELU values causes the mean of the activations to be pushed closer to zero. Moreover, learning and convergence occur more quickly when mean activations are closer to zero. ELU takes longer to compute because of its non-linearity with negative inputs.

Scaled Exponential Linear Units, or SELUs, are activation mechanisms that promote self-normalization. Neuronal activations in SELU networks naturally converge to a zero mean and unit variance. The output of the entire layer is maintained at a constant value thanks to the scaling of this function, which is comparable to the ELU function. By preventing the output from getting too small, this helps to lessen the issue of vanishing gradients. SELUs cannot die compared to ReLUs. Instead of requiring additional processing, SELUs learn more quickly and more effectively than other activation mechanisms. It is not yet common practice to use this activation function because it is a newer technology.

There are advantages and disadvantages to each activation function. The vanishing gradient issue affects the sigmoid function despite its ease of understanding and implementation. Although the ReLU function is effective and aids with vanishing gradient relief, it is plagued by the issue of dead neurons. ELU is able to solve this issue, however it takes longer to test. SELU is capable of resolving current issues, but additional study is necessary before its function can be applied more generally. It's crucial to select the proper activation function based on the particular task and dataset. To enhance the functionality of deep learning models, researchers and practitioners are still investigating and experimenting with various activation functions.