Name: Tahsin Ferdous

ID: 20-43413-1

Section: C

Activation function in neural networks

Introduction:

In artificial neural networks, the activation function of a node defines the output of that node given an input or set of inputs.

An activation function is a function used in artificial neural networks which outputs a small value for small inputs, and a larger value if its inputs exceed a threshold.

Variants of activation function:

1) **Sigmoid function:** It is a function which is plotted as **'S'** shaped graph. It takes a real-valued number.
   Equation: $A = 1/(1 + e^{-x})$
   Nature: Non-linear.
   Value Range: 0 to 1
   Uses: It is especially used for models where we have to predict the probability as an output.
   It is helpful for neural network and logistic regression model. This function also have some problems like vanishing gradient problem, not being zero-centered etc.

2) **Step function:** The step or threshold activation function is a simple activation function.
   Equation: y=f(x)=[x]
   Nature: Linear.
   Value Range: 0 to 1.
   Step activation function is effective for binary classification problems and can be used in neural networks as an activation function. The discontinuous nature of this application results in more intricate issues.

3) **Tanh function:** The activation that works almost always better than sigmoid function is Tanh function also known as Tangent Hyperbolic function.
   Equation: (e^2x) -1/ (e^2x)+1
   Nature: Non-linear.
   Value Range: -1 to +1.

This function is capable of modeling complex relationships between input and output variables. The computation of exponential required for tanh can be computationally expensive.

4) **RELU function:** It Stands for Rectified linear unit. It is the most widely used activation function. Chiefly implemented in hidden layers of Neural network.
Equation: A(x)= max(0,x)
Nature: Non-linear
Value Range: [0, inf)
Uses: 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.
The ReLU is the most used activation function in the world right now. Since, it is used in almost all the convolutional neural networks or deep learning.

5) **ELU function:** ELUs saturate to a negative value with smaller inputs and thereby decrease the forward propagated variation and information.
Nature: Non-linear
The ELU function applies an exponential function to x, substracts 1 and multiplies by alpha when x is negative. This function is computationally expensive and difficult to implement in certain hardware and software environments.

6) **SELU function:** SELUs, or Scaled Exponential Linear Units, are activation functions that induce self-normalization. SELU network neuronal activations automatically converge to a zero mean and unit variance.
Nature: Non-linear
Unlike ReLU, it can get below 0, allowing the system to have a zero average output. As a result, the model may converge faster. SELU's are mostly commonly used in Self Normalizing Networks (SNN). The output of a SELU is normalized, which could be called internal normalization, hence the fact that all the outputs are with a mean of zero and standard deviation of one. The main advantage of SELU is that the Vanishing and exploding gradient problem is impossible and since it is a new activation function, it requires more testing before usage.

Activation functions have a major effect on the neural network's ability to converge and the convergence speed, or in some cases, activation functions might prevent neural networks from converging in the first place. Activation function also helps to normalize the output of any input in the range between 1 to -1 or 0 to 1.