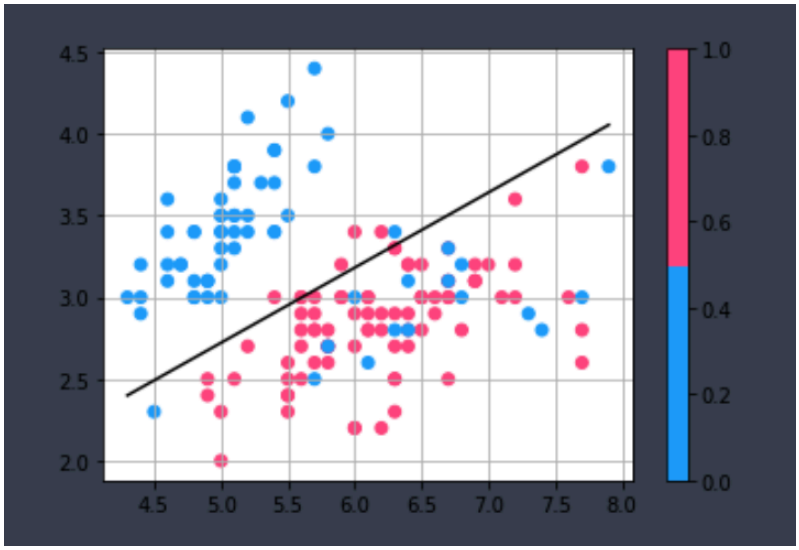
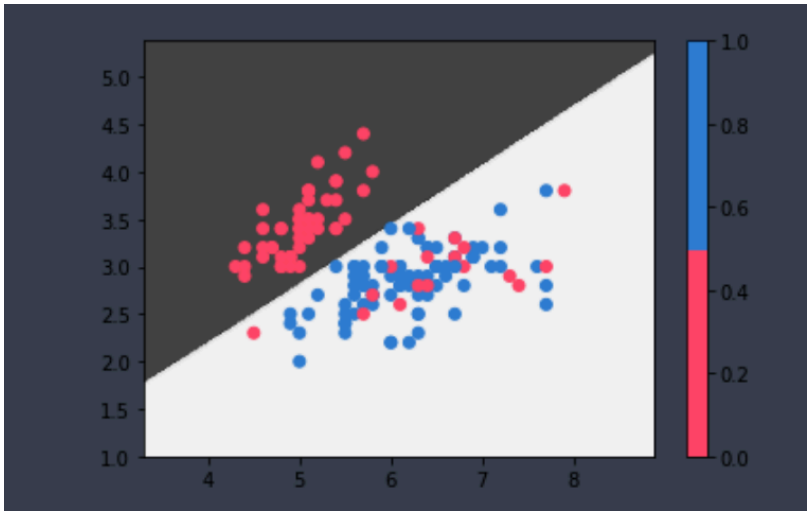
CS-1390 Assignment 3
NEERAJ PANDEY

Solution 1(a):

Test Error: 11.33% (11%-13%)
Code: ques1a.ipynb



Solution 1(b):
Test Error: 9.4% (6%-12%)
Code: ques1b.ipynb



Solution 1(c):
We can see that the test error for multi layered perceptron is usually less than the single layered perceptron. This is because a multi layered perceptron contains one or more hidden layers which is apart from the input and the output layers. Multi layered perceptrons includes hidden layers and can work with non-linear functions and also with quite complex relationships in a model. So, the hidden layers filters out some of the information from the inputs and passes onto the next layers In the model. This also improves the performance and makes the model network faster as it doesn't consider the outliers and work with the important information.

Solution 2(i):
When weights are initialised with the same values, each hidden unit will get exactly the same value and will yield same valued weights. This means that when the weights are initialised with the same value, all the units will receive same value which is equal to the sum of inputs. So, this will result in same values in the hidden layers as well. In other terms, this would make that the hidden layer units becomes symmetric. Since, the weights in all the layers are same, the multi layered perceptron is no better than the single layered perceptron.

Solution 2(ii):
When weights are initialised with very large/high values then the dot product of the weight and the train data (np.dot(W, X)) becomes significantly higher because of which, when an sigmoid activation function is applied on the model, the function maps its value to nearly 1, due to this saturation problem, the gradient changes slowly and increases the learning time (vanishing gradient problem).