Forecast case study

Keerthi Krishna PARVATHANENI

April 24, 2024

1 Model for prediction

The main step involved the prediction of any data can be organized as follows.

- Data Preparation
- Model Creation
- Model Deployment
- Model Prediction
- conclusion

Data Preparation

Data preparation combines many steps such as normalizing, resampling, removing tendency of the data. Normalizing is a common practice to bound the values most common may of normalizing data is to change the data range from 0 to 1. In time series, we also need to deal with the data tendency such as stationary and non-stationary.

Data Preparation for NonLinearRegression in PyTorch

Preparation of X-data:

Since year have 52-53 weeks, I thought it's a good idea to normalize the data by the corresponding week number divided by total number of weeks in that year $X_{data} = \frac{weeknumber}{totalweeeksinthatyear}$.

Preparation of Y-data:

For data to normalize is much straight forward, we just normalize the data based on the maximum and minimum value.

$$Y_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

Model Selection

There are a several forecasting models available from different open source python libraries, a few of them are *scikit-learn, PyTorch,* and so on. In this work we will use ***scikit-learn and PyTroch***, the primary reason for selecting this model is flexibility and easiness of this module.

The *PyTorch model* consists of four fully connected layers with *ReLU* activation functions between them. It's designed to map input features to output features for nonlinear regression tasks, where the relationship between input and output is nonlinear. The number of hidden layers and the number of neurons in each hidden layer can be customized based on the complexity of the problem and the available data.

The hyperparameters can be tuned by visually assisted model performance visualizer runs on streamlit app. You can run this app in your local ZSH terminal using command *poetry run streamlit run visualize _data* .

The hyperparameters such as number of Epochs, Learning rate, Split between train and test data, Number of hidden neurons can be entered in the slide-box given on the right. Use can also choose to enter only specific data that can be choice by department and business ID. We can also truncate the data provided to the model by closing the start date and end date figure(1).

2 Preliminary questions and EDA

- Which department made the highest turnover in 2016?

  Answer: Department with the highest turnover in year 2016 is, 127.

- What are the top 5-week numbers (1 to 53) for department 88 in 2015 in terms of turnover over all stores?

  Answer: Table 1

Table 1: Top 5 Performing Weeks

Week	Business Unit ID	Department Number	Turnover
2015-01-10	132	88	2306.643525
2015-07-11	4	88	2073.152091
2015-07-25	4	88	2070.041381
2015-07-04	4	88	2028.924124
2015-09-12	11	88	1935.623844

- What was the top performer store in 2014?

  Answer: Table 2

Comment	Value
Maximum turnover:	166858.495
Department with maximum turnover:	127
Business unit with maximum turnover:	121

Table 2: Top performer store in 2014.

- Based on sales can you guess what kind of sport represents department 73?

  Answer: Figure 2

  Looks like the summer sales, I guess sportswear like swimming, trekking and some shoes and cloths.

  In the following figure (deparment-73 and business-18), we can see the average sales were highest during summer and before fall.

- Based on sales can you guess what kind of sport represents department 117?

  Answer: Figure 3

  Looks like winter sales, I guess it should be related to ski, winter sports, Jackets, wool socks, Shoes, Skies and so on ..

- Are the time series stationary? Is this a problem?

  Answer: Time series data for certain stores/business units is stationary and others are not. To judge this I looked at the de-trended data, and also we conducted a Dickey-Fuller test.

  Stationary time series are easier in prediction when compared to non-stationary time series.

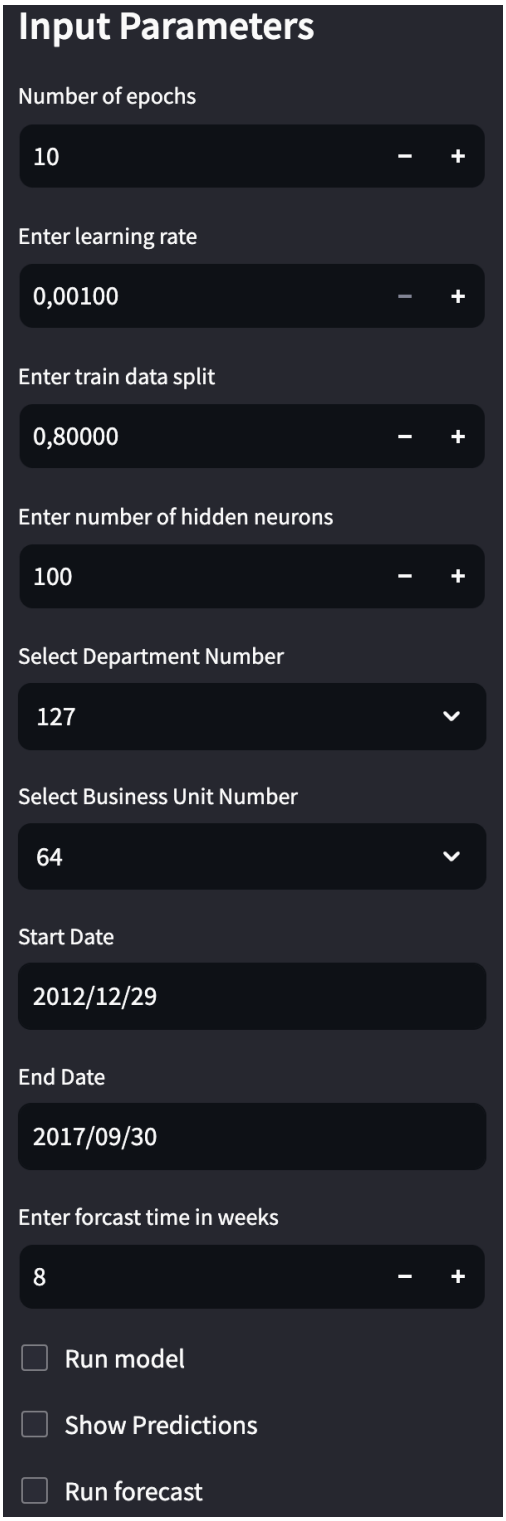
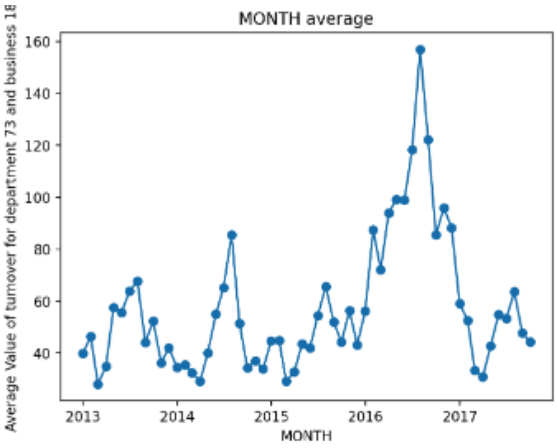
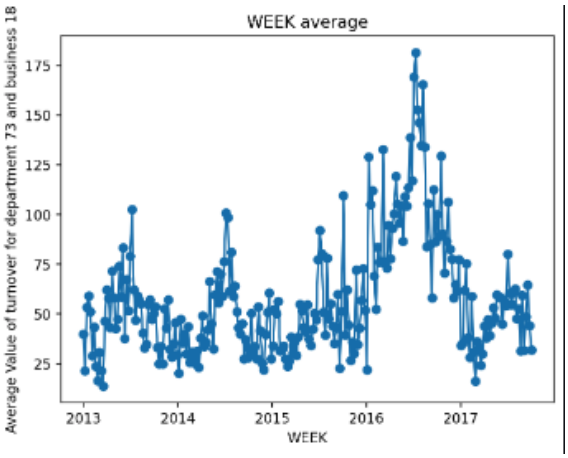


Figure 1: Options for tuning the PyTroch model

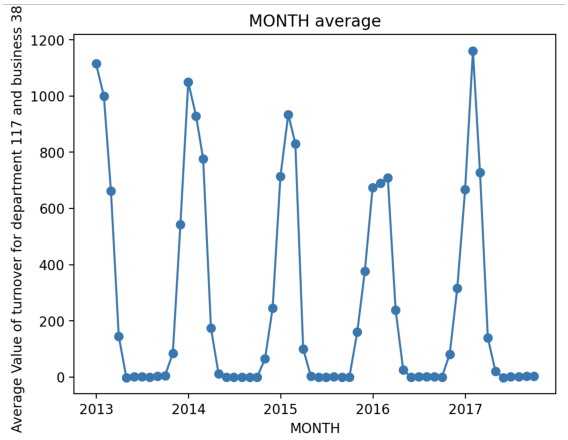


(a) Monthly average of department 73 business 18

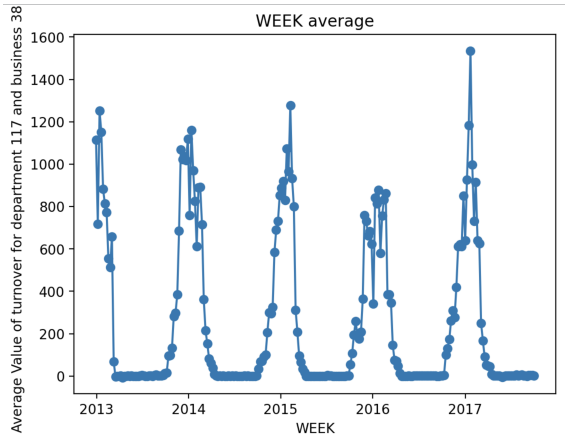


(b) Monthly average of department 73 business 18

Figure 2: Average turnover for department 73 business 18



(a) Monthly average of department 117 business 318



(b) Monthly average of department 117 business 318

Figure 3: Average turnover for department 117 business 318

- Can the time series be decomposed into several key components?
  Answer: Yes, time series can be decomposed into trends, seasonal and residual components. You can find a code that will help to visualize the trends for a sample data.