

Activation Function and its parameter plays very important role in neural networks, without it the entire network will work as a single linear network it bring non-linearity between layers for e.g. If you chain several linear transformation you get linear transformation i.e. if $f(x) = 3x + 5$, $g(x) = 2x + 2$ is chained u get $f(g(x)) = 3x(2x+2) + 5(2x+2) = 6x+6+10x+10 = 16x+16$ which is a linear equation.

Deep Neural network consists of many network, it is important to reduce cost function and that is done by Gradient Descent .First the input is passed to input layer and the output from that layer is passed on to the next layer and so on till the last layer. This is forward pass it is just like prediction and then there is calculation of error between output predicted value and actual value and then again it moves in backward direction and measures the error contributed by every single weight at each step this is called Back propagation. As it finds error contributed by each parameter it performs Gradient Descent to tweak the parameters at each level to reduce error.

For this algorithm to work perfectly it is important to have activation layer between each step. The popular activation function are Log Function also known as sigmoid function, hyperbolic tangent function(tanh), Rectified Linear unit(Relu) we have performed with all the three activation function in our assignment. The weight and bias initialization should be random if it is not random and all the bias are same the algorithm won't be able to perform complex task. We have used Softmax Function on the outer layer of network as the outputs are exclusive. The input to the Softmax function is a vector of real values and the output is a vector of the same length with values that sum to 1.0 like probabilities.

The error function used in this assignment is categorical cross entropy as it is classification task, it is essential to convert the labels into dummy variable i.e. one hot encoder 1 for true 0 for rest of others. The model created is by using one hidden layer. The first line creates a Sequential model.

This is the simplest kind of Keras model, for neural networks that are just composed of a single stack of layers, connected sequentially. The model is then compiled, optimized and trained and evaluated. The train loss and validation loss (validation is part of train model which is used for better generalization on test set) as well as train accuracy and validation accuracy calculated gives the performance of the model. If the train loss is more than the validation loss the model can be over fitting it can be solved by using proper regularization, or by increasing the data for training purpose. The model in this assignment can be improved by using more complex model.