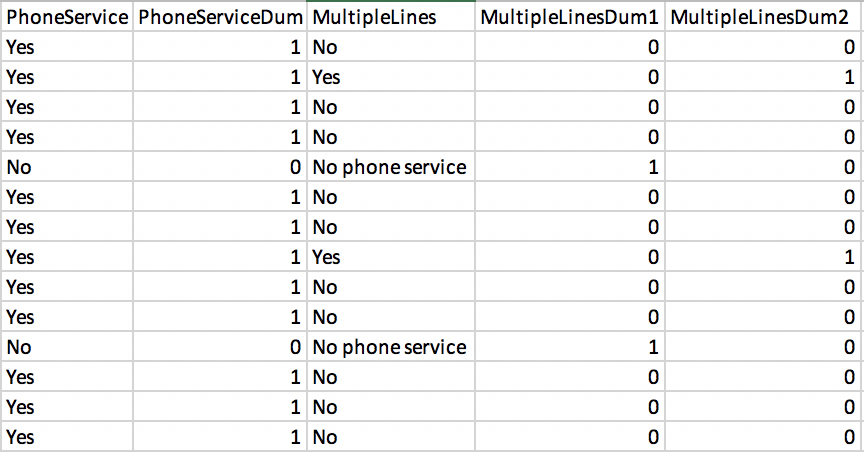
Judd Daniel Co Ang Homework # 1

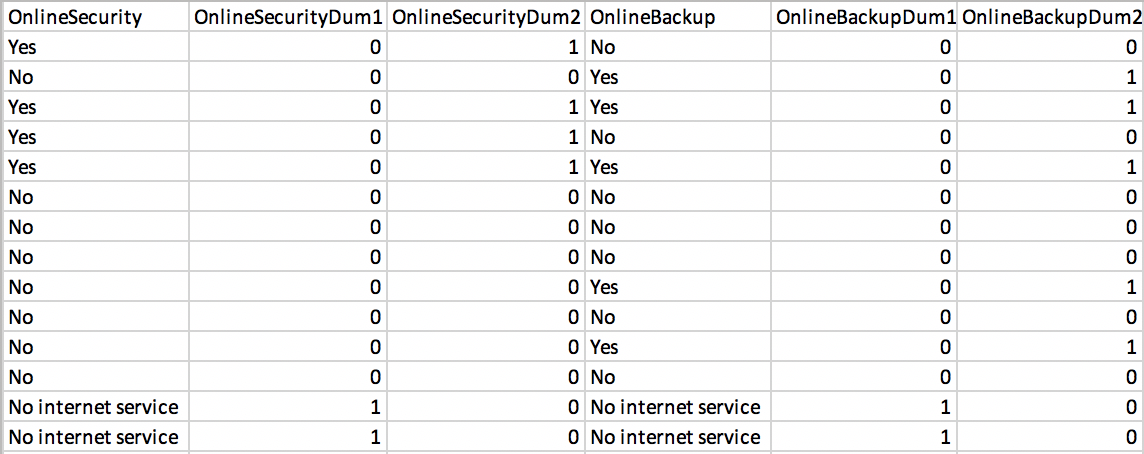
DEC 130 Section X Churning Prediction

In order to predict whether a subscriber from a certain telecommunications company will churn or not churn from his / her subscription, I have chosen to use a logistic regression analysis. The reasons for this are as follow:

1. The dependent variable only has 2 possible outcomes (i.e. Churn or Not churn).
2. The logistic regression model is semi-parametric measure that does not need to satisfy the general assumptions in a linear regression model.
3. A probability of churning or not churning can directly be derived from the analysis.

I used the SPSS software to aid me in conducting a logistic regression analysis. In order to formulate a logistic regression equation, I used the training data set provided during class; it should be noted that I decided not to conduct a data split on the training data set as a separate test set was provided by the professor. The first step was to reclassify any categorical variables into dummy variables as this is required in order for SPSS to be able to run the analysis. For variables with only “Yes” or “No” as possible categories (i.e. Partner, Dependents, PhoneService, PaperlessBilling, Churn), these were reclassified into 1 for “Yes” and 0 for “No”. Likewise, for gender with only two (2) possible categories, it was reclassed to 1 for “Male” and 0 for “Female”. Moreover, for the other variables with multiple categories (i.e. MultipleLines, InternetService, OnlineSecurity, OnlineBackup, DeviceProtection, etc.), “No” was used as the baseline for consistency, except for the “Contract” and “PaymentMethod” variables wherein “Month-to-month” and “Electronic check” were respectively used as baselines. A visual presentation of the reclassed dummy variables can be found below.

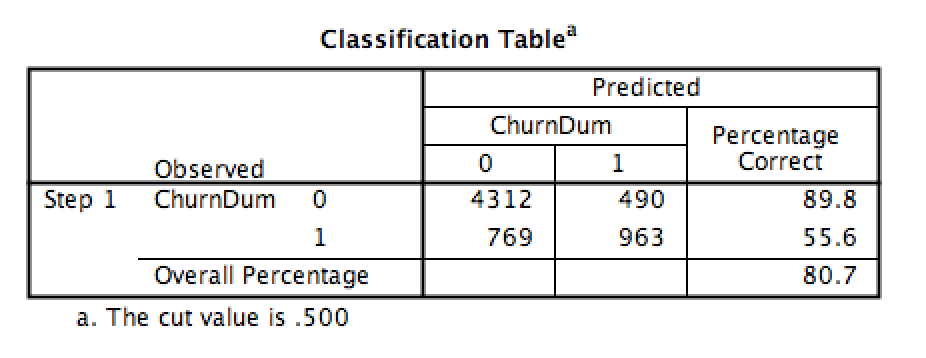




After the reclasses, a logistic regression analysis was run through SPSS and the following equations can be derived on which the reclassed dummy variables can be plugged in and an odds ratio can be obtained.

From here, a certain churning probability can be obtained through the following relationship:

Finally, the cutoff for the churning probability was obtained through the SPSS output of the logistic regression analysis as can be seen through the figure below. Customers with a corresponding probability of greater than 50% will most likely churn, while customers with a corresponding probability of 50% or less will most likely do otherwise.



\*Please refer to the attached excel file for reference on computations and reclassifications.

However, predictions using the above logistic regression model must be utilized with caution as the Hosmer and Lemeshow test yields a significant value of less than 0.05.

