MSBD 5013 Project 1: M5 Forecasting

Hao Wu, Yuhan Zhou and Lingjun Guo {hwucc, yzhoudv, lguoal}@ust.hk
Department of Computer Science and Engineering, HKUST

1. Introduction

We have applied different methods to predict sales of specific items in specific stores in the next 28 days, and finally we choose the simple exponential smoothing method with smoothing parameter 0.2 to encode due to the consideration of computing time and accuracy.

Simple Exponential Smoothing (SES): The simplest exponential smoothing model, aimed at predicting series without a trend, defined as

$$\hat{Y}_t = aY_t + (1-a)\hat{Y}_{t-1}.$$

The smoothing parameter alpha is selected from the range [0.1, 0.3] by minimizing the insample mean squared error (MSE) of the model, while the first observation of the series is used for initialization.

2. Dataset (Kaggle Competition)

The datasets consist of previous 1941 days (starting from 2011-1-29) sales information of 3049 items in 10 stores of 3 states in US(CA, TX and WI). We also have a calendar dataset that gives information about dates and events.

Compare to the benchmark

Comparing on the kaggle private leaderboard that is ranked by the Weighted Root Mean Squared Scaled Error (WRMSSE), our WRMSSE using the SES method is 0.01085 lower than the benchmark submission, which means our prediction is more accurate.

Score: 0.95829

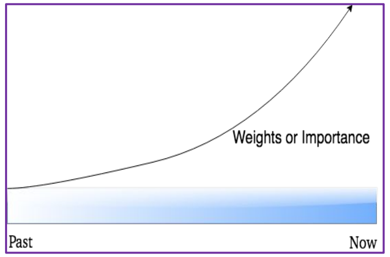
Our WRMSSE on the private leaderboard

SES.csv	0.96914
Benchmark Submission MLP_l.csv	0.97683
RF_l.csv	1.00952
RF_g.csv	1.06513
Naive.csv	1.75200

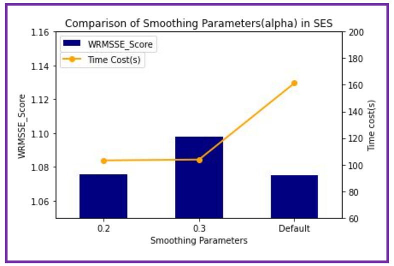
The benchmark submission on the private leaderboard

3. SES model and the Parameter

We use the SES model in the statsmodels.tsa API to fit our data. Although the price of some products may have trends or seasonal effects, the SES model is timesaving, and it assigns more weight to more recent data which seems also reasonable to our dataset. The SES model selects the smoothing parameter automatically, but the overall behavior is much better when we assign a value between 0.1 and 0.3. Finally, according to the comparison of WRMSSE score, we decide to use default smoothing parameter despite of relatively long running time.



SES assigns more weight to more recent data



Computing time under different smoothing parameters alpha

4. Evaluation

According to the rules of the Kaggle competition, we apply Weighted RMSSE (WRMSSE) to evaluate the different models we have tried, using following formula:

$$RMSSE = \sqrt{\frac{1}{h}\frac{\sum_{t=n+1}^{n+h}(Y_t-\hat{Y}_t)^2}{\frac{1}{n-1}\sum_{t=2}^{n}(Y_t-Y_{t-1})^2}}, \qquad WRMSSE = \sum_{i=1}^{42,840} w_i * RMSSE,$$

where wi is the weight of the i-th series of the competition.
The models with lower WRMSSE scores have better performances.

5. Compare to other methods

	SES	LSTM	LGBM
Running Time	103	204	20055
WRMSSE Score	0.95829	1.04530	0.72530

Compared to performance of other models including LSTM and LightGBM, SES have relatively good performance and high efficiency.

6. Conclusion

We have tried several different models including LSTM, SES, Prophet and LGBM in this project. Because of the time constraints, we have to consider not only the accuracy but also the efficiency of a model. Therefore, we finally choose SES to finish the project even through LGBM model have a better performance in prediction accuracy.

For the future work, we will concentrate on the improvement of accuracy of the prediction regardless of running time.

7. References

https://mofc.unic.ac.cy/wp-content/uploads/2020/02/M5-Competitors-Guide_Final-1.pdf

https://medium.datadriveninvestor.com/how-to-build-exponential-smoothing-models-using-python-simple-exponential-smoothing-holt-and-da371189e1a1

https://github.com/Deshram/Sales-Forecasting-for-Retail-Chains

8. Contribution

Model and Parameter Selection - Hao Wu, Lingjun Guo
Coding and Evaluation - Lingjun Guo, Yuhan Zhou
Poster writing - Yuhan Zhou, Hao Wu