**Artificial Neural Networks**

Artificial Neural Network is a system composed of interconnected input data sets that is used to train a machine for their dynamic response with optimum accuracy.

Basic computational unit of neural networks is referred as nodes. Each node is associated with the weight which decides its significance for the next input. Each node takes input from the preceding node and provide the output to the next node by calculating the weighted sum of all the inputs.

**Activation Function**

Activation function is basically used for controlling the output traffic i.e it basically decides whether the output need to be activated or not on the basis of the weighted sum calculated for that particular node.

The deciding factor for the activation of a node is the threshold value. If the weighted sum is beyond the threshold value, that node will get activated.

**Sigmoid Activation Function Vs ReLu Activation Function**

|  |  |
| --- | --- |
| **Sigmoid Activation Function** | **ReLu Activation function** |
| Graph looks like S- Curve that extends between 0 to 1. | Graph looks like a straight line and is considered to be half rectified. |
| Since the end lies between 0 and 1, Sigmoid activation Function is preferably used to determine the probability | In case of ReLu activation function if f(z) becomes zero if value of z is negative. |
| Function is monotonic but derivative is non-monotonic | Both function and derivative are monotonic. |
|  | All negative values become 0 due to which the model cannot be trained on the data set properly. |

**Error Function/Loss function**

Use to determine a node in the graph where there lies the minimum value of the error using “Gradient Descent” methodology.  
It basically tells us that how well our model will predict the expected output.

**Bias** is an input node which is constant. It allows graph to fit better in the predictions made with the available datasets. Bias help activation function to travel through the layer with an optimized graph fit.

**Early stopping**

Early stopping is done when we have prior idea that our graph is going to be overfit the nodes and to cease the unnecessary calculations while optimizing the error function using Gradient descent.

**Network and Hyperparameters on which accuracy of the graph depends-**

**Number of neurons in a layer** is the network parameter on which the accuracy of our model depends. We can also introduce hidden layers to make it possible for the network to exhibit non-linear behavior.

**Number of Epochs**

Epochs is a 1 forward pass and 1 backward pass for each data input in the training dataset is referred as Epochs.

Therefore, number of Epochs in the code signifies the number of one complete propagation of a node.

**Batch Size**

Batch size is the number of samples we need to the network at one time. We cannot pass whole dataset at once due to its large size so we divide the data into the Batch size.

**Learning rate**

Learning rate is an instance for optimizing the loss function. It basically tells how fast our function converges to the minimum value.

**Optimizer function** is another network parameter on which the accuracy of the model depends by processing the Learning rate.

|  |  |  |
| --- | --- | --- |
| **Inputs** | **Outputs** | **Graphs** |
| Number of neuron hidden layer=1000  Learning Rate=0.08  No. of epochs=5000  Batch size=850 | Errors: 2  Correct :98  Testing Accuracy: 98.0 |  |
| Number of neuron hidden layer=1000  Learning Rate=0.07  No. of epochs=5000  Batch size=850 | Errors: 1  Correct :99  Testing Accuracy: 99.0 |  |
| Number of neuron hidden layer=1000  Learning Rate=0.07  No. of epochs=4000  Batch size=600 | Errors: 3  Correct :97  Testing Accuracy: 97.0 |  |
| Number of neuron hidden layer= 2000  Learning Rate= 0.02  No. of epochs= 7500  Batch size= 140 | Errors: 0  Correct :100  Testing Accuracy: 100.0 |  |
| Number of neuron hidden layer= 2000  Learning Rate= 0.04  No. of epochs= 6000  Batch size=700 | Errors: 2  Correct :98  Testing Accuracy: 98.0 |  |
| Number of neuron hidden layer= 2000  Learning Rate= 0.01  No. of epochs= 6000  Batch size=100 | Errors: 1  Correct :99  Testing Accuracy: 99.0 |  |