

# Project 2 Report

*Tianqi Wu*

*11/8/2018*

## Model

**Naive:** I implemented a naive model for each store and department combination. In particular, missing values for the weekly sales were imputed with zeros, then the last value of the weekly sales during the train period was used to predict each new week in the testing period.

**TSLM\_SVD:** The second model computes a forecast using linear regression and seasonal dummy variables. Also, preprocess of SVD is used which replaces the training data with a rank-reduced approximation of itself. Missing values for the weekly sales were imputed with zeros. The number of components kept is 12.

**TSLM\_STLF\_SVD:** The third model combines STLF with the second model. It uses stlf() from the forecast package for the observations after 2012-02-28. This date is chosen since stlf needs two years of data in this case. Hence, tslm is first used to reach two years worth of data. Then, there is finally enough data to do stlf. The number of components kept is 12 and the model typed used is 'ets'.

The fuctions are leverage from the sample code given by Josh and winner code[1] from Kaggle with modifications. For the hyperparameters, number of components(12,13,14,15) and model type('ets','arima') are tried. No holiday adjustment is performed. The best result chosen is indicated above.

## Result

From the table,TSLM with SVD achieves best result of WMAE(1614). The overall running time of whole script is around 10 minutes and the computer system is Macbook Pro 3.1GHz, 8GB memory.

Table 1: Error of Models

| Fold    | Naive | TSLM_SVD | TSLM_STLF_SVD |
|---------|-------|----------|---------------|
| 1       | 2079  | 1967     | 1967          |
| 2       | 2589  | 1377     | 1377          |
| 3       | 2254  | 1385     | 1385          |
| 4       | 2823  | 1550     | 1550          |
| 5       | 5156  | 2310     | 2310          |
| 6       | 4218  | 1640     | 1640          |
| 7       | 2270  | 1686     | 1778          |
| 8       | 2144  | 1392     | 1535          |
| 9       | 2221  | 1412     | 1381          |
| 10      | 2372  | 1425     | 1344          |
| Average | 2813  | 1614     | 1627          |

## Reference

[1] [https://github.com/davidthaler/Walmart\\_competition\\_code](https://github.com/davidthaler/Walmart_competition_code)