Predicting MTA ridership with Deep Learning

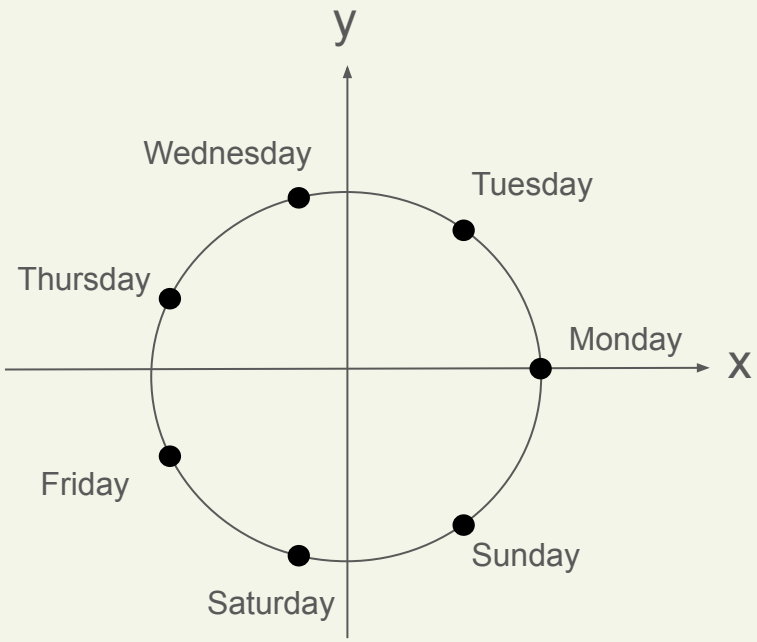
Jack Carlisle, Nick Haubrich

Erdos Institute Deep Learning Bootcamp Fall 2024

Goal: Use deep learning models and MTA open data to predict ridership of the NYC subway system.

Dataset and Feature Engineering

- MTA Subway Ridership Dataset
  - Hourly Ridership of each of 428 MTA Stations.
  - Spans from February 2022 to October 2024.
- Additional Features
  - Encoded hour of day, day of week, and month of year as “Fourier features”.
  - Included flag for National Holidays (Holiday = 1, non-Holiday = 0).



Day of week is encoded as the x and y coordinates of the corresponding evenly spaced points in the image above.

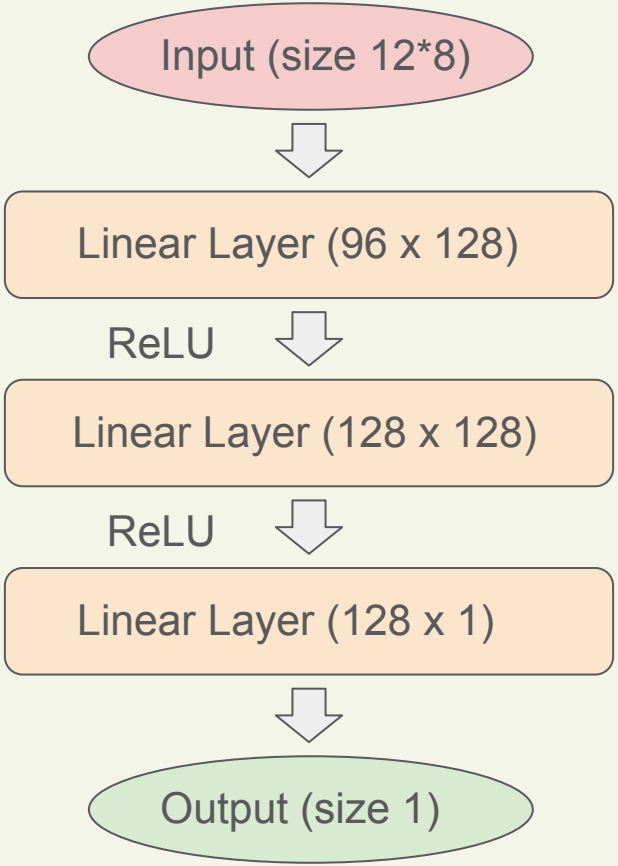
Modeling Total Ridership

Input: 12 hours of ridership data

Output: Next hours ridership

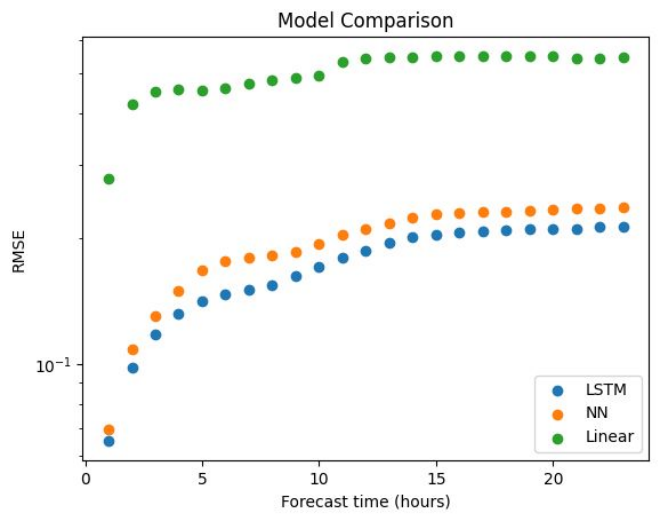
Models:

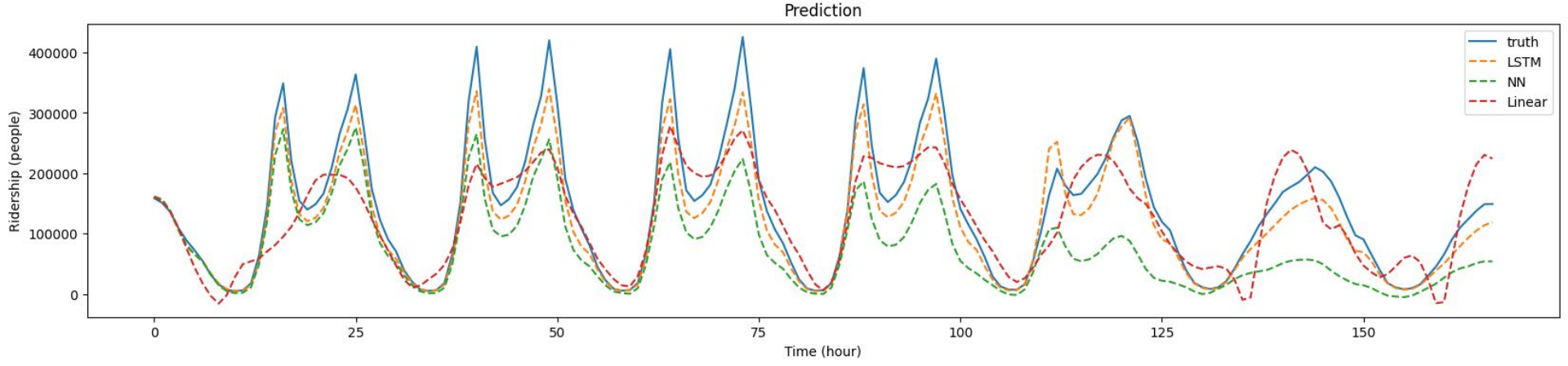
- Linear
  - Single linear layer of size 12*8 = 96.
- Dense Neural Network
  - 3 linear layers of size 128.
  - Activation function = ReLU(x).
- LSTM Neural Network
  - LSTM = ‘Long Short Term Memory’
  - Single LSTM Layer of size 128.



Architecture of Dense Neural Network predicting total subway ridership.

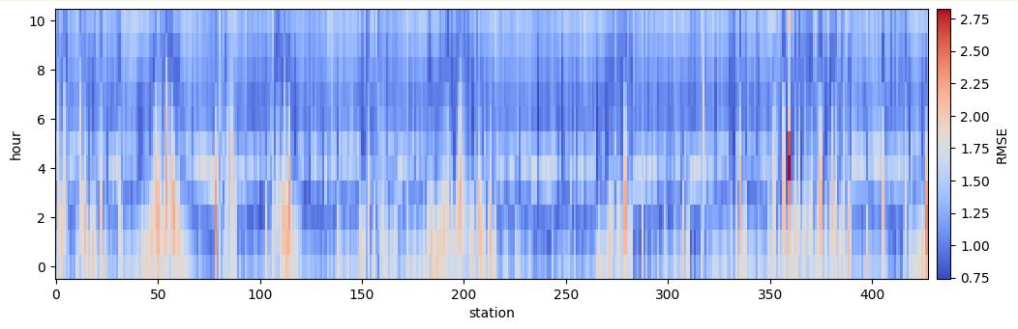
Performance Comparison

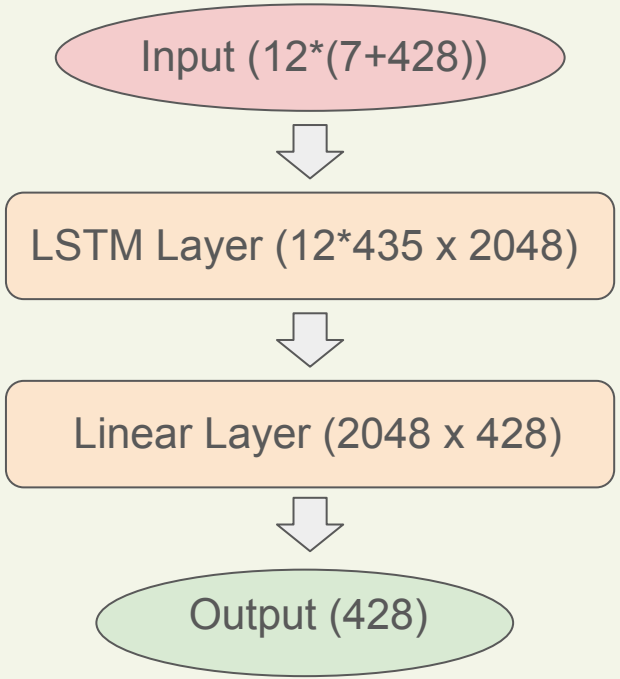




Per-Station LSTM Model

- LSTM Neural Network
  - Single LSTM layer of size 2048.
  - Uses past data from EVERY station to predict ridership of EVERY station.

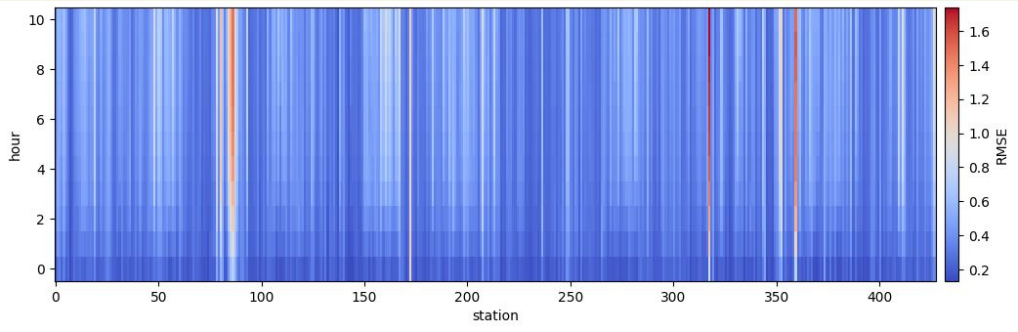


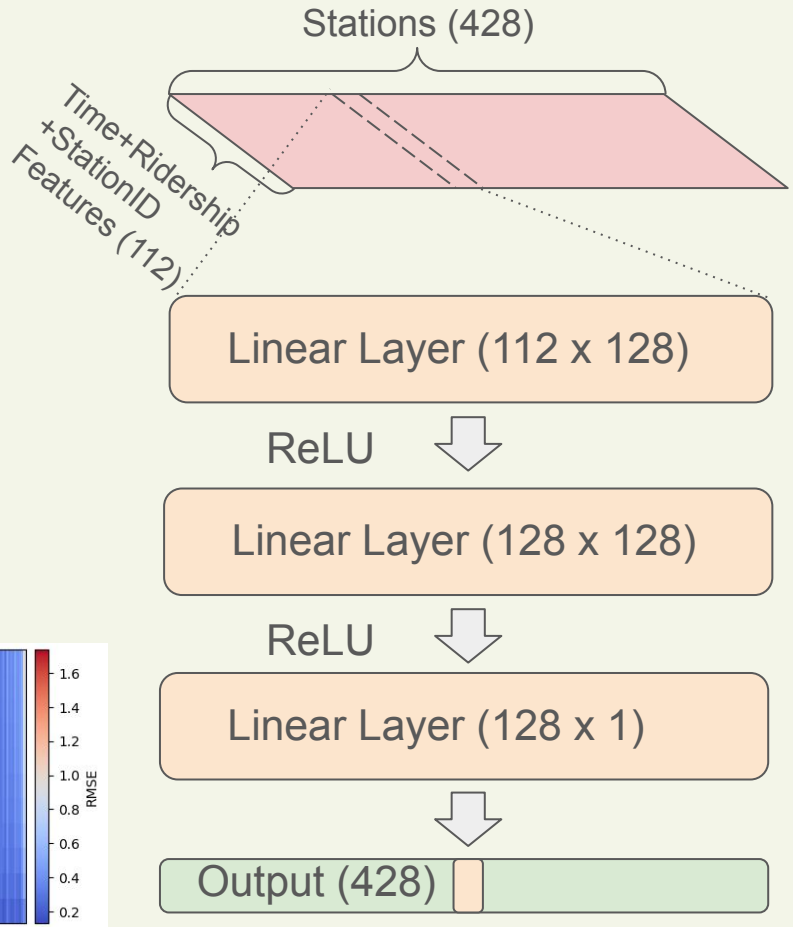


Architecture of LSTM predicting ridership of each station.

Per-Station CNN Model

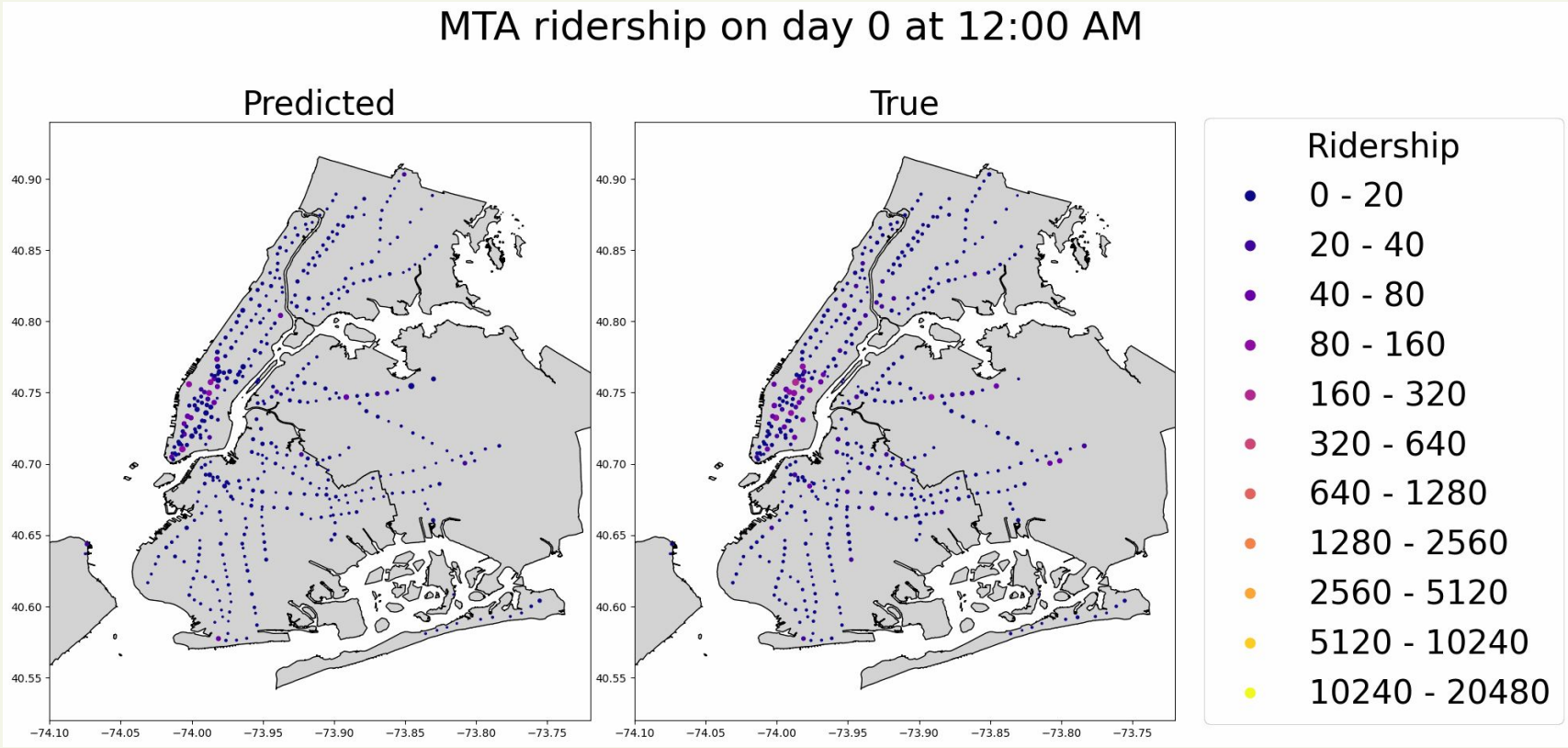
- 1-Dimensional CNN
  - CNN = ‘Convolutional Neural Network’
  - Acts like one dense network applied each SINGLE station.
  - Additional “station identifier” feature encoded as random vectors.





Architecture of 1-Dimensional CNN predicting ridership of each station.

Visualizing our CNN model's predictions



Conclusion: By implementing a variety of deep learning models, we can accurately predict the total and per-station ridership of MTA stations.