TIME SERIES

B Raja Narasimhan

January 2020

1 Introduction

Time Series is basically data points ordered in time. Some of the examples are stock price, sales of product based on time, ECG data etc. In this question we were asked to make our own skewed data-set and use a proper classifier algorithm to train it and evaluate the performance of classifier.

2 Dataset

For this assignment I have made the dataset which is amount of cars sold by a particular company in the particular month of the year. The parameter for if the quantity sold is more than their forecast then the value is 1 else their value is 0. This is a skewed dataset because forecast is done by professionals who have huge experience in their field and they take into consideration a lot of parameters(even some which seem insignificant).This happens because of sudden external and internal factors. Some months may be exciting for these company but there may be sudden decrease because of some natural disaster in manufacturing location or sometimes some months may be dull but there will be way more cars sold maybe because of news government may soon increase the tax etc. So there will lot more 0's than 1. Here I have made dataset where one column tells time(0 is first half of January, 5 next half, 10 is first month of February, 15 is the next and so on. . .). Next column tells the sales. Next column tells 0 or 1. I made 2 dataset for 2 different years.

3 Classifier

So XGBoost being the best classifier I applied it and got 75 percent accuracy and I evaluated the classifier and the results were as follows.
Accuracy: 0.75
Mis-Classification: 0.25
Sensitivity: 0.17
Specificity: 0.94
Precision: 0.94
Now I started tuning the parameters and I found out for learning rate =0.1, n es-

timators=4000, max depth=3, min child weight=1, gamma=0, subsample=0.9, colsample bytree=0.8,scale pos weight=6(In my dataset) my classifier was evaluavted as follows
Accuracy: 0.88
Mis-Classification: 0.12
Sensitivity: 0.67
Specificity: 0.94
Precision: 0.94

4 Future prediction

Here I tried to predict see pattern of sales in my dataset. I used exponential shift to see pattern. Here if alpha is between 0.35 to 0.6 we can see some patterns. If alpha is less than 0.35 graph is very vague, more than 0.6 it kind of traces the graph.