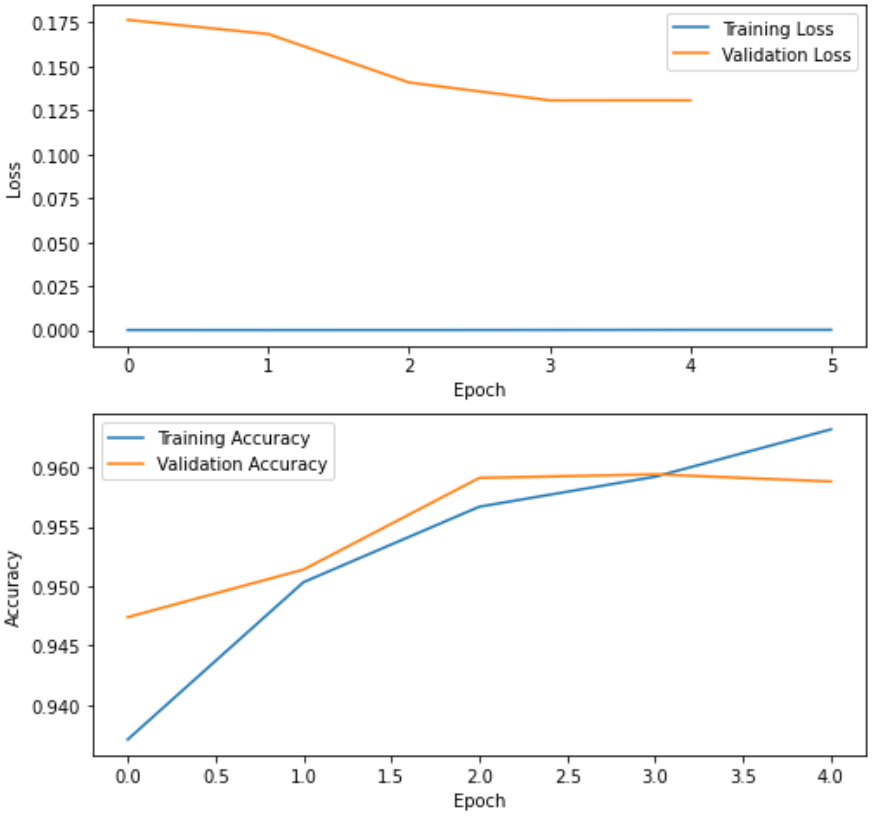
LAB ASSIGNMENT 9
Lab Report
Shubh Goyal (B21CS073)

QUESTION 1-

MNIST dataset was loaded and trained on a neural network for 5 epochs.
The neural network had 2 hidden layers. The first hidden layer had 128 neurons and the second hidden layers had 64 neurons. Both the layers had 'ReLu' activation function.
The total number of trainable parameters was 109,386.

The following are the graphs of loss vs epoch and accuracy vs epoch for both training and validation data.
Train loss is 0 on the graph because it is of order 10^(-4). The validation loss first decreased and then became consistent. Also the training accuracy consistently increased for the training set while it first increased then decreased for the validation set.

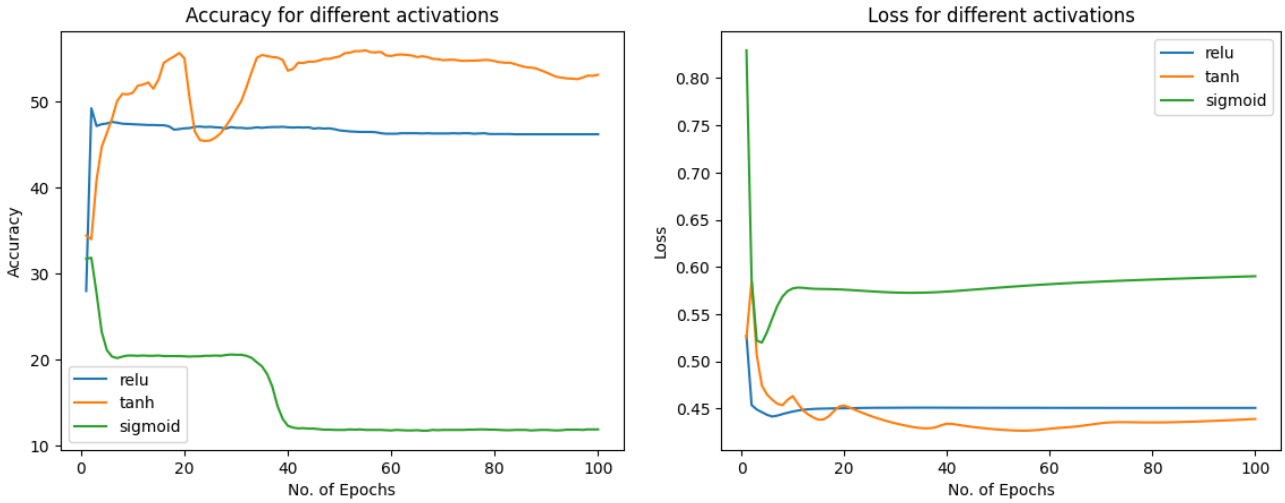


QUESTION 2-

Divided the dataset into three categories based on rings as given in the abalone.names file. Rings 1-8 corresponding to the first category, rings 9 and 10 to the second category and the rest belonging to the third category.

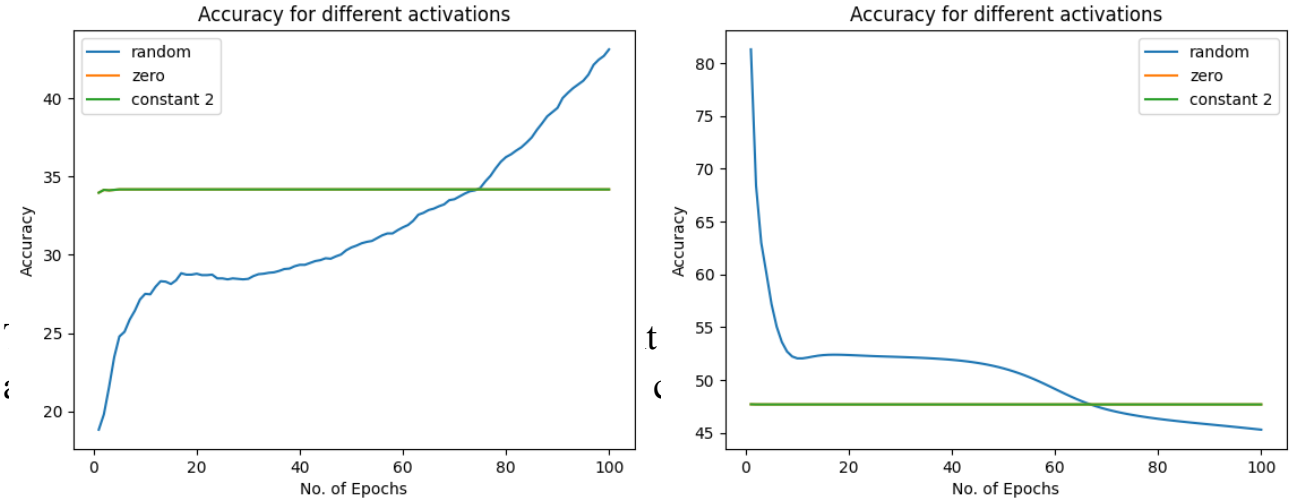
A class named NN() was created to implement an artificial neural network from scratch with the functionalities given as tasks.

Task 3: The highest accuracy was obtained with tanh activation function.



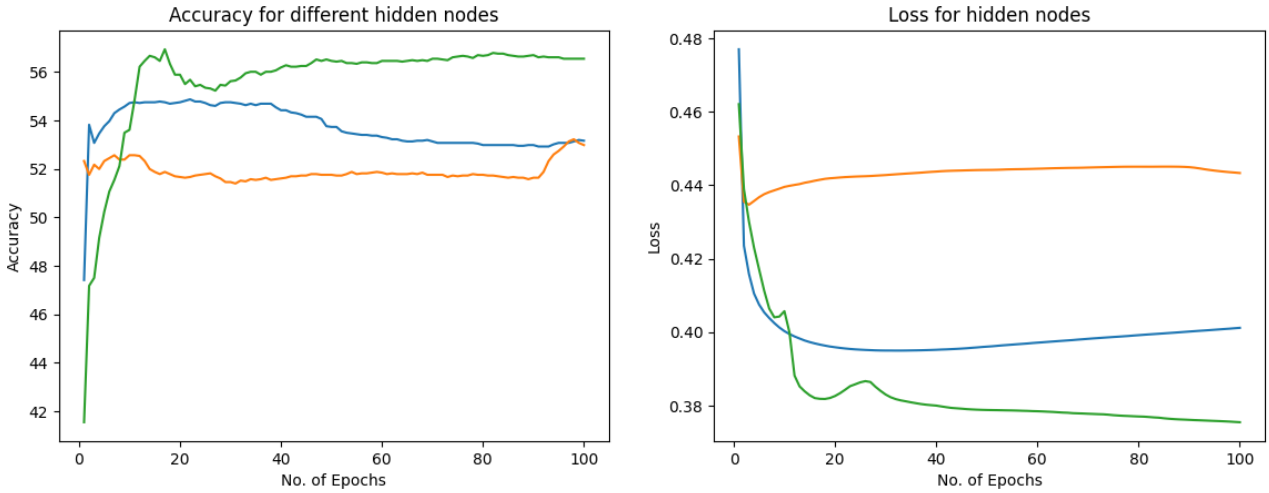
In case of sigmoid the accuracy decreased continuously while for relu the accuracy reached an almost constant value from the initial epochs itself.

Task 4:



training accuracy increased continuously and would have increased further on more epochs. The loss for random initialization is also decreasing continuously.

Task 5:



The model that fitted best had two hidden layers with 8 and 4 nodes respectively (represented by green line). Orange line represents model with one hidden layer having 8 nodes and blue line represents model having 2 hidden layers with 8 and 5 nodes respectively.

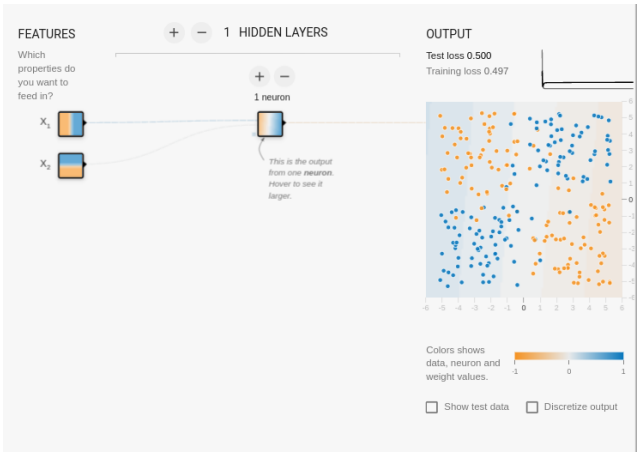
The second model's loss became constant at a comparatively high value and it did not converge further. For the other two models, the loss seems to have converged and the accuracies have also become almost constant.

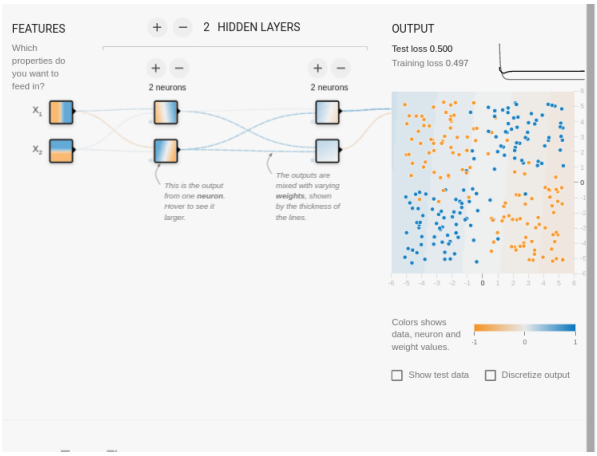
QUESTION 3-

Task A:

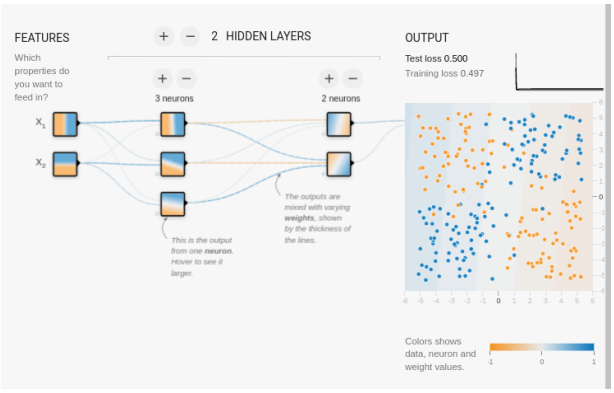
a. If the model is underfitting then increasing the size of the network will help but if the model is working fine then it may start overfitting on increase in size and will lead to a bad model. Change in convergence on increasing size depends on factors such as the type of dataset being used and also the number of training examples. Therefore, it is variable and cannot be said to only fasten or slow down the convergence in all cases.

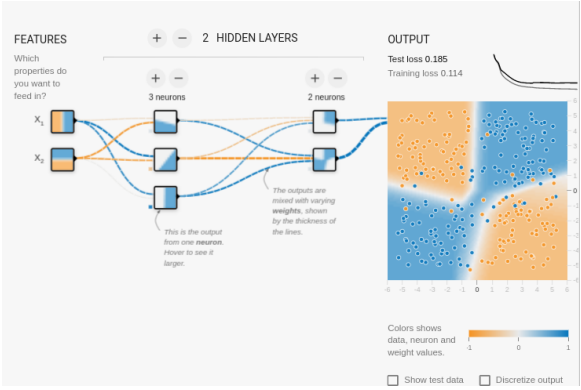
b. Each model converged in around 80-120 steps. The shape of convergence was similar in all the trials but in some trials the error value increased slightly after achieving minimum while in some it remained on an almost constant trend. Added one more layer with a single node and also added an extra node to the pre-existing hidden layer. In this case, the train loss decreased quickly at first and then remained constant while the test loss decreased a little bit and after hitting a minimum (which was still significantly more than the train loss at that epoch), increased significantly.





c. The best model was obtained with two hidden layers (3 neurons in first layer and 2 neurons in second layer). Activation was kept ReLU, the model was working well with tanh too but minimum loss was obtained with ReLU. Also regularization was kept to be L1. The minimum loss obtained was 0.119 for the train and 0.162 for the test. The test loss was around 0.01 higher along with the difference in loss in train and test being higher in case of L2 regularization and so L1 was chosen. The training surface had blunt edges.





Task B:

The model was more prone to changes at higher learning rates and the decision boundary changed very quickly with a number of iterations. But in the case of smaller learning rate the decision boundary changed very slowly and with learning rate less than 0.1 it seemed to be almost constant initially. Thus the training stability decreased with increase in learning rate as the model became more prone to changes with given inputs.