DataScience Homework 3

Logistic regression with Stochastic gradient descent

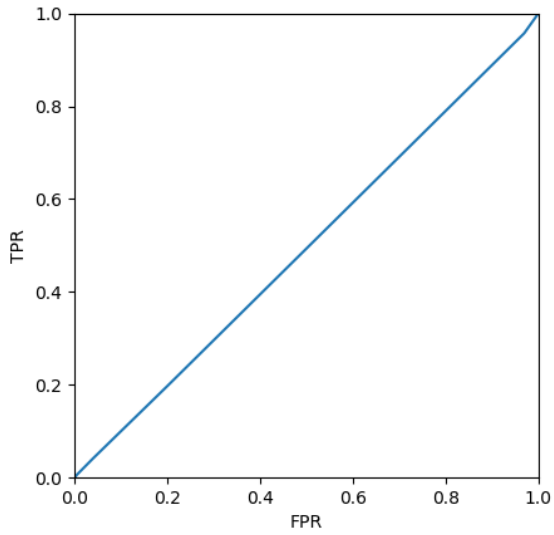
108525003 吳承儒

Before training : (theta will be 9x1 random vector)

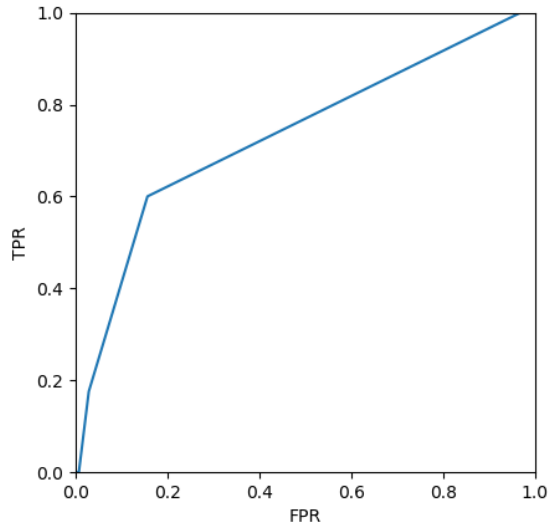
```
Logreg before-train train accuracy: 0.097329
Logreg before-train train precision: 0.097329
Logreg before-train train recall: 1.000000
Logreg before-train test accuracy: 0.085820
Logreg before-train test precision: 0.085820
Logreg before-train test recall: 1.000000
```

ROC curve :

1. Train data



2. Test data



Discussion :

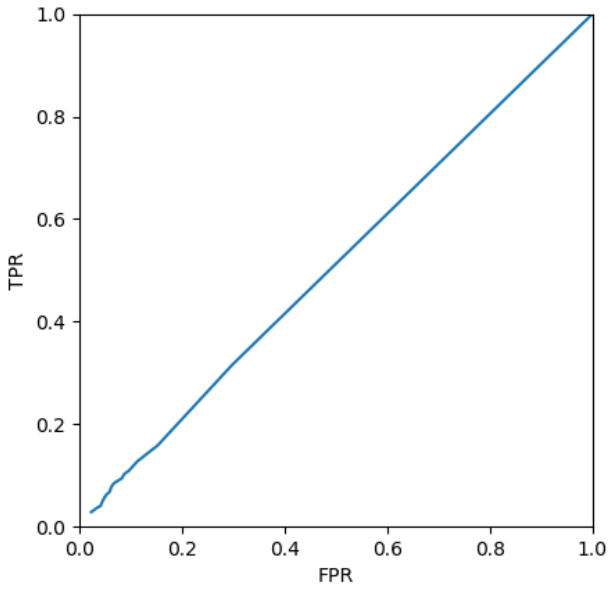
Random parameters seem it have excellent performance but the ROC curve is not really good

After training :

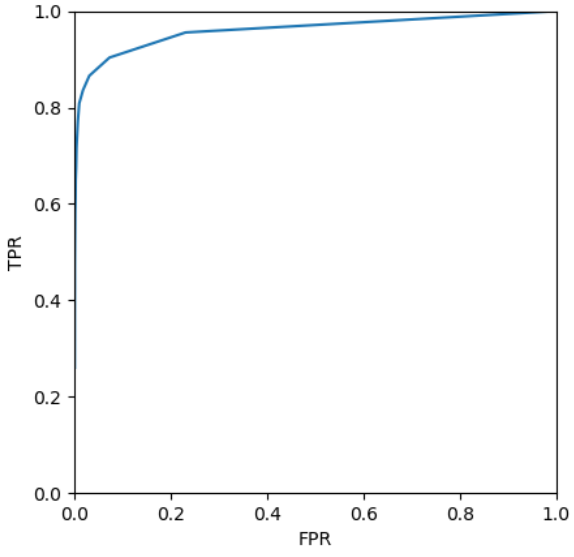
```
Logreg after-train train accuracy: 0.963795
Logreg after-train train precision: 0.962775
Logreg after-train train recall: 0.653272
Logreg after-train test accuracy: 0.967929
Logreg after-train test precision: 0.965184
Logreg after-train test recall: 0.649740
```

ROC curve :

1. Train data



2. Test data



Discussion :

1. Better theta will have large ROC curve area
2. Because of the lambda, it will decrease the recall but increase the precision