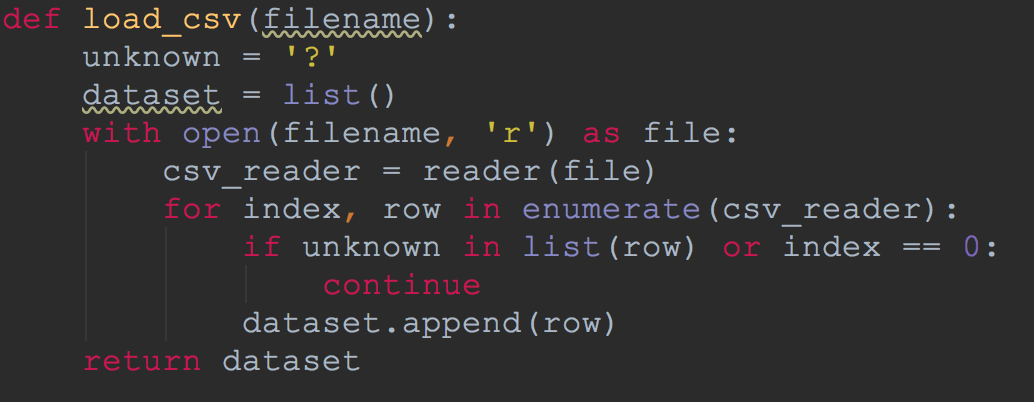
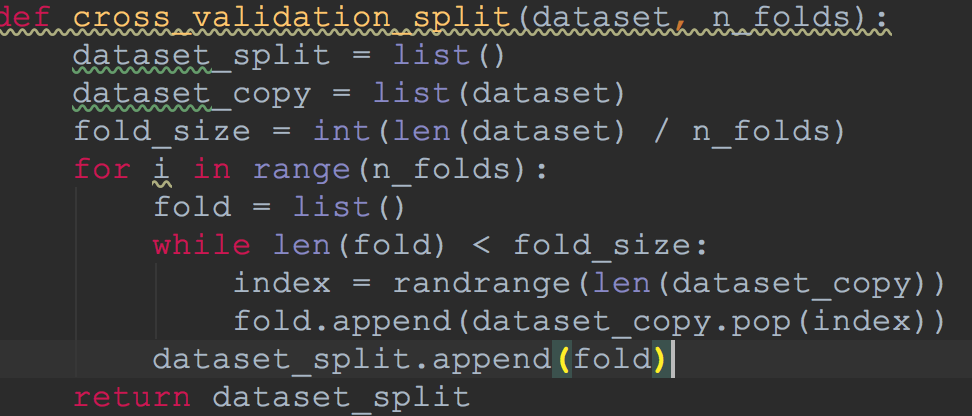
2.1



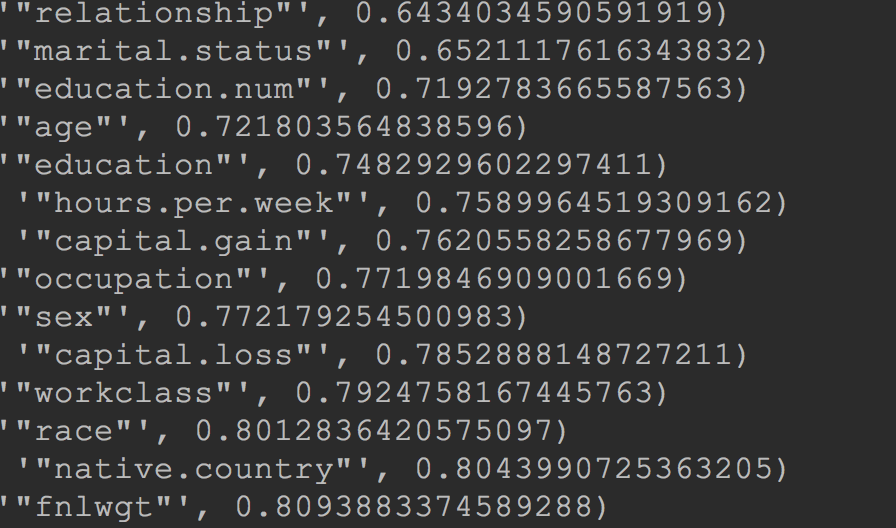
2.2

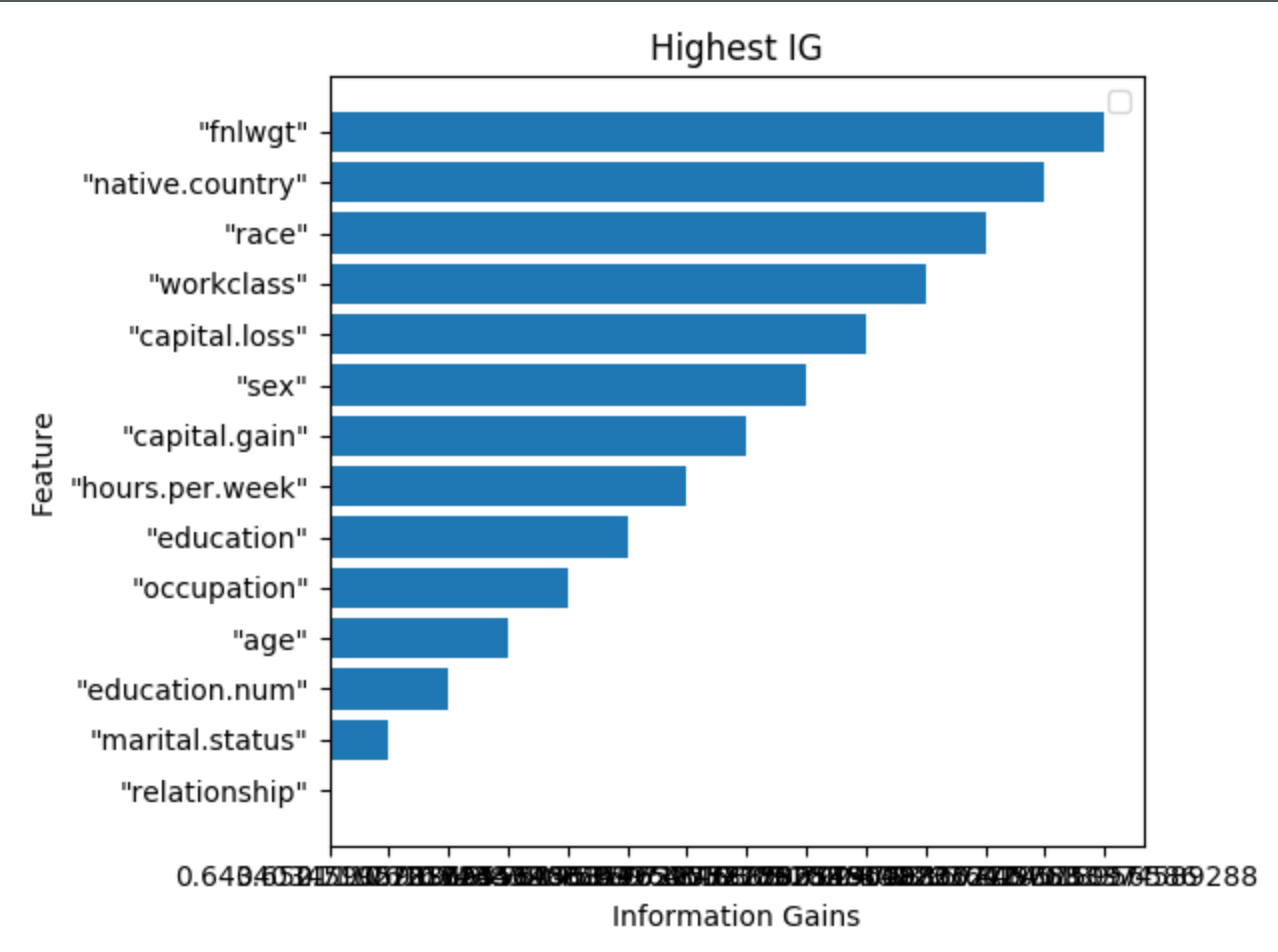
Because it has some strings in the dataset. We get the number of every category divided by the sum. Then the strings will convert to the number ranged between 0 and 1. Then normalized all the the data.

2.3



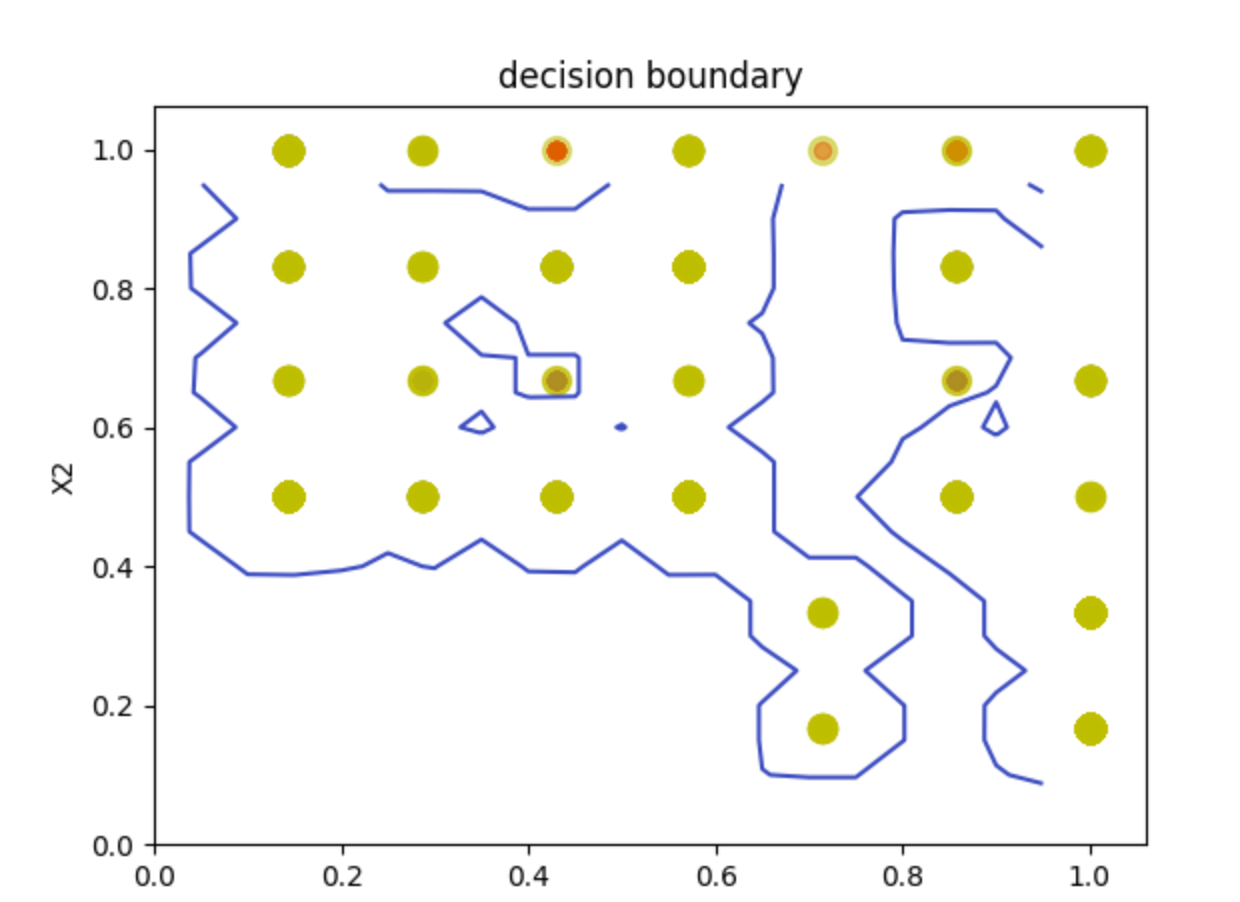
2.4





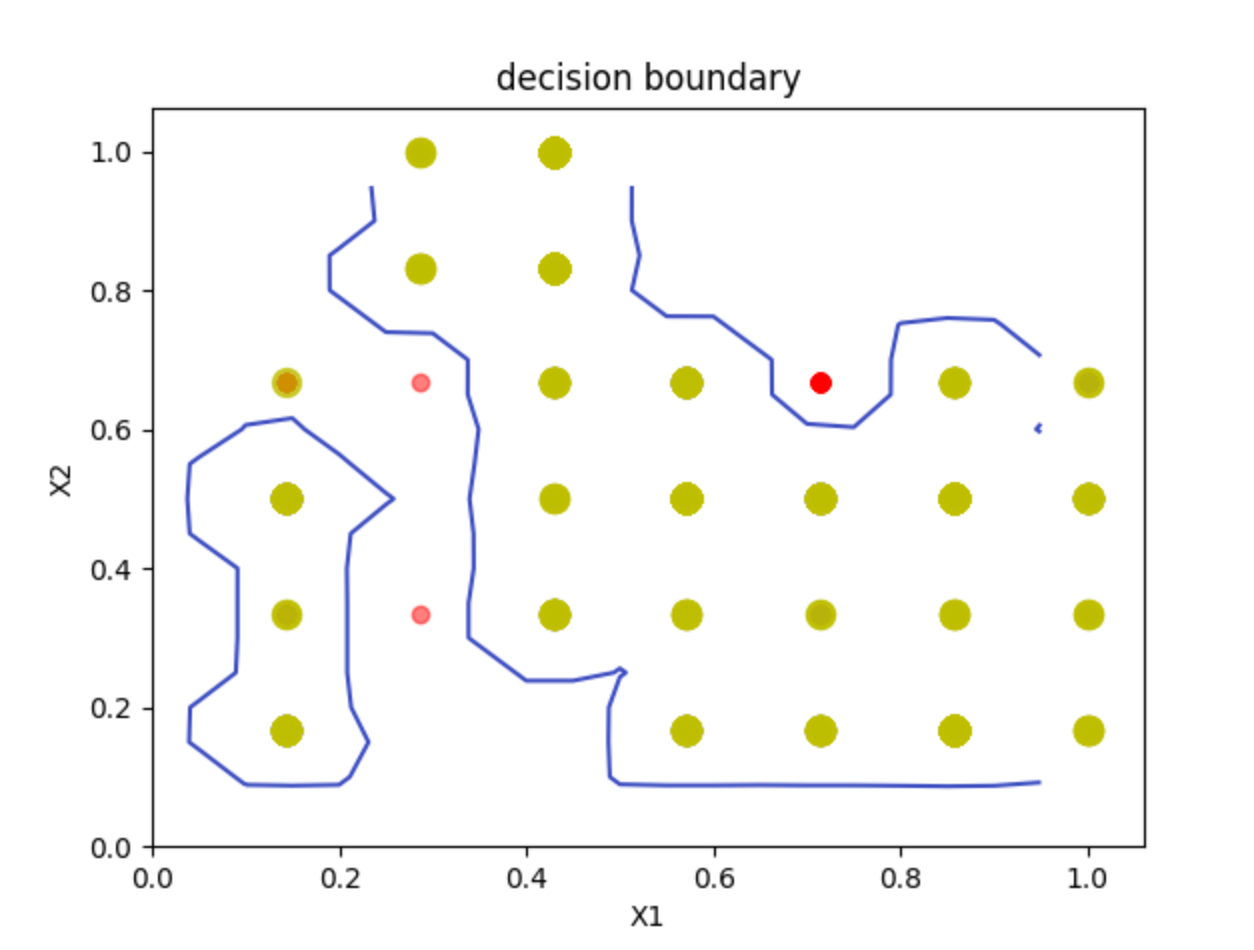
3.1

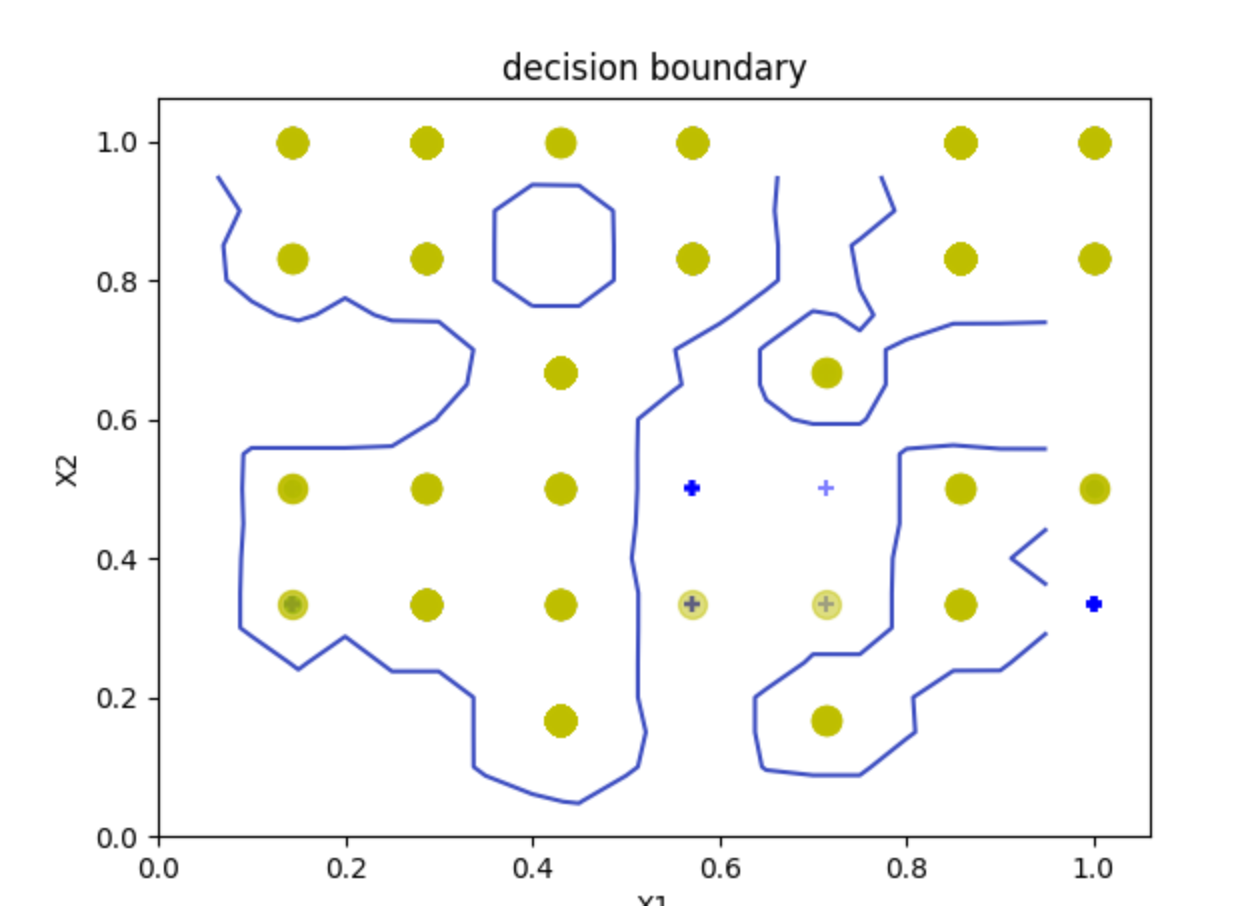
plot the second and the second reversely features as the X and Y. We plot the decision boundary.

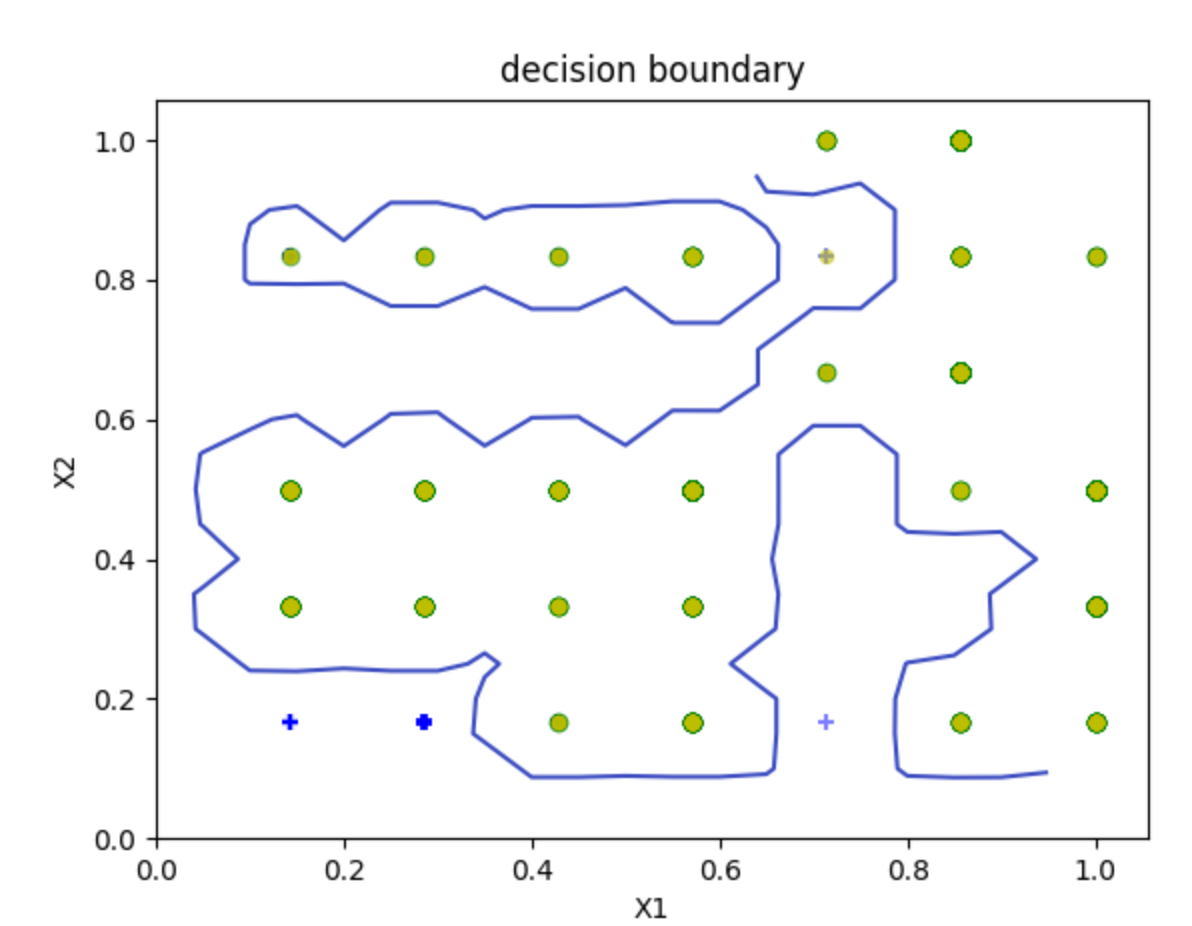


3.2

When we pick a higher C, it will have a better performance than a lower value of C. But if we pick infinite higher, it won’t make the decision boundary by SVM.

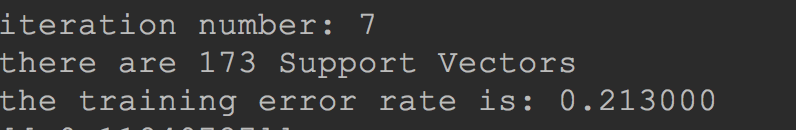






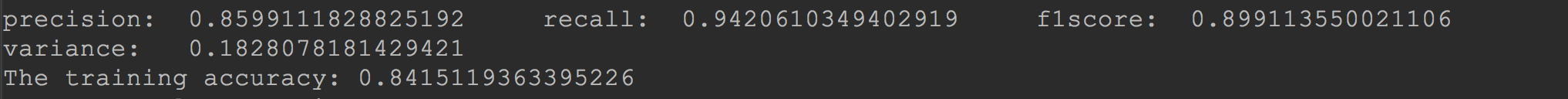
3.3

I just always change the value of C, and run the code to compare the accuracy. The best C maybe between 150 and 200.

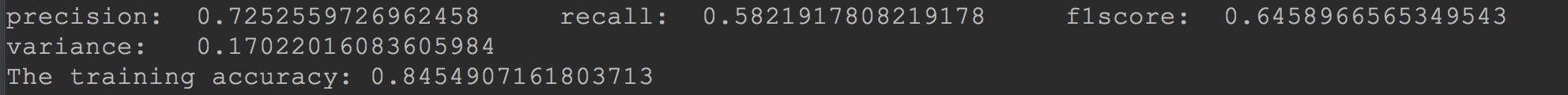


4.1

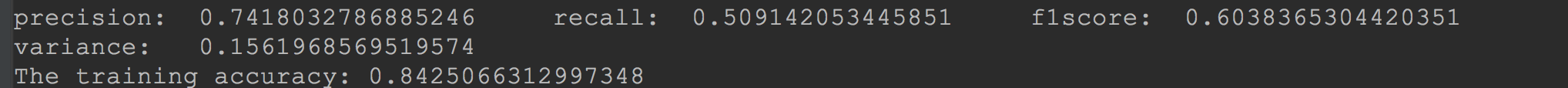
liner: C tol max\_ite



rbf: C gama recall : smop tol max\_ite



poly: C degree gama tol max\_ite



The accuracy is almost aways 0.80

4.2

add adaboosting to create weak classifiers.

