

Author: Ajit Bhanot (abhanot2@illinois.edu)

CS 598: Practical Statistical Learning: Project 2

Project II: Walmart Stores Forecasting

1. Introduction

The data given to us is the historical sales data for 45 Walmart Stores in different regions. The objective was to create predictions using 3 different models.

The starter code pack provided by the instructors was extremely helping in setting up the initial framework. The task was pretty much reduced to find the best 3 models. They were, `snaive`, `tslm`, and `stlf`. These are standard functions in R forecast package that is imported.

2. Model and Methodology:

There is a lot of information for this well-known challenge and many resources are available on the Kaggle site (<https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting/discussion/8023>), slack and of course numerous Piazza questions. The three models chosen were simple models that could achieve the goals set for the project. The approach for `stlf` was a bit different, after a few trial and errors and discussion on piazza, `stlf` was turned into `tslm` and `stlf` (`stlf` beyond the time we had enough data). Furthermore all NA's were considered to 0. For creating the time series, the frequency is 52, as given in the helper functions.

Model1: Simple `snaive` function as adapted from the template code given

Model2: Simple `tslm` (time series linear model)

Model3: (Model 2 for the time we do not get enough data and then switched over to `stlf`)

There was not much tuning required to achieve the desired goal of 1630 or better for average. The results are attached below along with the times.

The results for all three models are below. The execution was done on a Windows 10, Intel i7-7700 CPU @3.60GHz with 16GB RAM.

WMAE

Fold	Snaive	TSLM	TSLM Switch to STLF
1	2262.42	2042.40	2042.40
2	1787.08	1440.08	1440.08
3	1779.05	1434.72	1434.72
4	1716.12	1596.99	1596.99
5	2400.40	2327.64	2327.64
6	1696.90	1674.18	1674.18
7	2086.97	1718.58	1622.24
8	1750.28	1420.82	1348.01
9	1719.89	1430.80	1279.67
10	1680.96	1447.03	1245.37
Average	1888.01	1653.32	1601.13

Run times in Seconds

Fold	Snaive	TSLM	TSLM Switch to STLF
1	25.5	24.2	25.2
2	24.2	24.6	24.2
3	26.5	23.2	22.7
4	27.8	22.6	23
5	28.9	23.2	24.2
6	27.3	24.4	22.6
7	28.2	24.9	66.8
8	30.5	24.4	71.7
9	31.9	23.6	74.9
10	30.9	24.5	77.6