Activation functions

Activation functions are used in artificial neural networks (ANNs) to introduce non-linearity into the model. This allows the ANN to learn more complex relationships between the input and output data .The activation function decides whether a neuron should be activated or not by calculating the weighted sum and further adding bias to it.

There are many different types of activation functions, but some of the most common ones include:

1-Sigmoid:

The sigmoid function is a S-shaped curve that is often used in classification problems. It has a range of [0, 1], which makes it well-suited for representing probabilities.

Usually used in output layer of a binary classification, where result is either 0 or 1, as value for sigmoid function lies between 0 and 1 only so, result can be predicted easily to be ***1*** if value is greater than **0.5** and ***0*** otherwise

It is smooth and differentiable. However, it has a problem of vanishing gradient for large x values.

Equation : $A = 1/(1 + e^{-x})$

2- Tanh:

Tanh function also known as **Tangent Hyperbolic function**. It's actually mathematically shifted version of the sigmoid function. Both are similar and can be derived from each other but it has a range of [-1, 1]. This makes it more suitable for regression problems.

Usually used in hidden layers of a neural network as it's values lies between **-1 to 1** hence the mean for the hidden layer comes out be 0 or very close to it, hence helps in *centering the data* by bringing mean close to 0. This makes learning for the next layer much easier.

Equation : (2/(1+e^-2x))-1

3- ReLU:

It Stands for *Rectified linear unit*. It is the most widely used activation function. Chiefly implemented in *hidden layers* of Neural network

The ReLU function is a piecewise linear function that is often used in deep learning models. It is very efficient to compute, and it can help to prevent the ANN from overfitting the training data.

ReLu is less computationally expensive than tanh and sigmoid because it involves simpler mathematical operations. At a time only a few neurons are activated making the network sparse making it efficient and easy for computation. It doesn't have vanishing gradient issue. But it is not differentiable at x=0.

Equation : *A(x) = max(0,x)*. It gives an output x if x is positive and 0 otherwise

4- Leaky ReLU:

Leaky Rectified linear unit(Leaky Relu) is a variant of the ReLU function that allows for a small output value when the input is negative. This can help to improve the performance of the ANN on certain types of problems.

Relu return 0 if the input is negative and hence the neuron becomes inactive as it does not contribute to gradient flow. Leaky Relu overcomes this problem by allowing small value to flow when the input is negative. So, if the learning is too slow using Relu, one can try using Leaky Relu to see any improvement happens or not.

5- ELU:

The exponential Linear Unit is also similar to Leaky Relu but differs for negative input

The ELU function is another variant of the ReLU function that has a more gradual slope than the ReLU function. This can help to prevent the ANN from dying, which is a problem that can occur with the ReLU function. It has the same purpose that of Leaky Relu and convergence of cost function towards zero is faster than Relu as well as Leaky Relu. For example, neural network learning on Imagenet using Elu is faster than using Relu.

6- Selu function:

Scaled Exponential Linear Unit is the scaled form of Elu. Just multiply the output of Elu by a predetermined “scale” parameter and you will get the desired output which selu gives. This activation function is used in Self-Normalizing Neural Networks (SNNs) which is used to train a deep and robust network less effected from vanishing and exploding gradient problem.

7- Softmax:

The softmax function is also a type of sigmoid function but is handy when we are trying to handle multi- class classification problems . It is used as the final activation for classification problems with multiple classes. It converts the net output into a probability distribution over the classes.

Usually used when trying to handle multiple classes. the softmax function was commonly found in the output layer of image classification problems .The softmax function would squeeze the outputs for each class between 0 and 1 and would also divide by the sum of the outputs.

8- **Softsign function**:

Softsign function is an alternative to tanh function where tanh converges exponentially and softsign converges polynomially. It is mostly used in the regression problem and can be used in a deep neural network for text to speech conversion

Equation : softsign(x) = x / (1 + abs(x))

9- **Softplus function**:

Softplus function is a smoothed form of the Relu activation function and its derivative is the sigmoid function. It also helps in overcoming the dying neuron problem. Some experiments show that softplus takes lesser epochs to converge than Relu and sigmoid. It can be used in the speech recognition system.