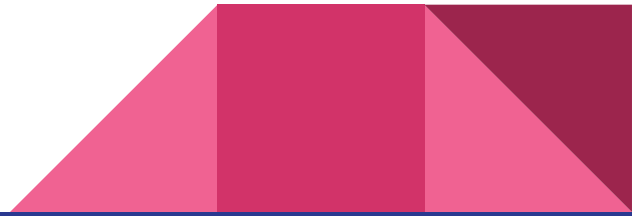
Reflexis Solution

TEAM: Code2Create
TEAM LEADER: Indraneel Ghosh(Contact: 9929862040)
OTHER MEMBER: Siddhant Kundu(Contact: 9829891415)
College: BITS Pilani, Pilani Campus

Problem Statement Interpretation

Machine Learning Interpretation

- Maximise sales: Find ideal system scheduled hours to maximise sales.
- Analyse effect of changes of manager scheduled hours.
- Make predictions using historical data.



Model Used

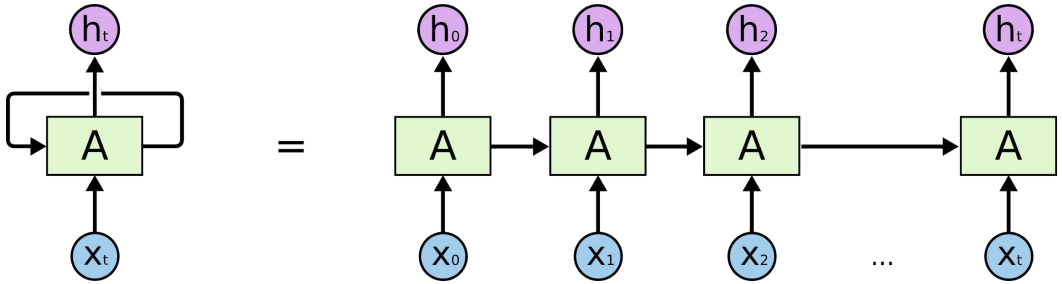
Recurrent Neural Networks(RNN)

Recurrent Neural Networks(RNN)

- RNN (With LSTM(Long Short Term Memory)(An integrated Machine Learning Approach)
- EDA (Study and clean the given dataset, find distribution of data)
- We have reasons to believe that the sales depend on seasonality and trends are expected to be present in the data. RNN(with LSTM) captures these trends the best.

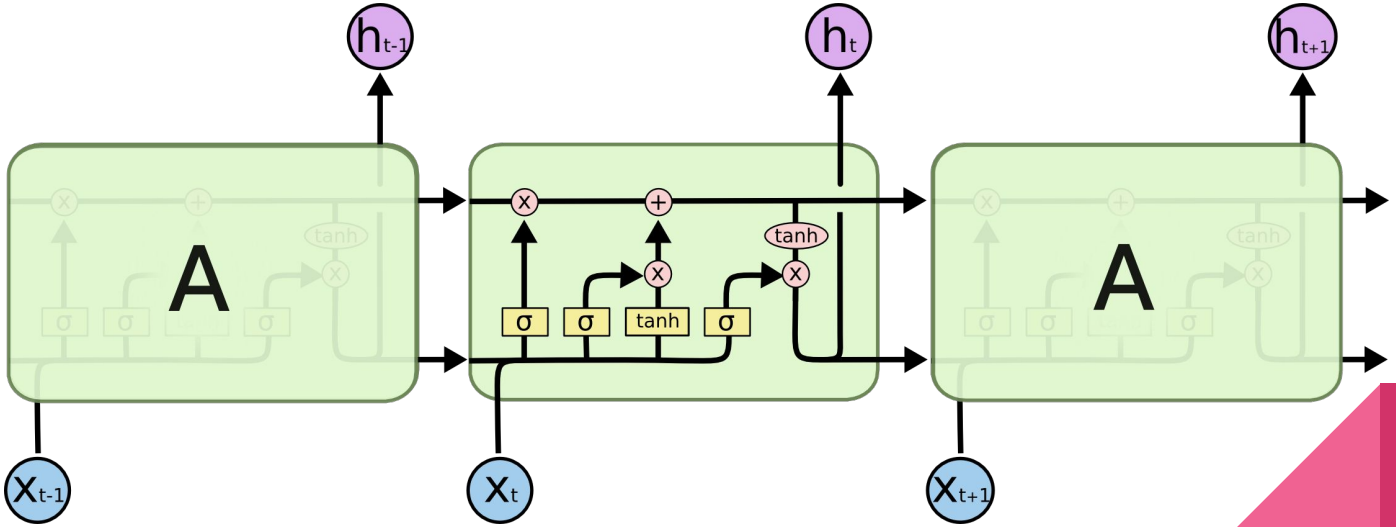
Recurrent Neural Networks

The neural network has previous inputs influencing the earlier outputs., Traditional RNNs are unable to handle “long-term dependencies.”



LSTM(Long Short Term Memory)

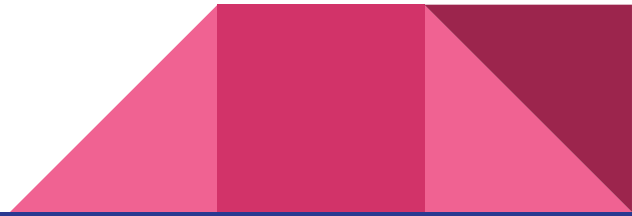
LSTMs have a chain like structure, but the repeating module has a different structure. Instead of having a single neural network layer, there are four, interacting in a very special way.



Assumptions

1. Maximum 12 hour work days

2. Increments made to work hours are in 0.25 intervals



Implementation Details

- Input: Store (Store number)
- RNN Model was designed using keras implementations.
- #Epochs = 50
- Groupby ‘Store’(Data grouped by store)
- Batch Size = 4
- cuDNN used to improve performance and reduce training time for the model.

RNN Structure

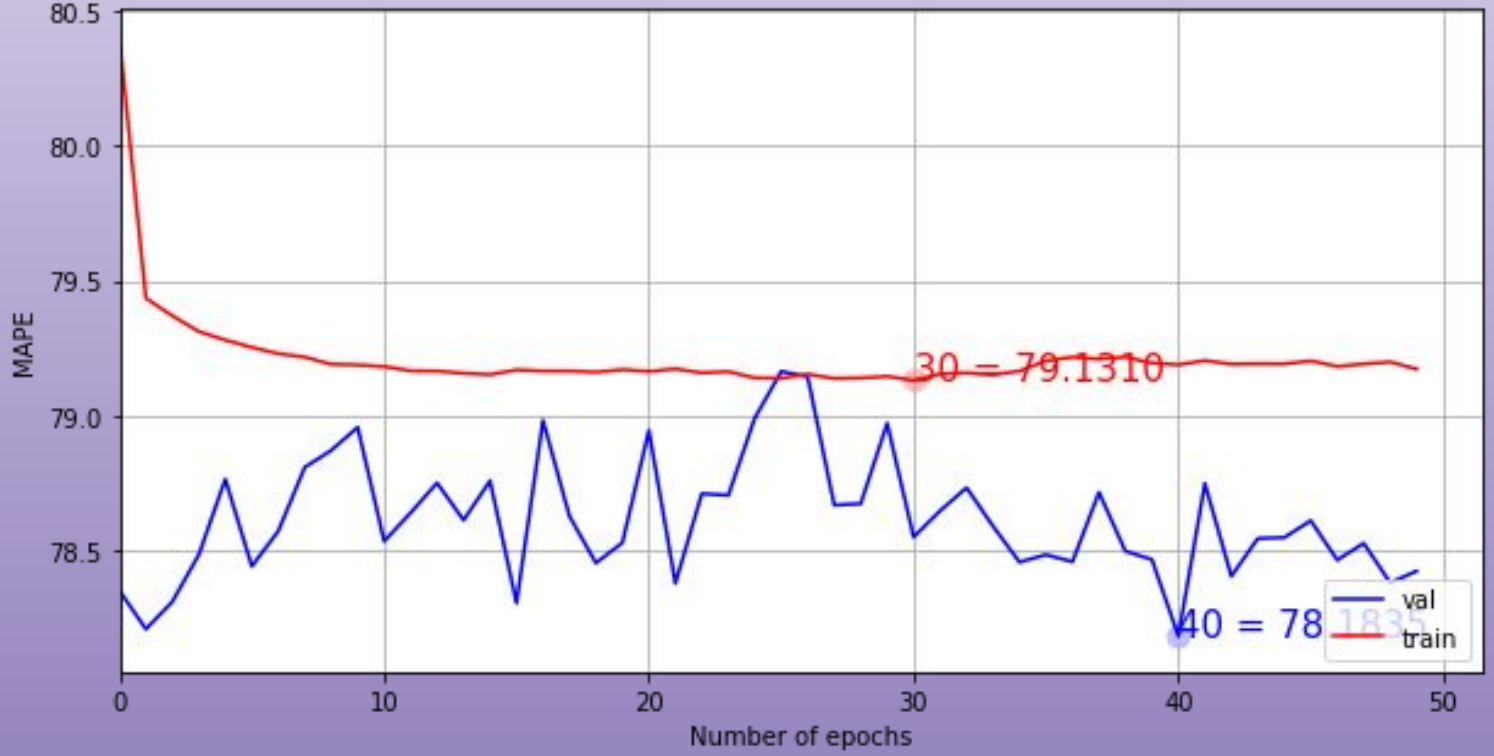
- RNN Has 3 Layers of LSTM with 64 cells in each layer.
- Every layer has a dropout of 0.25 to reduce overfitting.
- Data is fed as a sequential time series. This Time Series contains an instance of the list [<STORE>, <MANAGER_SCHED_HOURS>, <SYSTEM_SCHED_HOURS>] for every timestamp, which for us is every day for which we have a record.
- The output for this is a list giving a scaled version of the predicted value of [‘SALES_ACTUAL], between -1 and 1.

Final Output

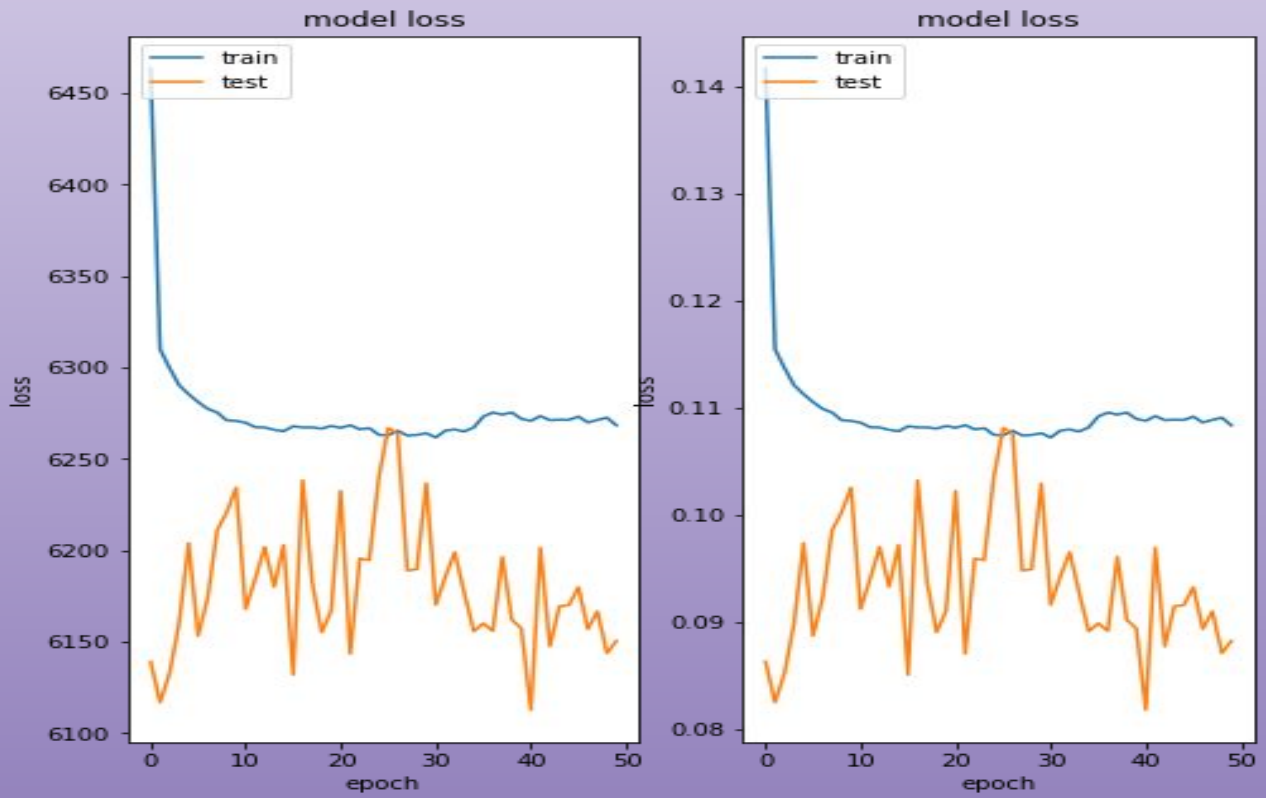
1. A single Dense layer is used at the end for the output generation.
2. Output generated by RNN predicts what the ‘SALES_ACTUAL’ value would be, given the ‘MANAGER_SCHED_HOURS’, ’STORE’, ’SYSTEM_SCHED_HOURS’.
3. We then find the optimal number of scheduled hours so as to maximise profits.
4. We further use this output to determine the effect of changes in manager scheduled hours.

Results(RNN)

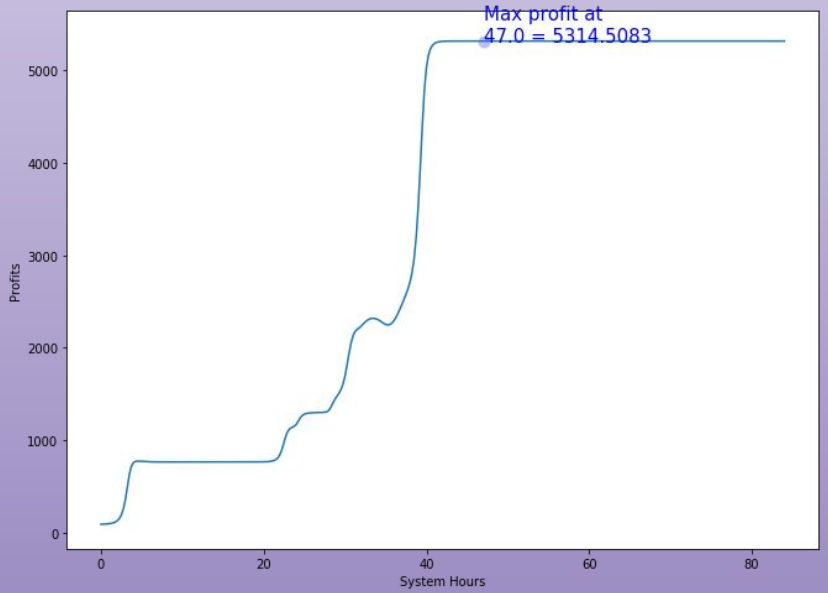
RNN Results

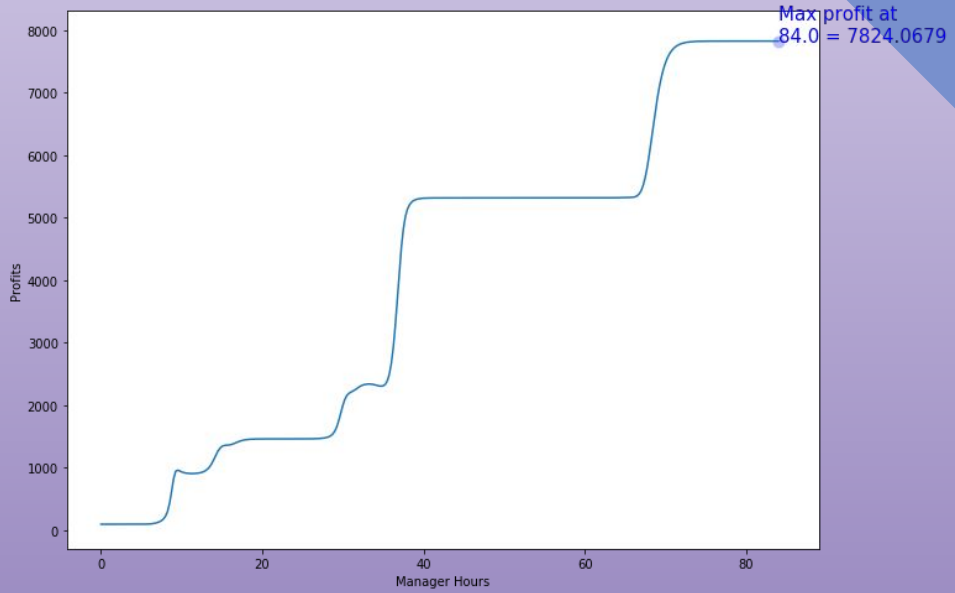

80.5
80.0
79.5
79.0
78.5
MAPE
0
10
20
30
40
50
Number of epochs
30 = 79.1310
40 = 78.1835
val
train

RNN Results


model loss
train
test
loss
epoch
6450
6400
6350
6300
6250
6200
6150
6100
0
10
20
30
40
50
model loss
train
test
loss
epoch
0.14
0.13
0.12
0.11
0.10
0.09
0.08
0
10
20
30
40
50

RNN Results


Max profit at
47.0 = 5314.5083
Profits
System Hours
0
1000
2000
3000
4000
5000
0
20
40
60
80

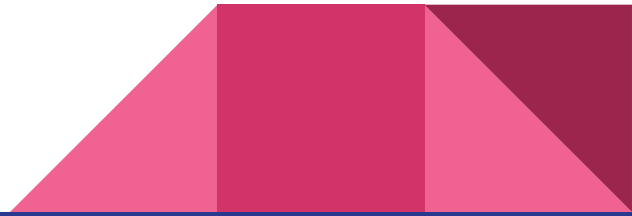

Max profit at
84.0 = 7824.0679
Profits
Manager Hours
0
1000
2000
3000
4000
5000
6000
7000
8000
0
20
40
60
80

Interpretation of Results

Profits maximization condition :

1. System Scheduled Hours = 47 Hours

2. Manager Scheduled Hours = 84 Hours



Our Opinions

- The dataset provided is too small, which is leading to overfitting.
- A larger dataset(100000+ entries for every store) can provide better results. This is because the latent features determining the results are highly dependent on the store.
- Alternative approach we could think of was using SARIMA.(Cons: Your output only depends on time, not on manager or system scheduled hours for such a modelling paradigm)

Additional Reference

Code Base: https://github.com/ighosh98/aic-reflexis.git

[Note: Access would be made available on request.]

Thank You