Activation function

An activation function is a mathematical function used in neural networks to introduce non-linearity into the output of a neuron or a group of neurons. The non-linear transformation provided by the activation function is important because it allows the neural network to learn and model complex relationships between the inputs and outputs.

The activation function is applied to the weighted sum of the input values and bias term of a neuron to determine the output of that neuron. This output is then passed on to other neurons in the next layer of the network.

There are several types of activation functions used in neural networks, some of which are:

Sigmoid/Logistic Function: The sigmoid function is defined as f(x) = 1/(1 + e^(-x)). It maps the input to the range [0, 1], which makes it useful for binary classification problems where the output is a probability score between 0 and 1. However, the sigmoid function suffers from the vanishing gradient problem when the input values are too large or too small, which can slow down the learning process.

Hyperbolic Tangent (tanh) Function: The tanh function is defined as f(x) = (e^x - e^(-x))/(e^x + e^(-x)). It maps the input to the range [-1, 1], which can be useful for classification problems where the output is bipolar. The tanh function is also differentiable, making it useful for backpropagation during the training of neural networks.

Rectified Linear Units (ReLU): The ReLU function is defined as f(x) = max(0, x). It maps the input to the range [0, inf) and is known for its simplicity and effectiveness in deep neural networks. The ReLU function is non-linear and computationally efficient, which makes it a popular choice for neural network models.

Leaky ReLU: The Leaky ReLU function is a modified version of the ReLU function that addresses the problem of "dead" neurons that can occur when the input is negative. The Leaky ReLU function is defined as f(x) = max(ax, x), where a is a small positive value. This allows some of the negative values to pass through the neuron, which can help to prevent "dead" neurons from occurring.

Softmax Function: The softmax function is used in multi-class classification problems to normalize the output scores of the neurons so that they represent probabilities that sum up to 1. The softmax function is defined as f(x_i) = e^(x_i) / sum(e^(x_j)) for all j in the output layer.

Activation functions are used in various types of neural networks, including feedforward networks, convolutional neural networks (CNNs), and recurrent neural networks (RNNs). The choice of activation function depends on the problem being solved, the architecture of the neural network, and the nature of the input and output data.

For example, the sigmoid function can be used in a binary classification problem to predict whether a given input belongs to one of two classes. The ReLU function can be used in a deep neural network to learn complex features in image data for object recognition tasks. The softmax function can be used in a multi-class classification problem to predict the probability of an input belonging to each of the possible classes.