

CS598 - Project 2

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 are used to generate the prediction:

- Seasonal naive model
- Time Series Regression model: with trend and season.
- Dynamic: for fold 1 to 6, same time series regression model as 2nd approach is used. Starting 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.

Pre-processing

- Run SVD (first 12 components) on each by-department sales data and then transform it back to the original matrix size.
- Missing value handling, I simply replace missing Weekly_Sales with 0 in the svd function.

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

Test results

Fold	Naive	SNaive	Dynamic
1	2216.924	1969.731	1969.731
2	1744.291	1379.503	1379.503
3	1743.753	1389.163	1389.163
4	1665.263	1550.923	1550.923
5	2384.856	2313.864	2313.864
6	1629.668	1642.299	1642.299
7	2024.295	1688.378	1593.354
8	1677.858	1392.058	1324.968
9	1652.767	1413.899	1266.187
10	1624.973	1426.096	1233.962
Average	1836.465	1616.591	1566.395

Computation time: 1889.326 seconds