

Time Series Forecasting with RNNs

Thomas Keate

UTU - CS 4320

April 30, 2024

- Time series are sequences of data points collected at constant time intervals.
- Time series forecasting is the use of a model to predict future values based on previously observed values.
- Time series forecasting is used in many fields such as stock prediction, weather forecasting, and sales forecasting.
- Recurrent Neural Networks (RNNs) are a popular choice for time series forecasting.

Dataset and Data analysis

- The data is from the kaggle, "Store sales - Time Series Forecasting" competition.
- The data consists of multiple rows ordered by date, along with multiple stores and product categories.
- I had to cleanup and combine the data, the final dataset consisted of two different forms, one which was the sum of sales for all the stores based on the day, and the cleaned data.

id	date	store_nbr	sales	onpromotion	family_AUTOMOTIVE	...
1262933	2014-12-11	44	9.0	0	False	...
1815844	2015-10-18	9	27.0	0	False	...
565947	2013-11-14	38	2.0	0	False	...
1702487	2015-08-16	28	1.0	0	False	...

Table: First 4 lines of the dataset

Dataset and Data analysis (cont.)

The data I used for the time series forecasting model consisted of the sum of sales for all the stores based on the day. It is stored in an array and looked like this:

```
[2511.6189990000003  
496092.417944  
361461.23112400004  
354459.67709270003  
477350.121229  
519695.40108800004  
336122.801066  
318347.777981  
302530.80901799997  
258982.003049  
289737.6850845  
403258.2120105]
```

Problems with the Dataset

- The dataset had multiple problems that needed to be fixed before modeling.
- The dataset had missing values that needed to be filled in such as 0 sales for a particular day.
- I had to combine the sales for a particular day.

Modeling

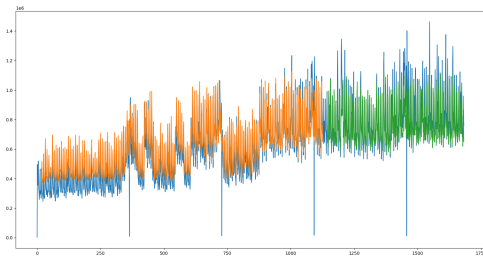
- I first created a neural network linear regression model to predict the sales for a particular day given if there was a holiday or not.
- Then I created a time series forecasting model using tensorflow and keras LSTM to predict the total sales for a particular day. I tried multiple different models and hyperparameters to find the best model.
 - The best model consisted of 3 layers, 2 dense layers and 1 output layer. The model was trained for 100 epochs.

Layer (type)	Output Shape
LSTM	(1, 1, 4)
LSTM_1	(1, 1, 4)
LSTM_2	(1, 4)
Dense	(1, 1)

Results

The model was trained on 66% of the data and tested on the remaining 33%.

- The best model had a training RMSE of 109284.68.
- The best model had a testing RMSE of 124407.87.

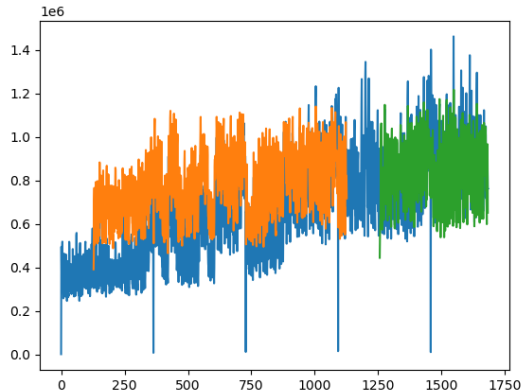


The graph above shows the predicted sales vs the actual sales. The orange line is the predicted sales on the training data and the green line is the predicted sales on the testing data.

Results (Cont.)

I ran the model again but using a lookback of 128 days. Instead of the 16 days I used before.

- The best model had a training RMSE of 222065.39.
- The best model had a testing RMSE of 123725.30.



Conclusion

- The model was able to predict the sales for a particular day with a RMSE of 124407.87.
- The model could be improved by using more data and more features.
- The model could be improved by using a more complex model.
- Overall, the model was able to predict the sales for a particular day with a reasonable amount of error.