Algorithm

- Build a Feed Forward neural network with configurable hidden layers. All the layers will have some 'x' Neurons each.
- The input layer matrix has no of rows equal to the size of the data set, and no of columns as the columns in the data set. The first hidden layer has no. of rows as the number of columns in the data set and any number of columns, while the last hidden layer has to have no of columns as the number of output classes.
- The internal layers may use any function, for example, Sigmoid, ReLU, tanh etc. The output layer uses softmax function .
- To compute the error in the last step of forward feed, cross entropy function is used.
- For backward propagation, to find out the weight and bias gradient, for all the internal hidden layers, we use the derivative of the function that we used.
- Repeat the whole process to compute no. of costs to get the minimum cost (Error), so as to use it for predicting test data.

- **Variation of Cost vs iteration graph with Number of Hidden Layers**

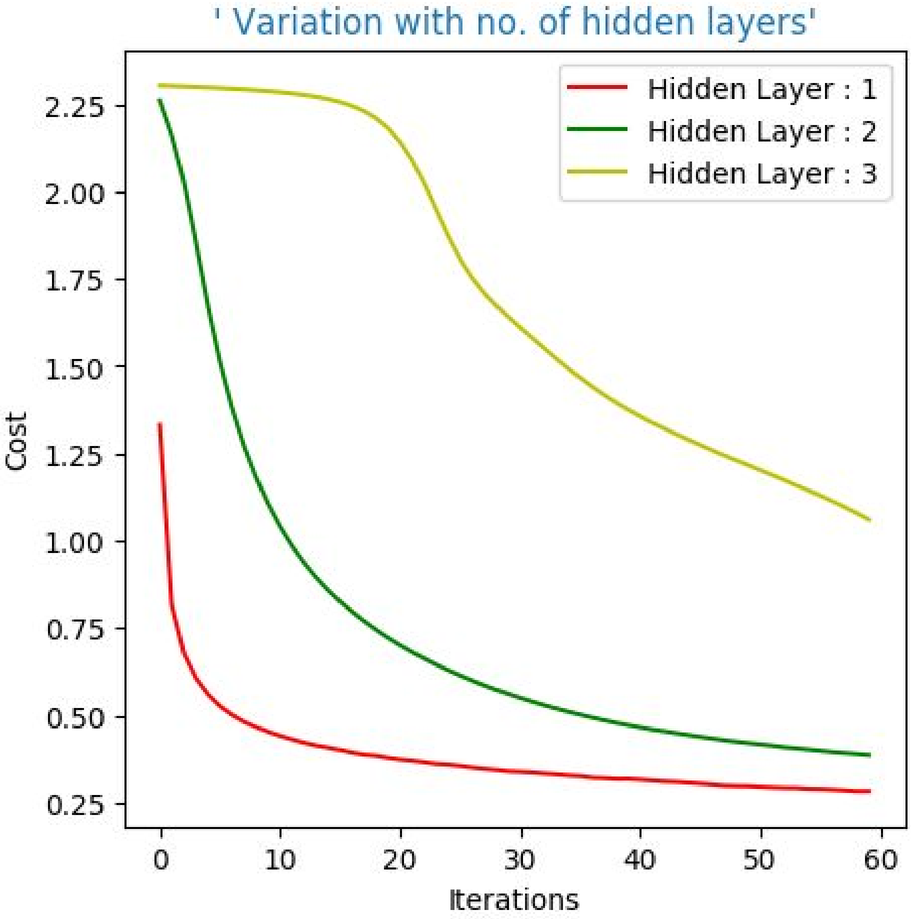
FUNCTION	LEARNING RATE	HIDDEN LAYER	ACCURACY	NEURONS	BATCH SIZE
Sigmoid	0.1	1	86.56 %	500	500
Sigmoid	0.1	2	84.77 %	500	500
Sigmoid	0.1	3	77.56 %	500	500
ReLU	0.01	1	84.56 %	500	500
ReLU	0.01	2	88.66 %	500	500
ReLU	0.01	3	86.22 %	500	500
tanh	0.01	1	84.28 %	500	500
tanh	0.01	2	82.22 %	500	500
tanh	0.01	3	79.25 %	500	500

Sigmoid Function : In case of Sigmoid Function, increasing the hidden layers, the accuracy decreases. Also, changing the learning rate is not creating much difference in the accuracy, though we can still find that the best result is obtained for learning rate 0.1, and number of hidden layers 1.
(It's not the best model though)

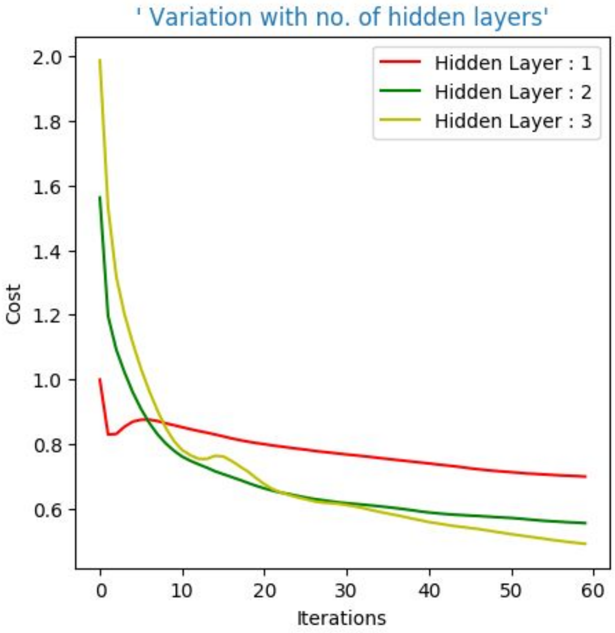
ReLU Function : The best result for ReLU is obtained when learning rate was 0.01, as for bigger learning rates the cost converges very fast giving us a bad accuracy. Also, the best result is obtained for 2 number of layers. This was observed to be the best model.

Tanh Function : The best case for tanh also is, with learning rate 0.01 and no. hidden layers 1.

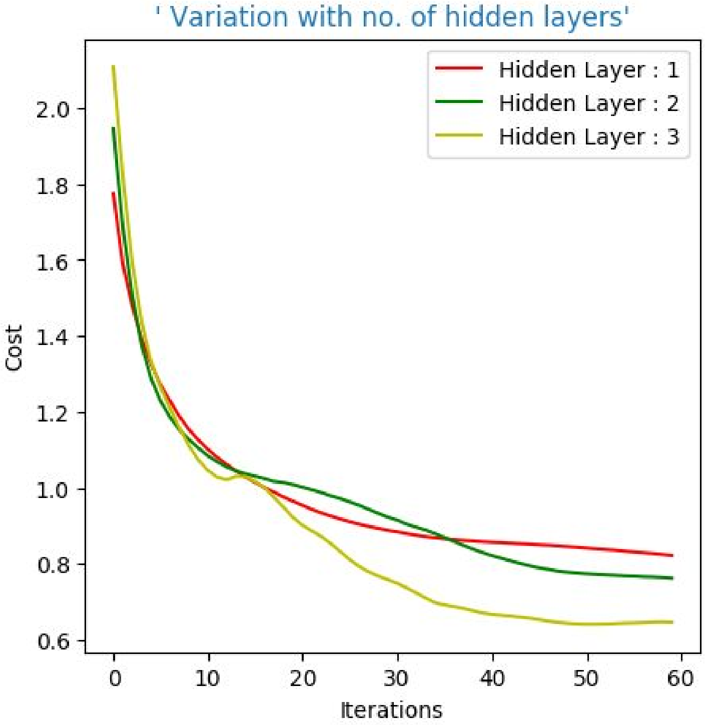
With increase in size of batches, or decreasing the size of neurons in a hidden layers, the accuracy decreases, so the best chosen batch size, and number of neurons is 500.
Below are the graphs , representing the behaviour of all the functions, with change in number of hidden layers



<u>ReLU</u>



Sigmoid



tanh