

Bayesian Estimation of the Churn Probability

The main objective of this stage is to build a python code to predict the probability of a customer churn given a number of parameters (values in columns of the data set). This project assignment will use the pdfs and CDFs developed in the previous project assignment and should use them for Bayesian estimation of the output .

Assuming that the word churn represents the Event Churn =yes and that the letters A,B,C,D.... represents events form all other columns in the data set then we are interested in calculating

$$\Pr(\text{Churn}|A, B, C, D, \dots)$$

Since it is very difficult to calculate the joint probability $P(B, C, D, \dots)$ we will use pays low to do the following

$$\Pr(\text{Churn}|B, C, D, \dots) = \frac{\Pr(A, B, C, D, \dots | \text{Churn}) P(\text{churn})}{\Pr(A, B, C, D, \dots)}$$

Assuming that variables are independent

$$\begin{aligned} & \Pr(\text{Churn}|B, C, D, \dots) \\ &= \frac{\Pr(A|\text{Churn}) \Pr(B|\text{Churn}) \Pr(C|\text{Churn}) \Pr(A|\text{Churn}) \Pr(B|\text{Churn}) \Pr(D|\text{Churn}) \dots P(\text{churn})}{\Pr(A) \Pr(B) \Pr(C) \Pr(D) \dots} \end{aligned}$$

In order to do that for a customer that is represented by a number of events (a number of values in the all the columns except the churn column) , we need to have a way to calculate the conditional probabilities such as $\Pr(A|\text{Churn})$ and the actual probabilities such as $P(A)$

To do that this project assignment has the following requirement

- 1- Evaluate the relation between different columns in the data set by finding if they are dependent or not and by calculating the covariance and correlation when applicable.
- 2- Find data anomalies and decide what to do with these anomalies
- 3- Fit the pdf of each numerical numbers to one of the known pdfs (e.g normal, exponential , beta ...)
- 4- Fit the pdf of the each numerical conditional pdfs to one of the known pdfs (e.g normal, exponential , beta ...)
- 5- Build a python code to evaluate bays rule and estimate the probability of churn given certain inputs
- 6- Build a python code to evaluate the accuracy of your prediction.
- 7- Investigate if removing some of the inputs will enhance your prediction.

Bonus

A bonus for this assignment would be to develop an extra method for predicting churn that is better that can provide a better performance