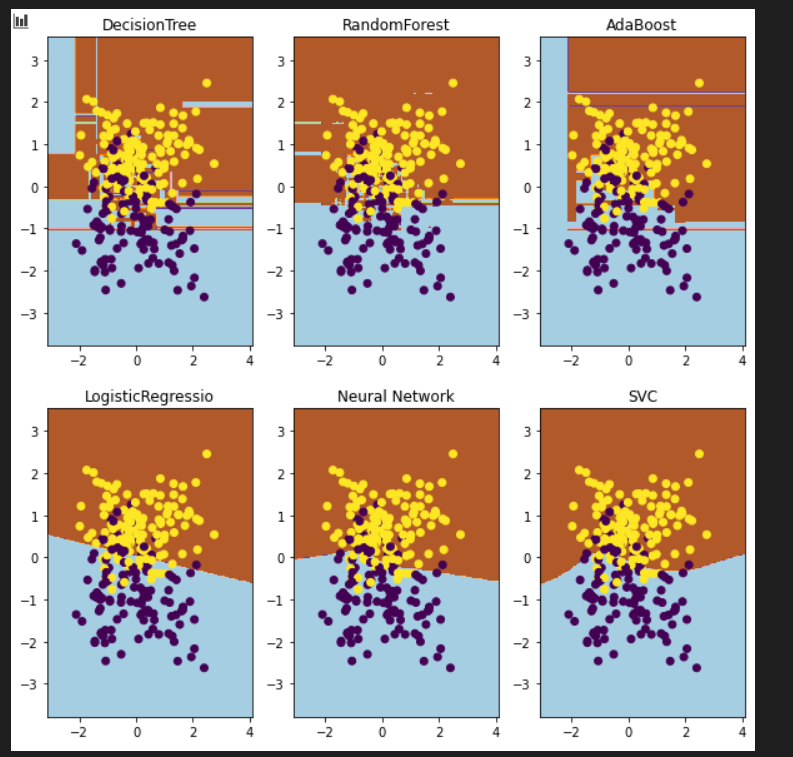
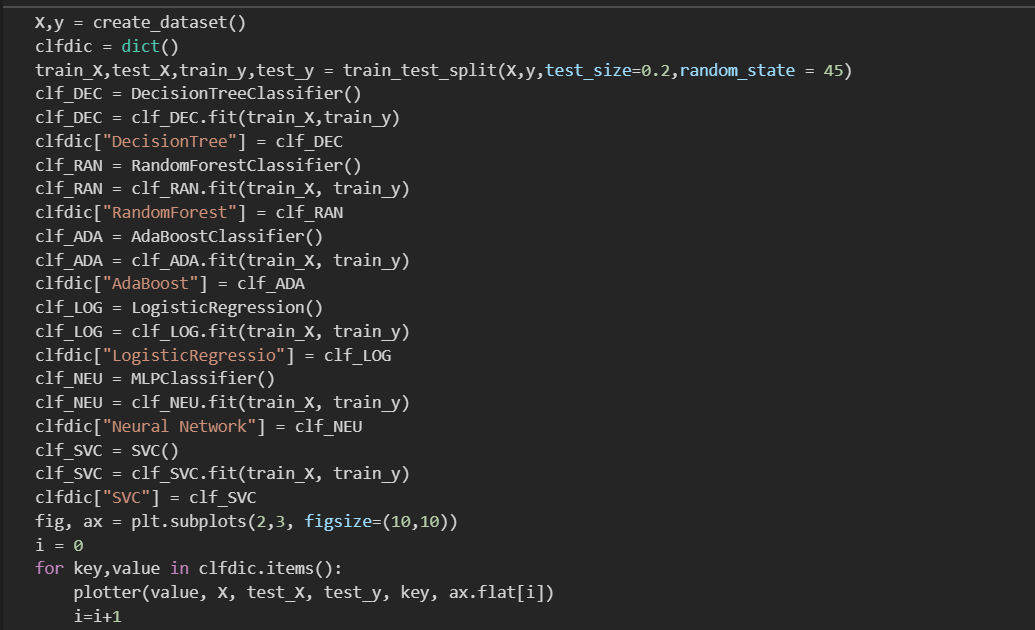
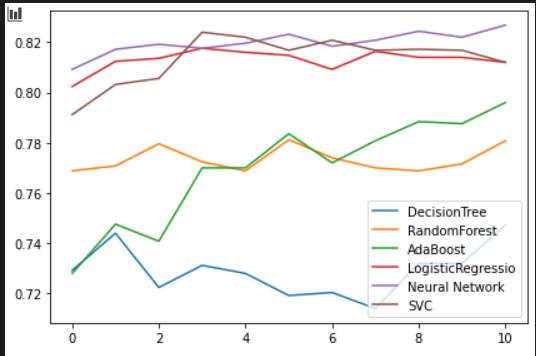
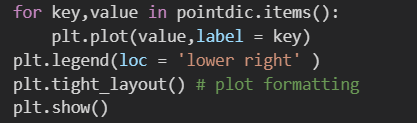
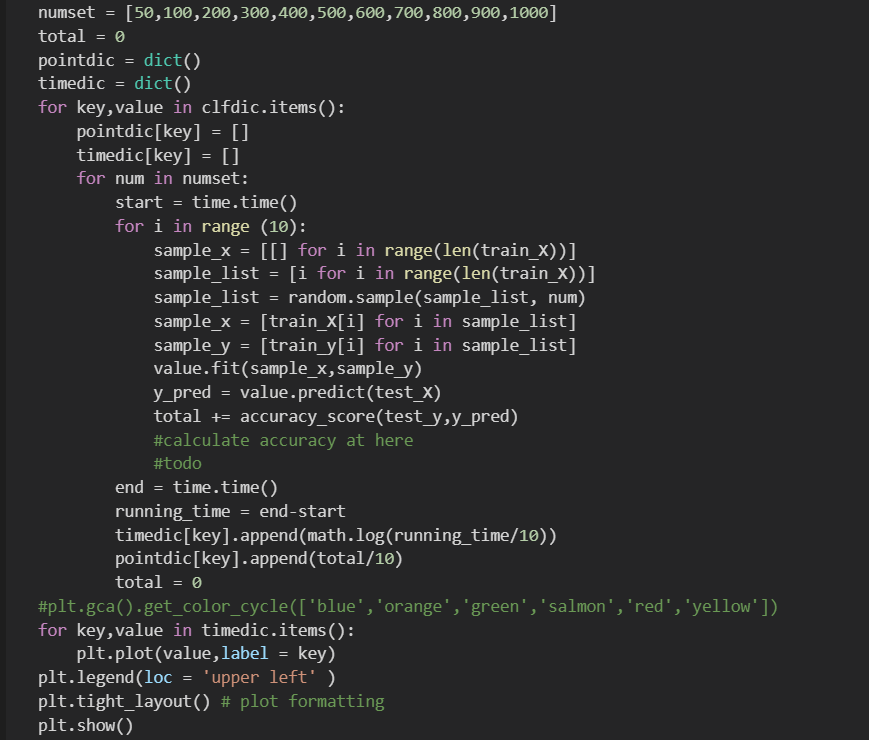
Q1

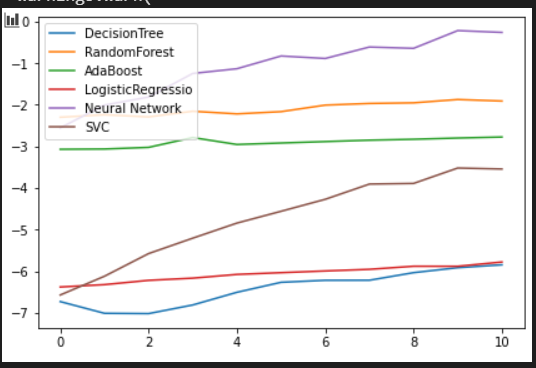
(a)





(b&&c)



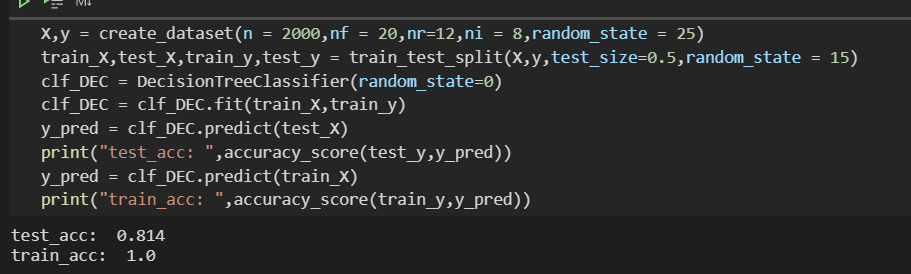
Firstly, we can find Decision tree is the simplest classifier which take shortest time and lowest accuracy.

Logistic Regression, Neural Network, random forest are more accurate than the others because base on bias-variance decomposition The less volatility, the better classifier and Neural Network takes the longest time in those classifier ,accuracy of random forest it too low, so logistic Regression takes not that much time which I think are the better classifiers

The accuracy of AdaBoost increase if the size of training set increase, the others wont be that much affected by change the size of training set

And also, time spend for every classifier will increase if the size of training set increase.

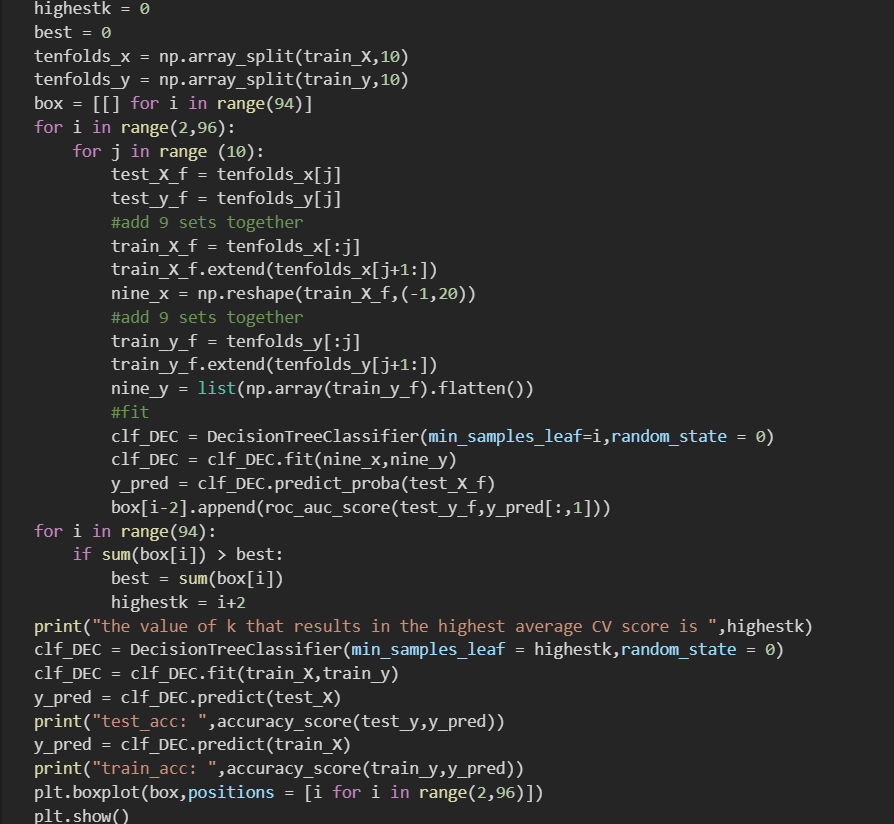
(d)

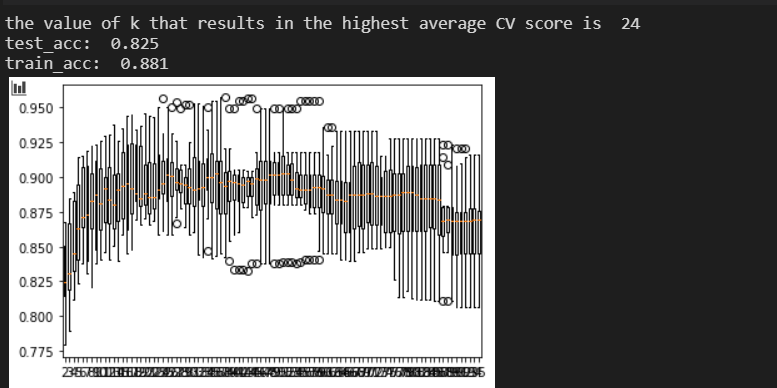


(e)

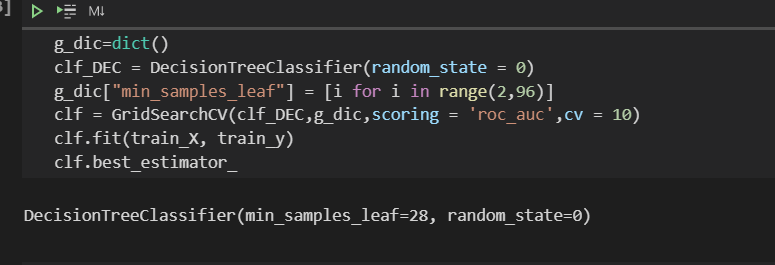


(f)



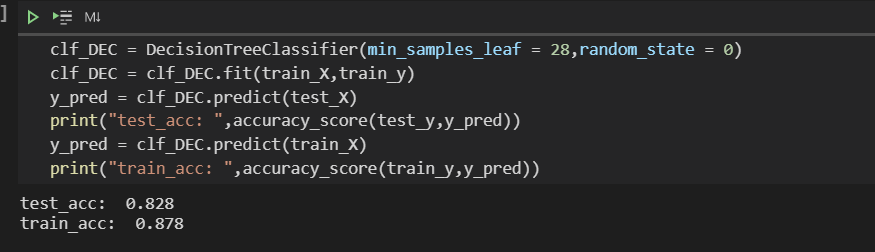


(g)



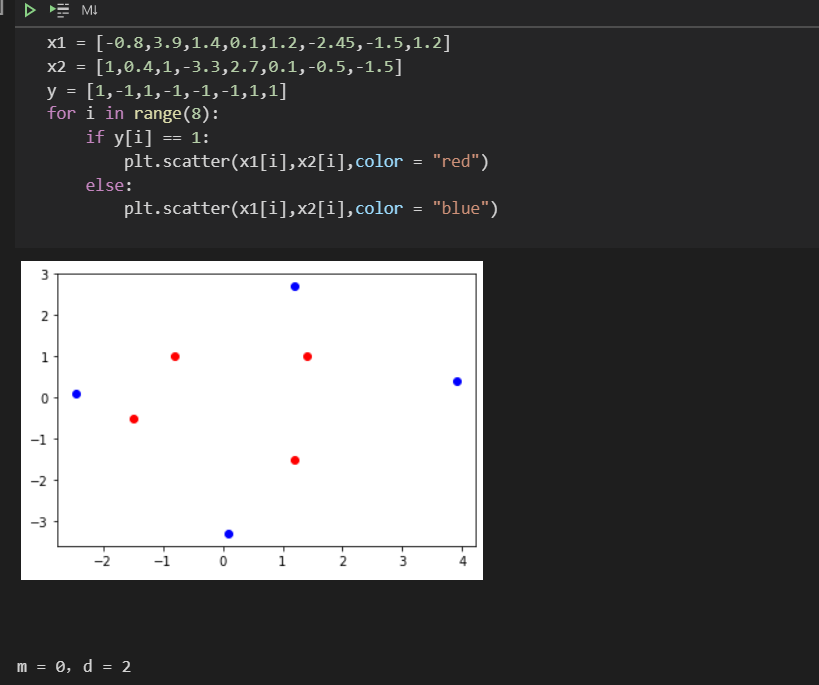
The reason why it different to mine result is we split the data in 10 and use one of each to be the test data is not strict, randomly select 100 data from the train\_data is better

.

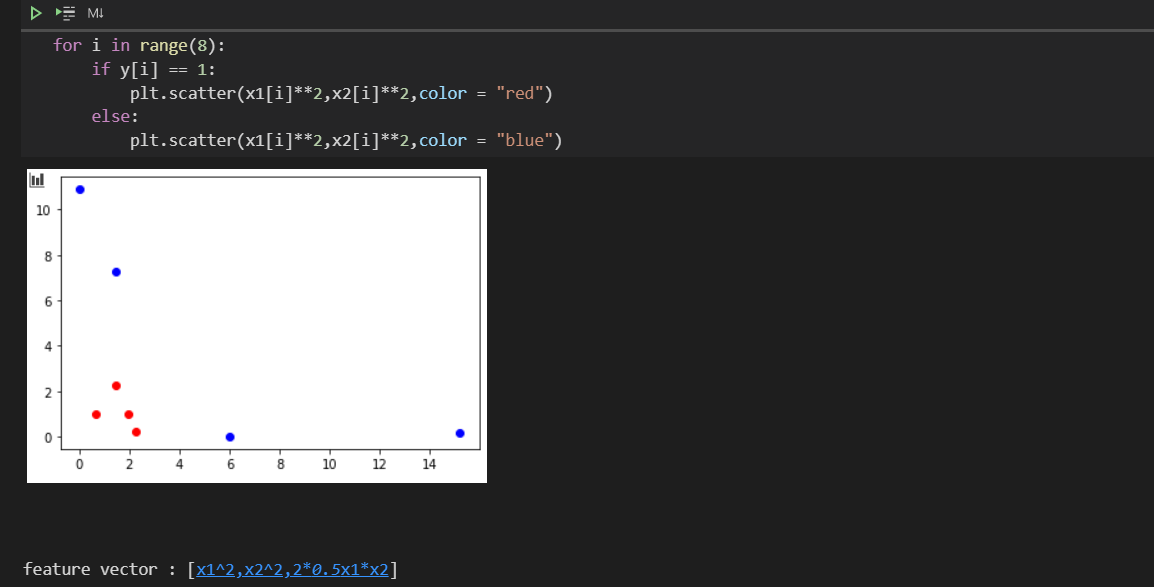


Q2

(a)

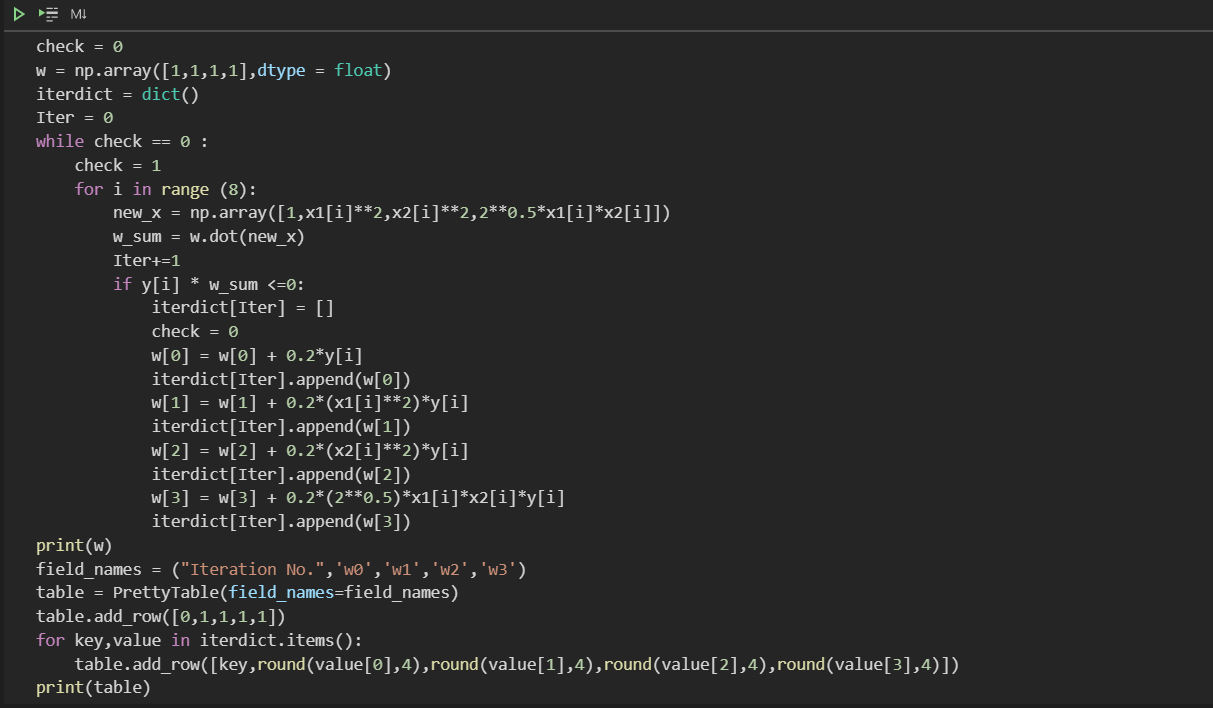


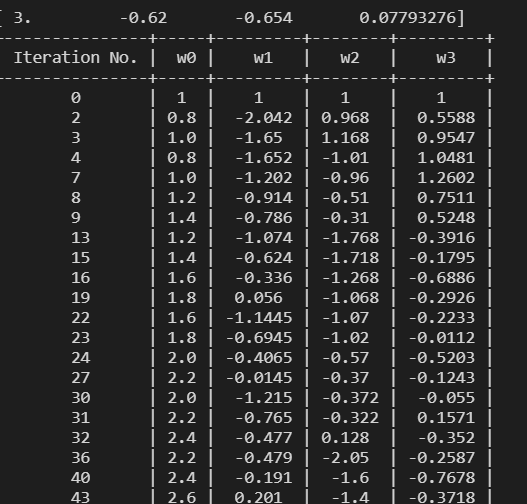
(b)

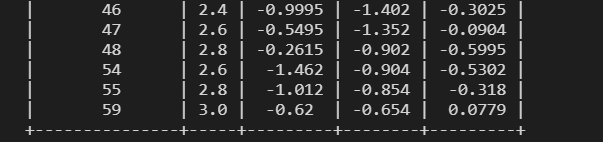


Yes it is linear separable

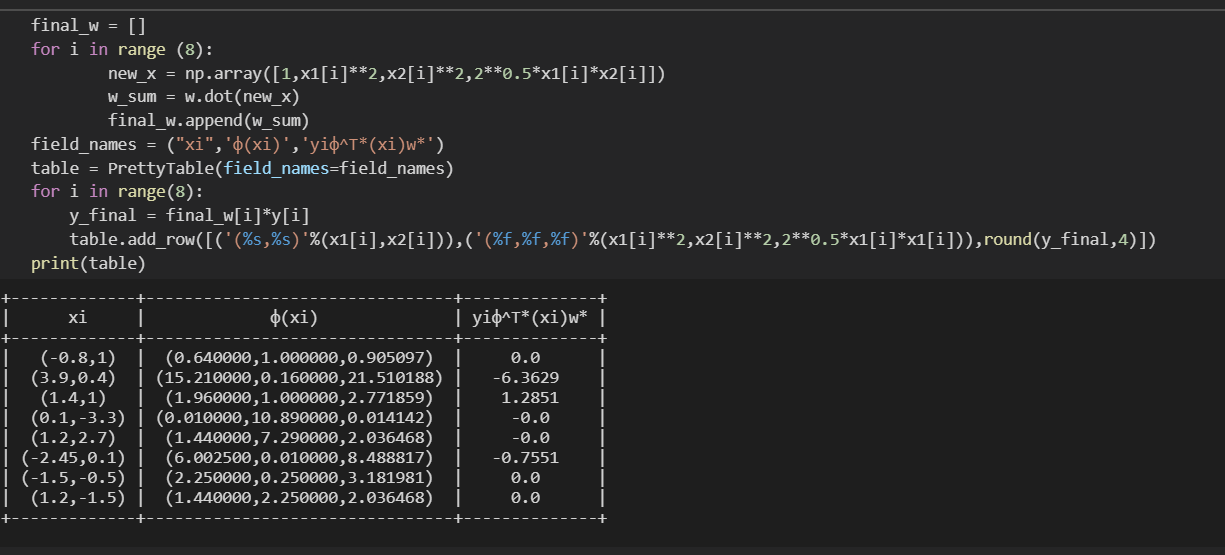
(c)



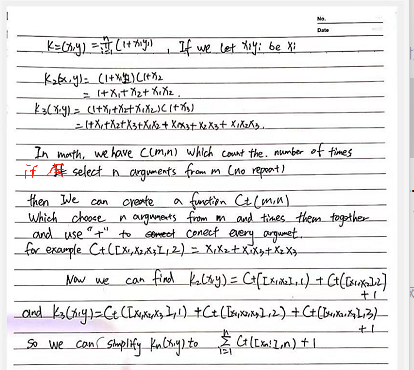




Last vector is [3,-0.62,-0.654,0.0779]



All the yi\*xi\*wi > 0 so it is converged

(d

so kn(x,y) = 1 + sum(from i = 1 to n) Ct([j for j in range(x1,xn)],n)