INFSCI 2725 Data Analytics

Assignment 4

Zheng Liu E-mail:zhengliu@pitt.edu

Luda Wang E-mail:luw20@pitt.edu

Kewei Li E-mail:kel137@pitt.edu

We chose Naïve Bayes Method and Neural Network Method with five-fold cross validation method. We compared the two methods by accuracy and ROC curve.

We import the data by using Pandas read_table method.

After import the data, we transfer the Char data to Int data like this:

Republican=1 ; Democatic=0;

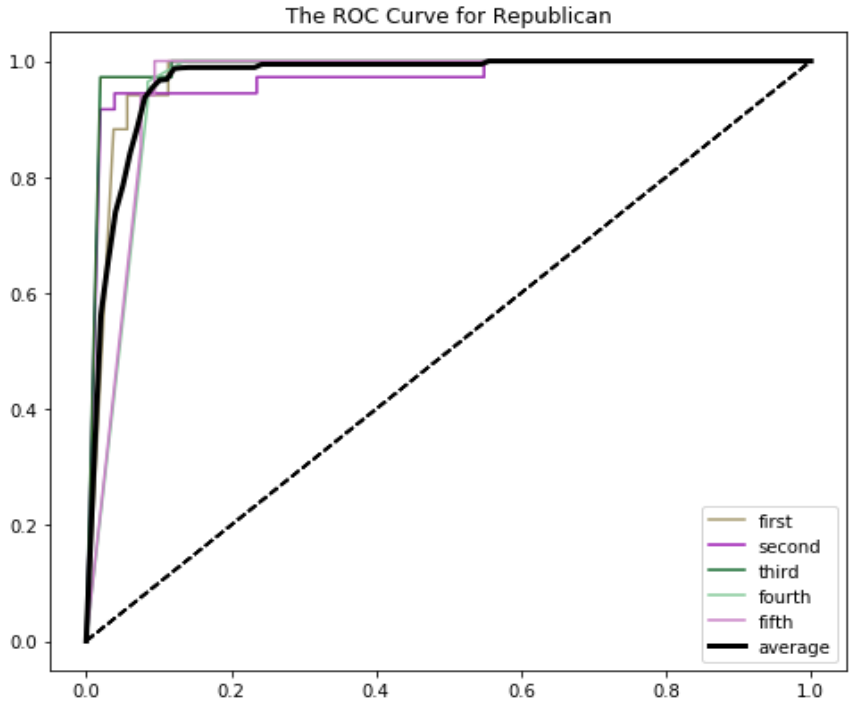
y=1; n=0; w=2.

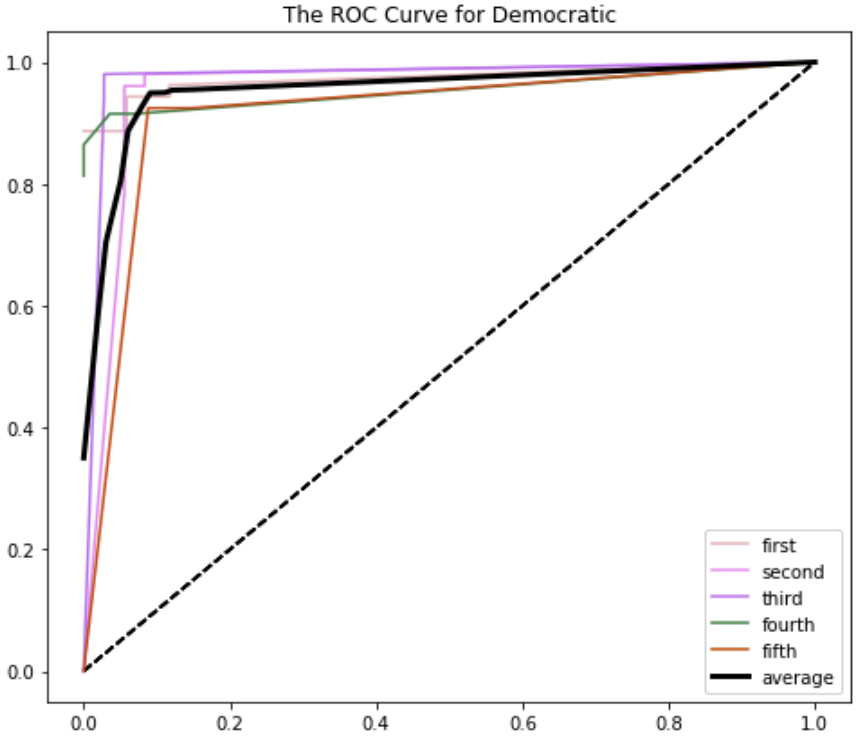
Then we import the RepeatedKFold, GaussianNB, MLPClassifier,recall_score, precision_score, metrics.

The result of Naïve Bayes Method is below.

Navie Bayes Method Result						
	1	2	3	4	5	Average
Accuracy	0.8851	0.9310	0.9655	0.9540	0.9540	0.9379
Number of wrong prediction	10	6	3	4	4	5.4
Sensitivity of Republican	0.8235	0.8889	0.9722	1.0000	0.9118	0.9193
Sensitivity of Democratic	0.9245	0.9608	0.9608	0.9322	0.9811	0.9519
Precision of Republican	0.8750	0.9412	0.9459	0.8750	0.9688	0.9212
Precision of Democratic	0.8909	0.9245	0.9800	1.0000	0.9455	0.9482

The ROC curve of Naïve Bayes Method is below.



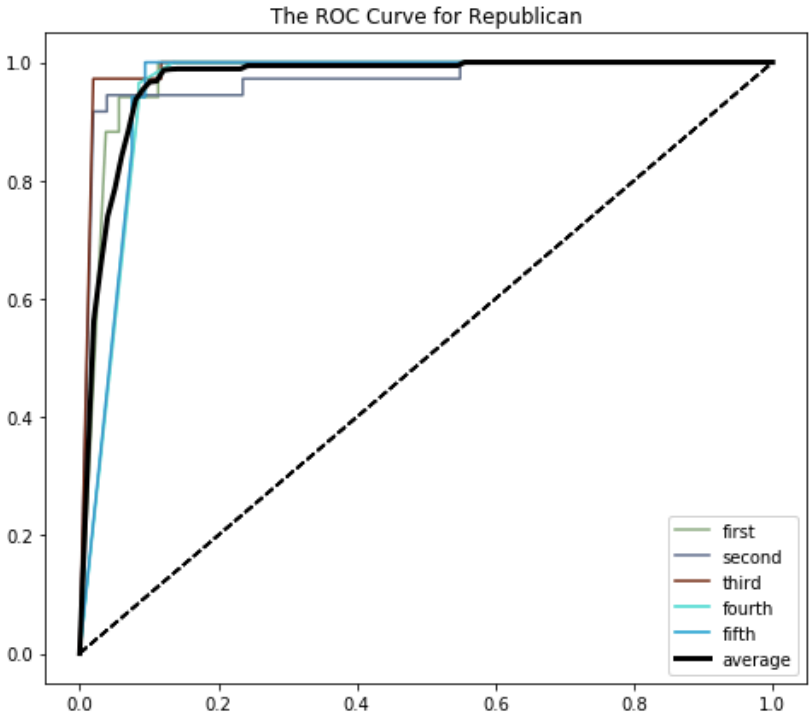


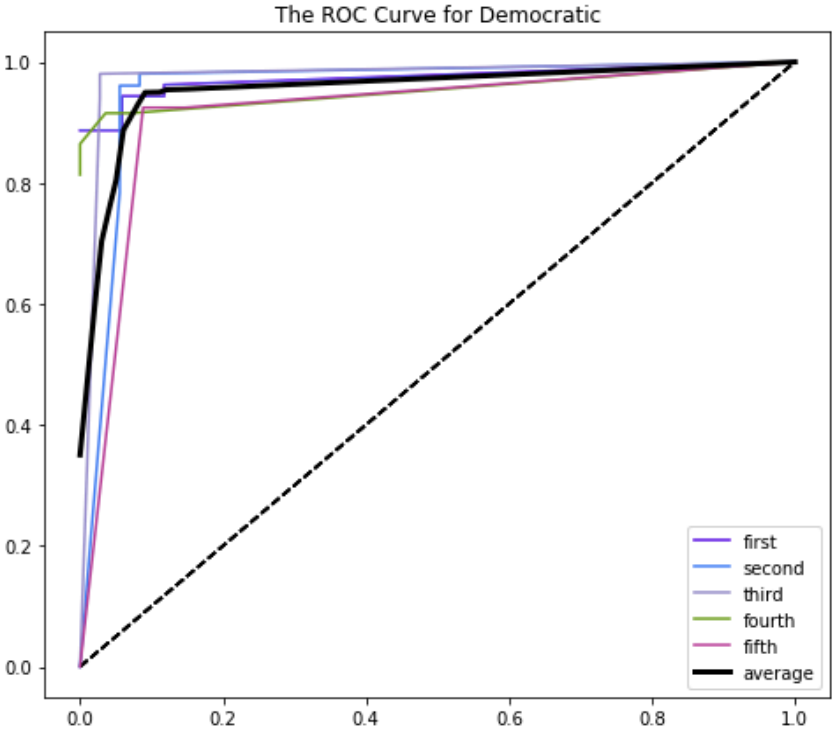
We also use the Neural Network Method to do the classification. We use two hidden layers. First layer has 5 units, and second layer has 2 units.

The result of Neural Network Method is below.

Navie Bayes Method Result						
	1	2	3	4	5	Average
Accuracy	0.9310	0.9425	0.9770	0.9195	0.8966	0.9333
Number of wrong prediction	6	5	2	7	9	5.8000
Sensitivity of Republican	0.8824	0.9167	0.9722	0.9286	0.8529	0.9106
Sensitivity of Democratic	0.9623	0.9608	0.9804	0.9153	0.9245	0.9486
Precision of Republican	0.9375	0.9429	0.9722	0.8387	0.8788	0.9140
Precision of Democratic	0.9273	0.9423	0.9804	0.9643	0.9074	0.9443

The ROC curve of Naïve Bayes Method is below.





By comparing the two methods, the accuracy is almost the same.