American International University-Bangladesh

Name: Ismail Hossain Rishad

ID: 20-43002-1

Course : CVPR[C]

Step Function: A step function is a function like that used by the original Perceptron. The output is a certain value, A_1, if the input sum is above a certain threshold and A_0 if the input sum is below a certain threshold.

Advantages:

- Great for binary classification
- It is a piecewise constant function which indicates that it maintains a constant value throughout a certain range of input and it is called constant function.
- The step function is computationally efficient , it is easy to calculate the output for any given input.

Disadvantages:

- The gradient of the step function is zero. This makes the step function not so useful since during back-propagation when the gradients of the activation functions are sent for error calculations to improve and optimize the results.
- It cannot be used for multi-class classification

Sigmoid : A sigmoid function is an activation function which is used to add the non-linearity in a model. It has a s shape curve. It f determines which value need to pass as output where the output needs to be a probability between 0 and 1 and what not to pass .

The sigmoid function is defined as: $\sigma(x) = \frac{1}{1+e^{-x}}$

Advantages:

- Gives you a smooth gradient while converging.
- One of the best Normalised functions.
- Gives a clear prediction(classification) with 1 & 0.

Disadvantages:

- Prone to Vanishing Gradient problem.
- Not a zero-centric function(Always gives a positive values).
- Computationally expensive function(exponential in nature).

Tanh: The tanh function is mainly used classification between two classes. Both tanh and logistic sigmoid activation functions are used in feed-forward nets. It is also called as Hyperbolic tangent. Tanh function has a range between -1 and 1, meaning that it squashes its input values into this range. This can be useful in neural networks as it allows the outputs of the network to be normalized and centered around zero.

The formula for the tanh function is tanh(x)= $\frac{e^{2x}-1}{e^{2x}+1}$

Advantages:

- Zero-centric function unlike Sigmoid.
- It is a smooth gradient converging function.
- It is differentiable which allows for gradient-based optimization methods used during training .

Disadvantages:

- Prone to Vanishing Gradient function.
- Computationally expensive function(exponential in nature).
- The output of the tanh function can saturate to the extremes(-1 or 1)for large input values which can limit the representational power of the neural network .

RELU: The full form of RELU is Rectified Linear Unit. It is a non-linear function that will output the input if it is positive otherwise it will output zero. It is especially used in Convolutional Neural Network and Multilayer perceptron.

The formula of RELU function is : f(x)= max (0,x)

Advantages:

- Can deal with Vanishing Gradient problem.
- Computationally inexpensive function(linear in nature).
- It is a simple function which helps to compute quickly and for this it makes it ideal for large neural network.

Disadvantages:

- Not a zero-centric function.
- Gives zero value as inactive in the negative axis.
- It is unbounded which means the output values can be very large for very large input values. This can lead to numerical instability and difficulty during training.

ELU: The full form of ELU is Exponential Linear Unit. It maps the input values to the output range of (-1,∞)

The ELU activation function is defined by the following equation:

$$ELU(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha(e^x - 1) & \text{if } x < 0 \end{cases}$$

Where x is the input to the function, α is the hyperparameter that determines the value that the function approaches for large negative values of x.

Advantages:

- Doesn't have the dying ReLU problem.
- The function tends to converge cost to zero faster and produces more accurate results(recently read it on a blog).
- More of a merger between good features of ReLU & Leaky ReLU

Disadvantages:

- Exploding gradient problem .
- It is lower to compute comparison to RELU

SELU:

$$f(x) = \lambda x \qquad \text{if } x > 0$$

$$f(x) = \lambda\alpha(e^x - 1) \qquad \text{if } x \leq 0$$

The SELU activation function is a variant of the Exponential Linear Unit (ELU) activation function that has been shown to work well in deep neural networks.

Advantage:

- Like ReLU, SELU does not have vanishing gradient problem
- SELUs learn faster and better than other activation functions without needing further procession.

Disadvantage:

- SELU is a relatively new activation function so it is not yet used widely in practice. ReLU stays as the preferred option.
- The computation of the SELU function involves exponential and logarithmic operations, which can make it computationally more expensive.