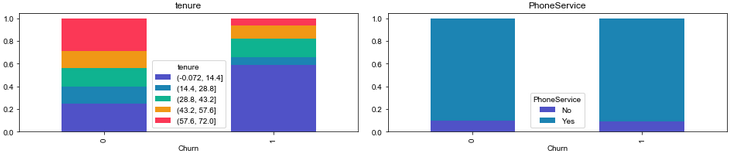
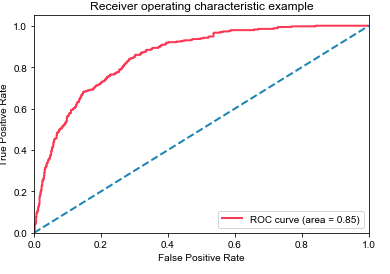
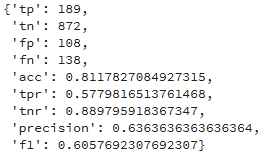
Hans Cainglet 22 September 2018

Dec 130- HW2: Churn Prediction using Logistic Regression Model

I attempted to predict whether customers would churn or not based on the given data. Before I processed said data, I cleaned the columns: converted cells into integers (ex: Female/Male to 1/0), and created dummy cells. I performed exploratory research to get a better look at the data. From the graphs derived, I dropped the “customer ID”, “gender”, “PhoneService” and “MultipleLines” columns as they did not provide deeper insight on whether customers would churn or not. Attached below is a snippet of the exploratory results, wherein tenure (left) is a variable worth considering whereas PhoneService (right) is not.



Afterwards, I used the .Logit function to calculate for the Beta coefficients with “Churn” as dependent variable. Attached below are snippets of the ROC curve along with the True Positive, True Negative, Precision, Accuracy, and F1 Score, which all show good results.



Overall, the training phase proved promising. However, I am still struggling to plug the data for the “Test” file and I am unsure on how to use the function to predict whether customers will Churn or Not in the new file. Attached below are the .Logit results for the training data. Nonetheless, From the Bi results, raising it to e (eBi) would suggest on the likelihood of that person to churn or not. For instance, senior citizen (e0.1784)= 1.195. When a person is a senior citizen, odds of churning are multiplied by 1.195 which is a high number!

