Shawn Ong

Dec 130 X

Logistic Regression

In order to predict whether a customer will churn or not, I first looked at the data and saw that a huge chunk of churners are those who do not avail of particular services such as Online Services, Tech Support, etc. The type of internet subscription is also a factor. I then created dummy datas for each of the possible category since there are three namely Yes, No, and No internet service. I then removed the other predictors so that the data frame will only consist of the numerical predictors. I then divided the data into testing and training. One problem I ran into is that when the data is divided into test and train, the logistic regression function will return an error statement stating that a model cannot be made as there is a singular matrix. Because of this, I would have to use a different method to linear regression. I would have to use Scikit Learn's Linear Regression instead. After creating the model, I prepared the test data similar to the train data removing the non-numerical predictors, leaving the same predictors as the train data. I then used the model to predict and placed the results in a csv next to the CustomerID.