CS598 - Project 1

Xiaoming Ji

Computer System

Hardware

- Dell Precision Tower 5810
- CPU: Intel Xeon E5-1607 @ 3.10GHz
- Memory: 32GB

Software

- OS: Windows 10 Professional 64bit
- R: 3.5.1
- R Packages:
  - forecast_8.4
  - tidyverse_1.2.1
  - lubridate_1.7.4

Models

3 approaches (total 4 models) are used to generate the prediction:

- Naive model
- Seasonal naive model
- Dynamic model: for fold 1 to 6, regression model (tslm) is used. Start from fold 7, since the training data has more than 2 years of data, STL+ARIMA (method='arima', ic='bic') model is built to make the prediction.

Missing value handling

- Weekly_Sales: replace missing value with 0.
- IsHoliday: search through the training data to find the IsHoliday of same date. See function: `fill_missing_holiday`

Note: my testing show more sophisticated imputation approach won't improve the performance.

Test results

Fold	Naive	SNaive	Dynamic
1	2688.673	3996.682	4234.129
2	4982.033	1335.095	1374.007
3	3673.091	1554.424	1833.845
4	7647.495	1615.619	1621.523
5	6631.979	2478.130	2534.620
6	5785.446	1973.211	2515.315
7	3725.379	3728.204	3263.735
8	1696.924	1686.528	1340.781
9	4285.989	2492.405	2241.240
10	7131.503	2059.334	2132.104
Average	4824.851	2291.963	2309.130

Computation time: 4.68 seconds