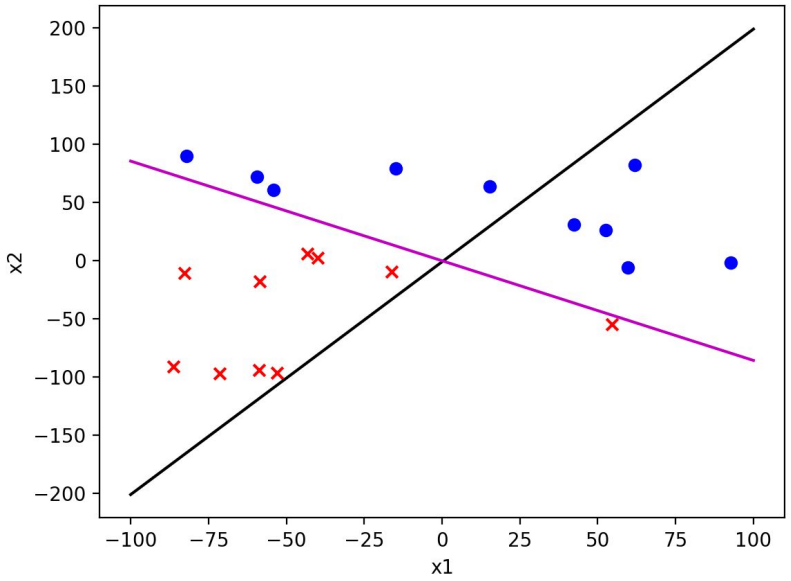
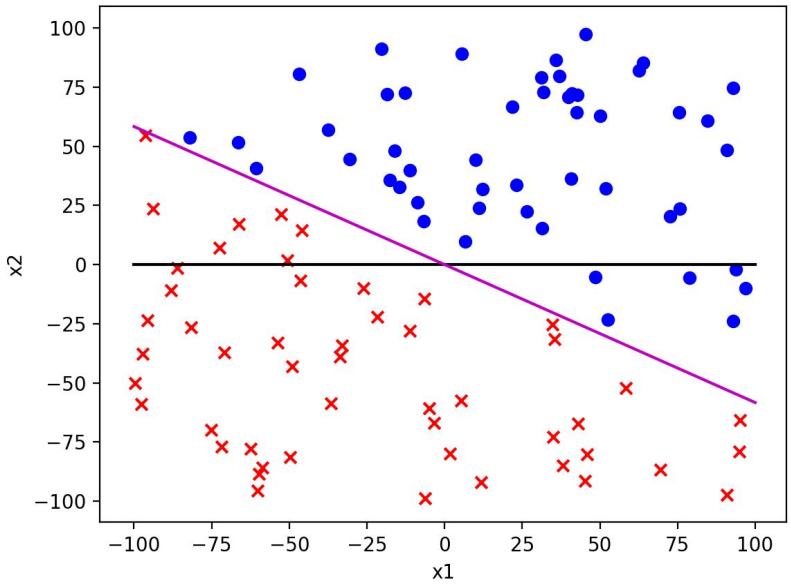
EE5434 – Machine Learning for Signal Processing Applications

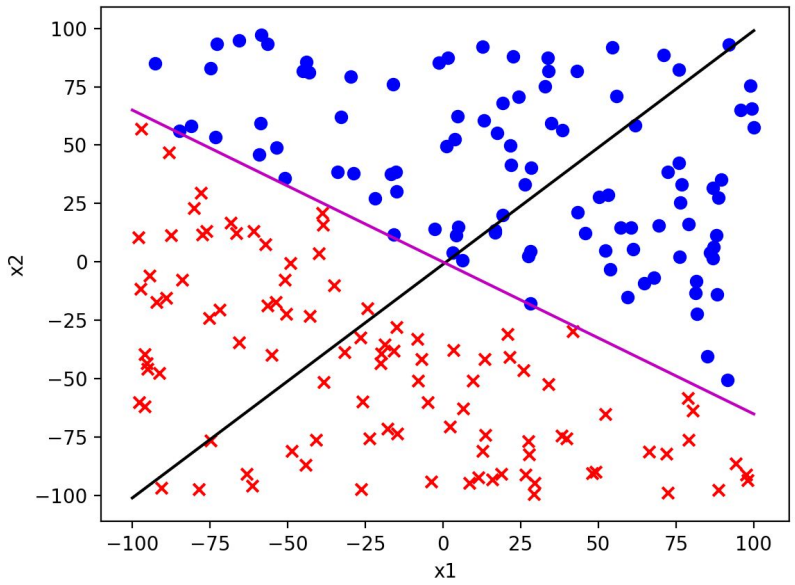
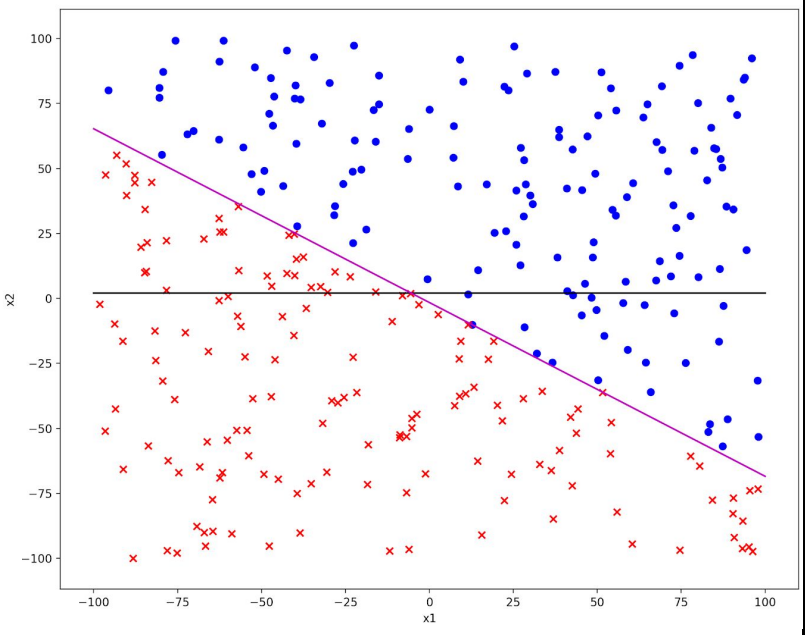
Assignment 1 – Due date: September 23, 2019

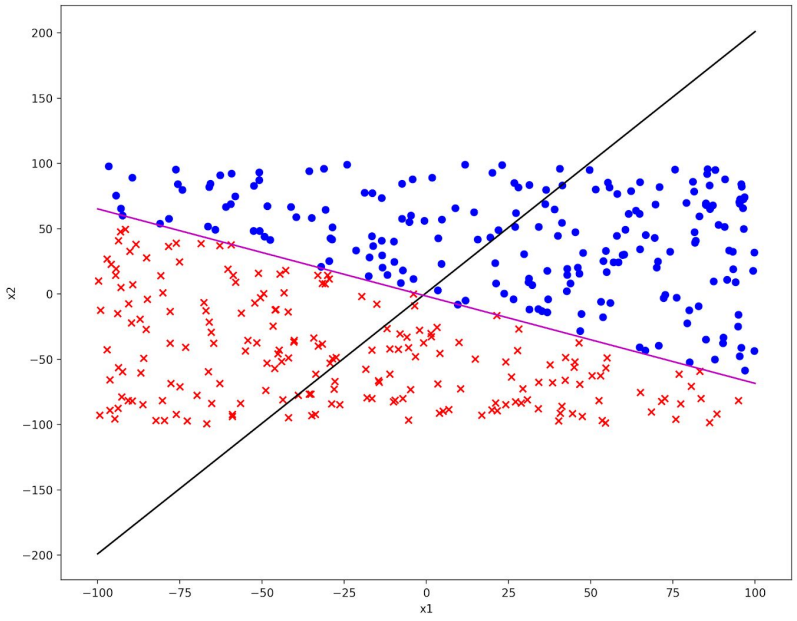
Report

Name: _____Sin Bo Chi_____ Student ID: ___54412731____

Part 1: Designated vector w

DataEmit [5,2,3] 10 10	
DataEmit [5,2,3] 50 50	

DataEmit [5,2,3] 100 100	
DataEmit [5,2,3] 150 150	

DataEmit [5,2,3] 200 200	

In the above figures, the black line is the randomly generated line before running PLA algorithm, and the purple line is the known line after running PLA algorithm. From the figures, we can know that PLA algorithm adjusts the black line into purple line, which is a line to divide all positive and negative data points. According to the figures, if the size of the training data increases, the running time of program becomes longer, but the accuracy of PLA line becomes higher. In the figure of DataEmit [5,2,3] 10 10, even there are few data points, the purple line touches one of the negative points. In the figure of DataEmit [5,2,3] 200 200, even there are in total 400 data points, the purple line can find a suitable way to divide positive and negative data points.

Part 2: Chosen own vector w

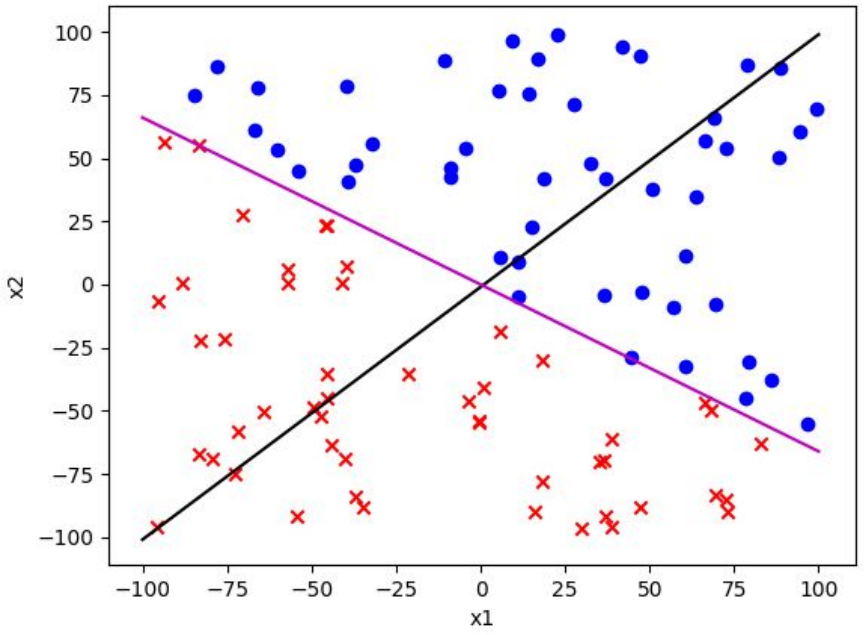
The vector w = <2,4,6> is chosen and 2 tests are implemented:

1. Increase the training data
2. Change ratio of the two labels in training data

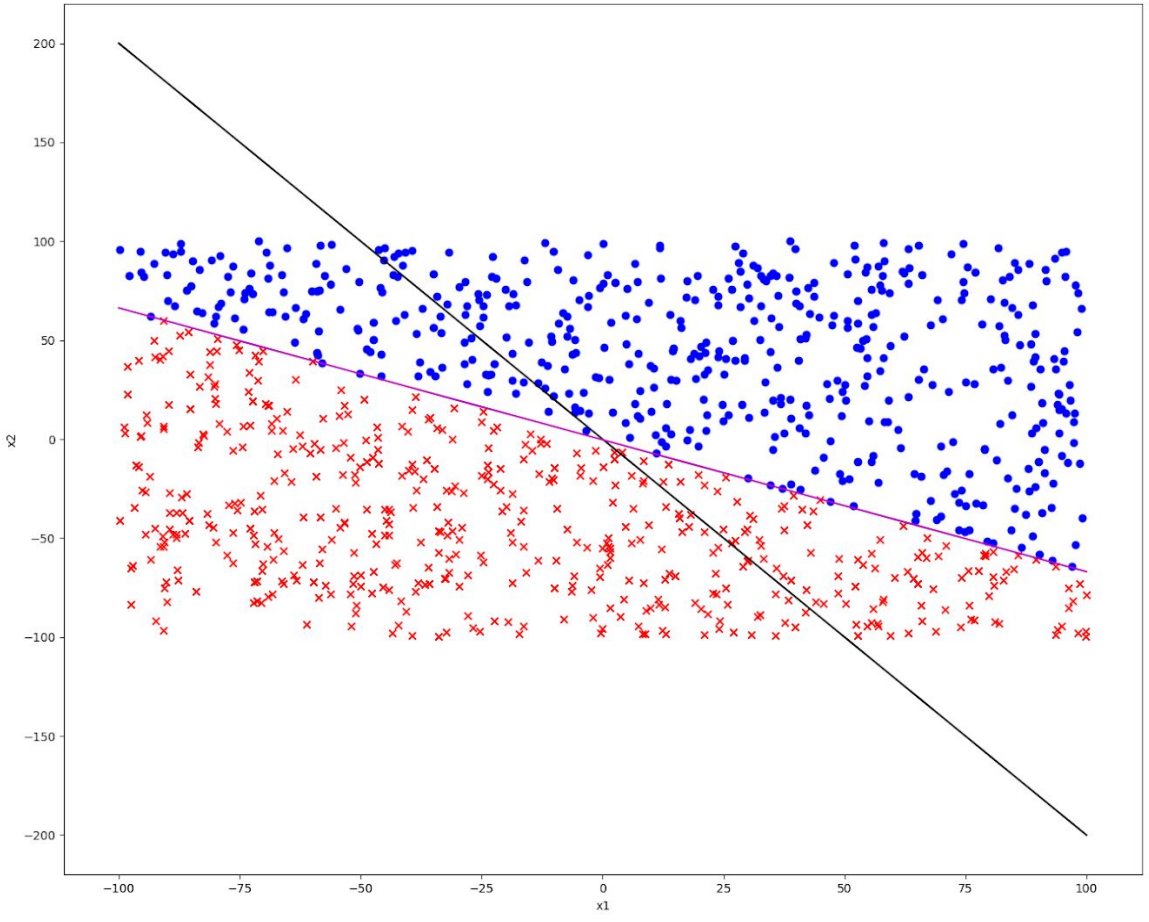
1) Increase the training data

In this experiment, two sets of data are generated by DataEmit program. The first set of data includes 50 positive and 50 negative data points, and the second set of data includes 500 positive and 500 negative data points. The results are summarized below:

50 positive and 50 negative data points

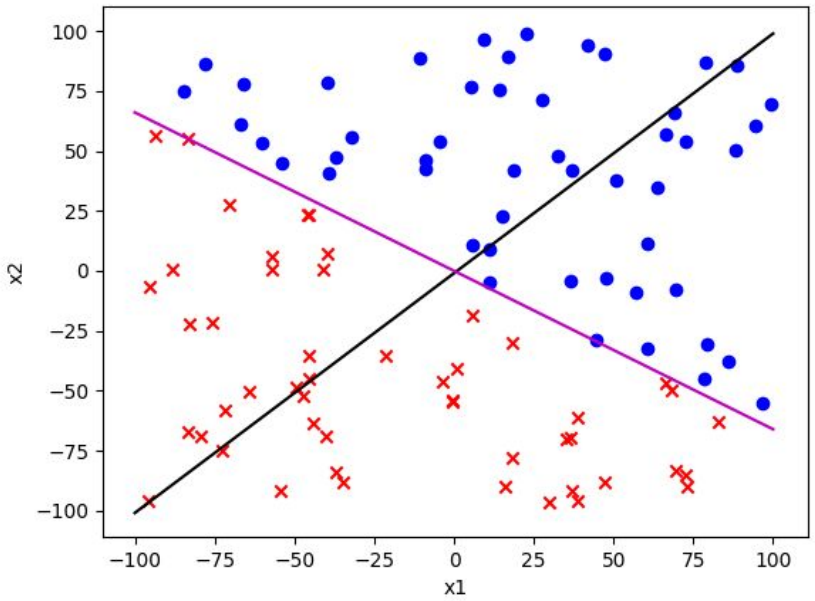
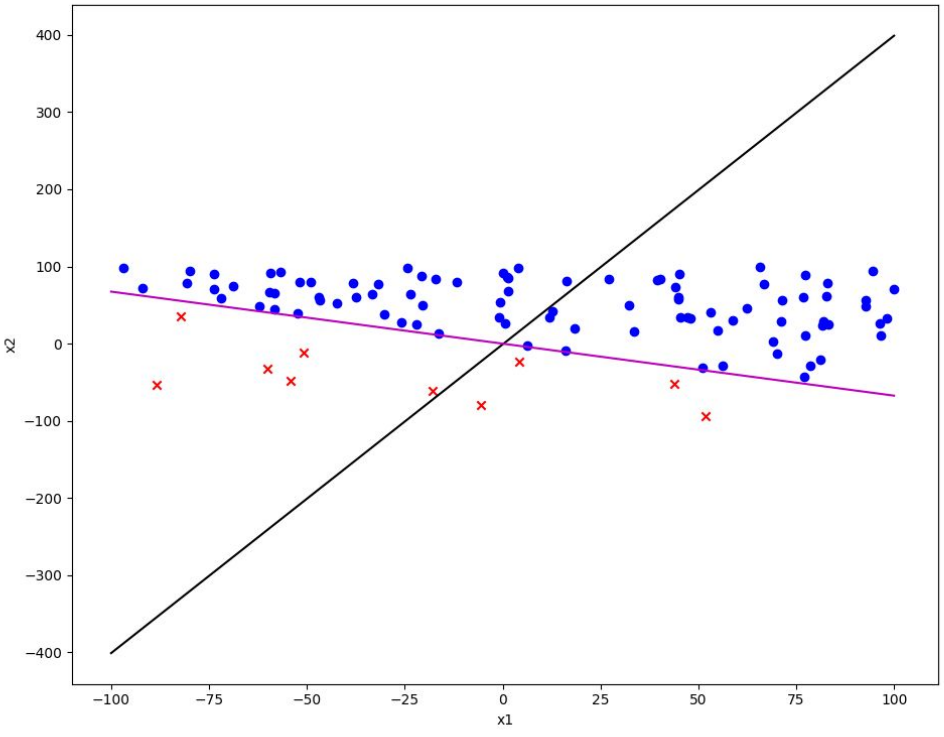


500 positive and 500 negative data points



2) Change ratio of the two labels in training data

In this experiment, two sets of data are generated by DataEmit program. The first set of data includes 50 positive and 50 negative data points, and the second set of data includes 90 positive and 10 negative data points. The results are summarized below:

50 positive and 50 negative data points

90 positive and 10 negative data points


To conclude, since all the data points are linearly separated by the equations $(w1x1+w2x2+w0)>0$ and $(w1x1+w2x2+w0)<0$, both (1) increase the training data and (2) change ratio of the two labels in training data do not affect the accuracy of the PLA line (purple line) by the program.