香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

M5 Forecasting - Uncertainty

Group 1:

Chen Tian

Li Muxiao

Qiu Qingqing

Introduction

- Background
  - A sales forecasting is an estimation of the quantity and value of products to be sold at a specified time in the future.
  - Sales forecasting based on historical data plays an important role in helping enterprises make better business strategies.
- M5 Forecasting which covers Walmart from three US states(California, Texas, and Wisconsin)
- Complementary to first project (Accuracy):
  - Focusing on the uncertainty

    Do probabilistic forecasting for the corresponding median and four prediction intervals (50%, 67%, 95%, and 99%).

Dataset

- **train-evaluation.csv**

  The historical daily unit sales data per product and store; [d1,d1941].

- **calendar.csv**

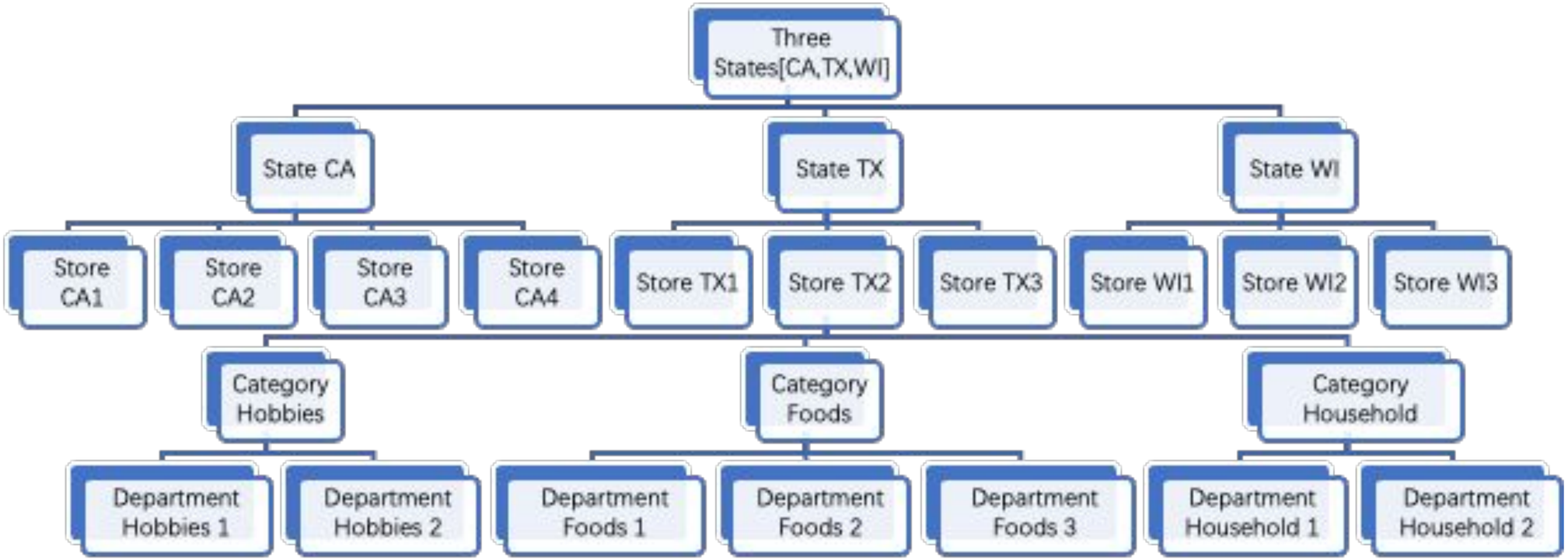
  The dates on which the products are sold along with associated functions (such as day of the week, month, year) and 3 binary markers that indicate whether stores in each state are allowed to purchase SNAP food stamps on that date.

- **sell-prices.csv**

  Information about store, item ID, and average weekly prices of the products sold per store.

Dataset

Dimension hierarchy relation in the dataset



Theory

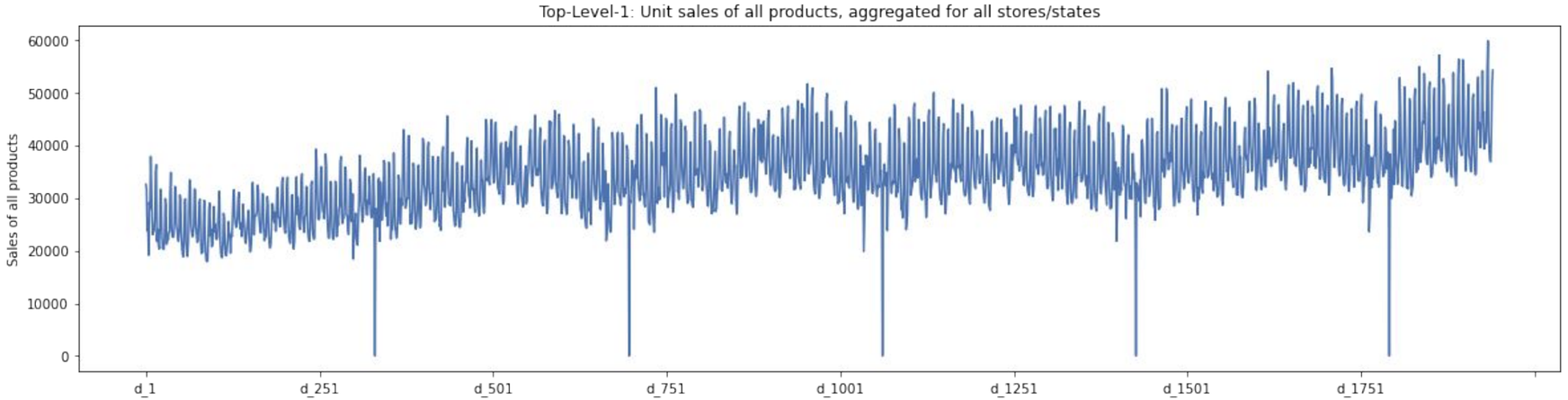
- **Artificial Neural Network (ANN)**
  - ANN is a computational model that is inspired by the way biological neural networks in the human brain process information. It is an extensive and interconnected network of adaptive simple neurons whose organization mimics the interactive responses of the biological nervous system to real-world objects.

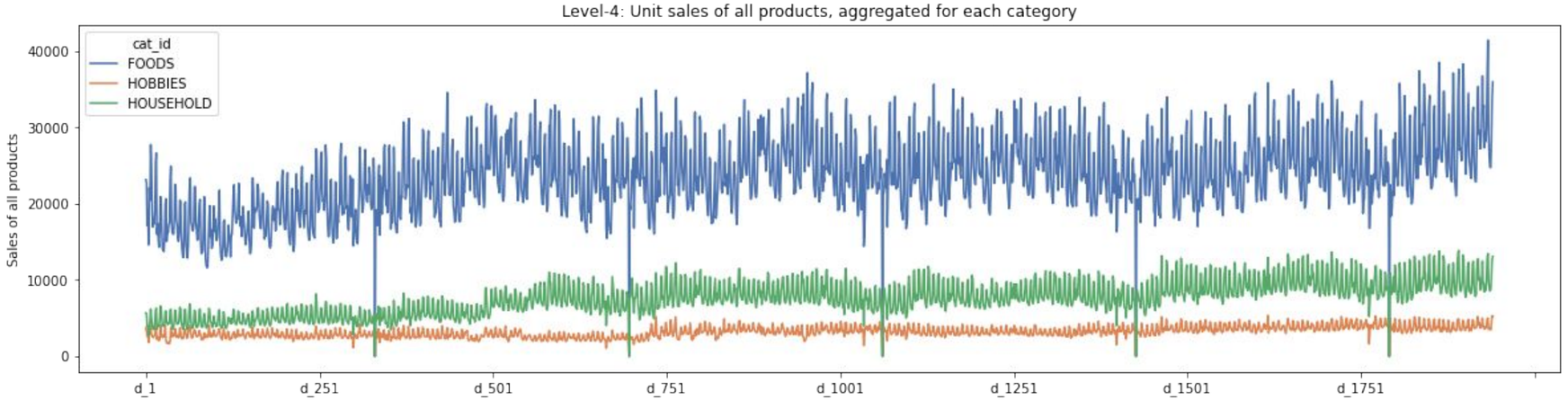
- **Long Short Term Memory Network (LSTM)**
  - LSTM is one of the Recurrent Neural Networks that consists of repeated modules of neural networks. It add the explicit handling of order between observations when learning a mapping function from inputs to outputs, not offered by ANN.

Data Analysis

Level id	Aggregation Method	The number of time series	Level id	Aggregation Method	The number of time series
1	Unit sales of all products, aggregated for all stores/states	1	7	Unit sales of all products, aggregated for each State and department	21
2	Unit sales of all products, aggregated for each State	3	8	Unit sales of all products, aggregated for each store and category	30
3	Unit sales of all products, aggregated for each store	10	9	Unit sales of all products, aggregated for each store and department	70
4	Unit sales of all products, aggregated for each category	3	10	Unit sales of product x, aggregated for all stores/states	3049
5	Unit sales of all products, aggregated for each department	7	11	Unit sales of product x, aggregated for each State	9147
6	Unit sales of all products, aggregated for each State and category	9	12	Unit sales of product x, aggregated for each store	30490

Data Analysis

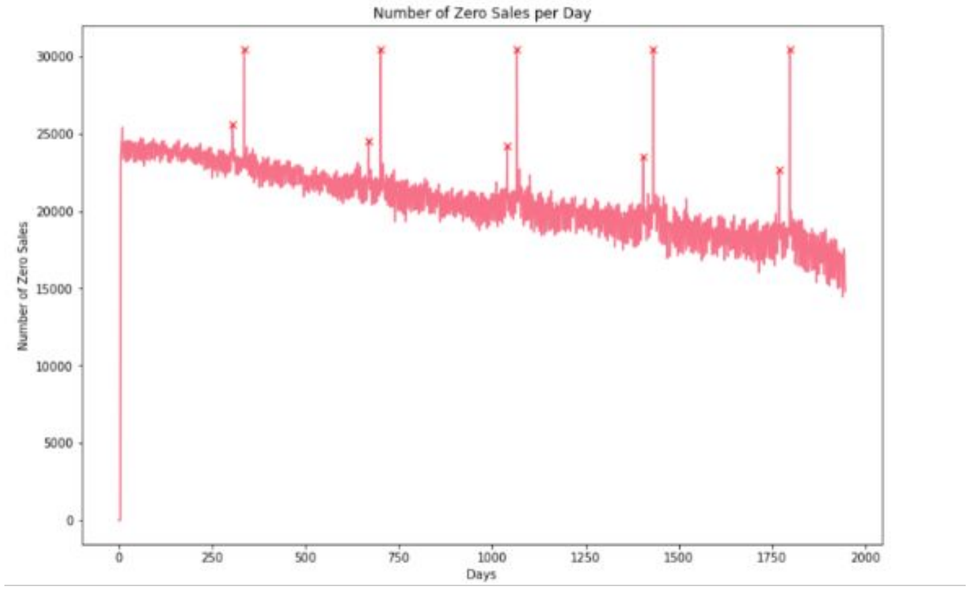




- Remarkable periodic patterns.
- Within each level, there exist differences of groups.

Data Preprocessing & Feature Engineering

- **Data Preprocessing**
  - Replace outliers(sales on Thanksgiving Day and Christmas Day) using the mean value.
- **Feature Engineering**
  - Encoding the categorical variables ‘event name’ and ‘event type’ into numbers
  - Add Lag Feature
    - A lag is a fixed amount of passing time.
    - [28, 35, 42, 49, 56, 63]



LSTM Model

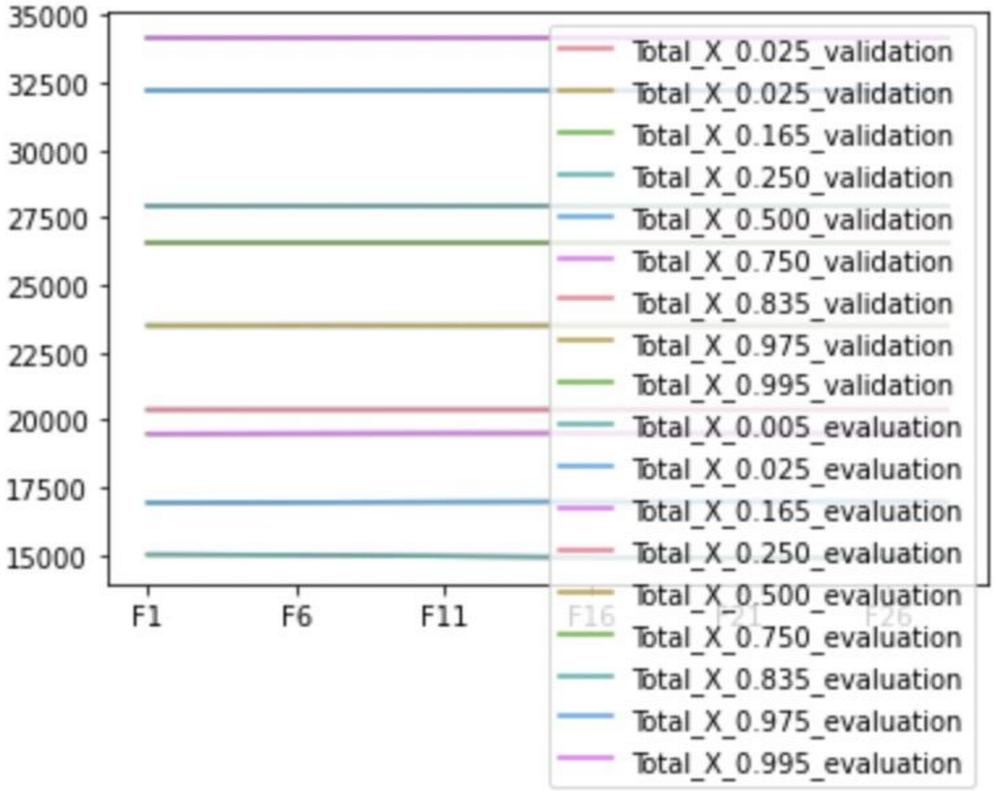
$$\mathbf{SPL(u)} = \frac{1}{h}\frac{\sum_{t=n+1}^{n+h}(Y_t - Q_t(u))u\mathbf{1}\{Q_t(u)\leq Y_t\} + (Q_t(u) - Y_t)(1-u)\mathbf{1}\{Q_t(u) > Y_t\}}{\frac{1}{n-1}\sum_{t=2}^{n}|Y_t - Y_{t-1}|}$$

The settings of parameters are shown in the table. We defined pinball loss as loss function and choose Adam as optimizer.

It is noteworthy that there may be better settings without memory storage and time limitation. The epoch is about 30 with batch size 44.

Layer	First Layer	Second Layer	Third Layer	Fourth Layer
Units	40	400	400	Dense(42840)
Dropout	0.2	0.2	0.2	

LSTM Model



- The specific changes of predicted values are not obvious visually.

- The model needs improvement.

ANN Model

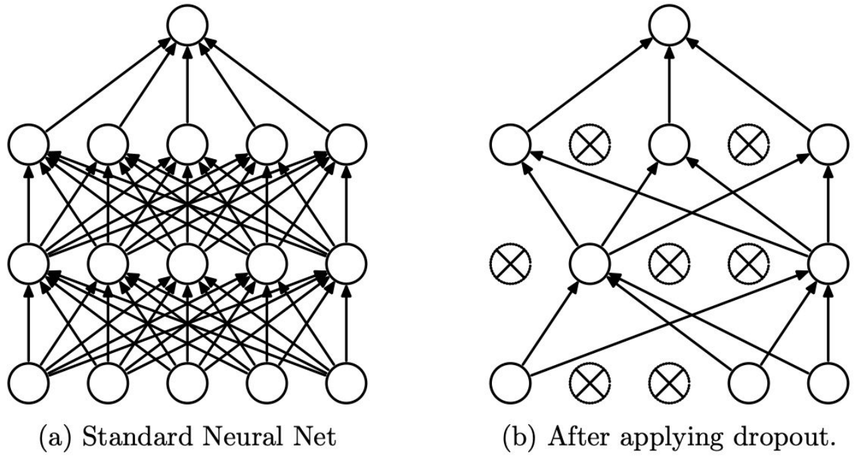
- **New Dataset Construction** ➡ Useful attribute: start and scale

- start: the starting selling date of the product.
- scale: the normalization constant.

- **Model Structure**

- Layers: Input ➡ Embedding ➡ Flatten ➡ Dense ➡ Dropout ➡ Output
- Main Parameter: tensorflow.keras.layers.Dense(500, activation="relu")

  tensorflow.keras.layers.Dropout(0.3)



(a) Standard Neural Net (b) After applying dropout.

ANN Model

- **Model Performance**

- Achieve better performance.

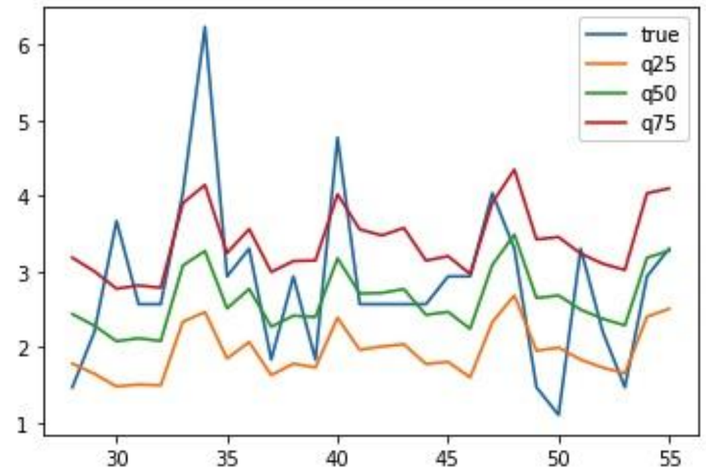
```
In [67]:  1  nett.evaluate(xv, yv, batch_size=50_000)

IOPub data rate exceeded.
The notebook server will temporarily stop sending output
to the client in order to avoid crashing it.
To change this limit, set the config variable
`--NotebookApp.iopub_data_rate_limit`.

Current values:
NotebookApp.iopub_data_rate_limit=1000000.0 (bytes/sec)
NotebookApp.rate_limit_window=3.0 (secs)

Out[67]: 0.21776057838026136
```

- **Quantile Prediction**

- Upper bound -> 0.25

  Lower bound -> 0.75

- Interval confidence is around 80%.



Comparison

	Public Score	Private Score
LSTM	**0.54083**	**0.56559**
ANN	**0.16162**	**0.17087**

In the competition, **Weighted Scaled Pinball Loss** is used for evaluation, so attributes ‘start’ and ‘scale’ in the new dataset helps to improve the score.