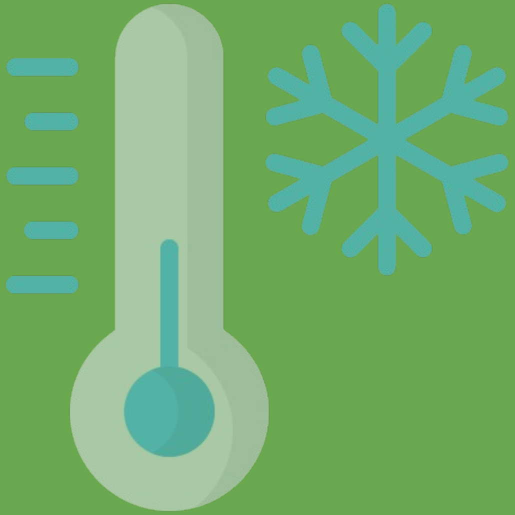
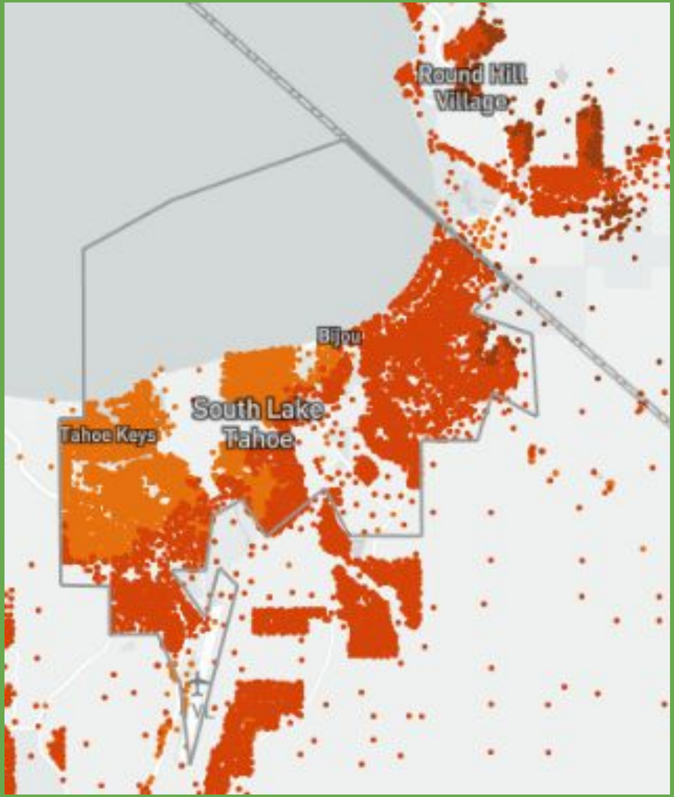
Predicting Temperature Trends in Tahoe



Sid, Joan, Teja, Alex
DATASCI 207, Summer 2024
Instructor: John Santerre PhD

Background

- Research and Environmental Monitoring
- Public Safety and Infrastructure
- Energy
- Economic Impact
    - Transportation Operations
    - Agriculture
    - Recreation



Data Set

- National Centers for Environmental Information Global Climate Database
- Global Forest Watch Data Set
- Processing Steps:
  - Data Wrangling
  - Assessing yearly averages for deforestation
  - Data normalization
  - Merging Datasets
  - Split data

Problem Statement

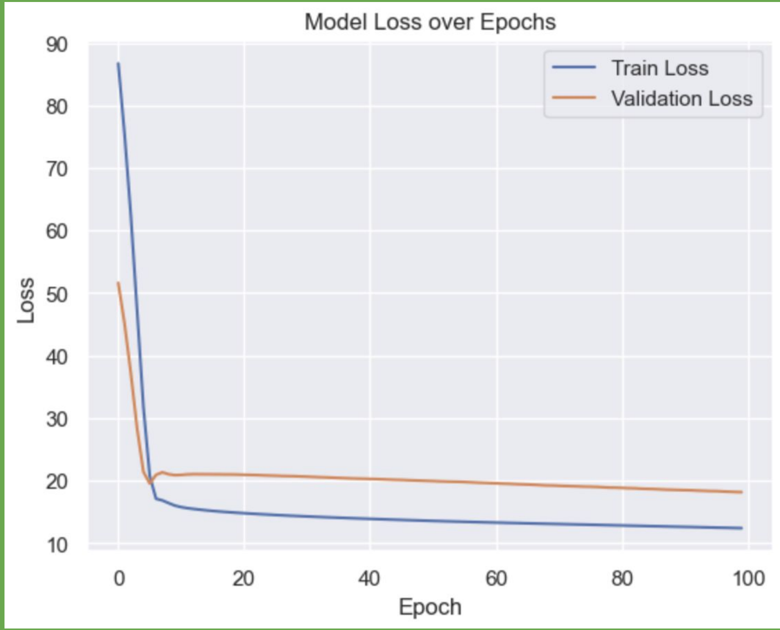
Developing Machine Learning models to accurately predict the average temperature in Tahoe using the NCEI's historical climate data and meteorological factors, with the aim of improving public safety, optimizing infrastructure management, and supporting economic and environmental planning

Baseline models

- Baseline takes in features without a time component to predict `tavg`
- Models evaluated: linear regression, neural network regression, and XGBoost regressor

Baseline loss and RSME

Neural Net



Model Type	Loss
Linear Regression	RSME: 4.7482
Neural net regression	RSME: 3.992
XGBoost regressor	RSME: 3.0593

Time Series using ARIMA

AR

Autoregressive: forecasting future values using a linear combination of previously observed values

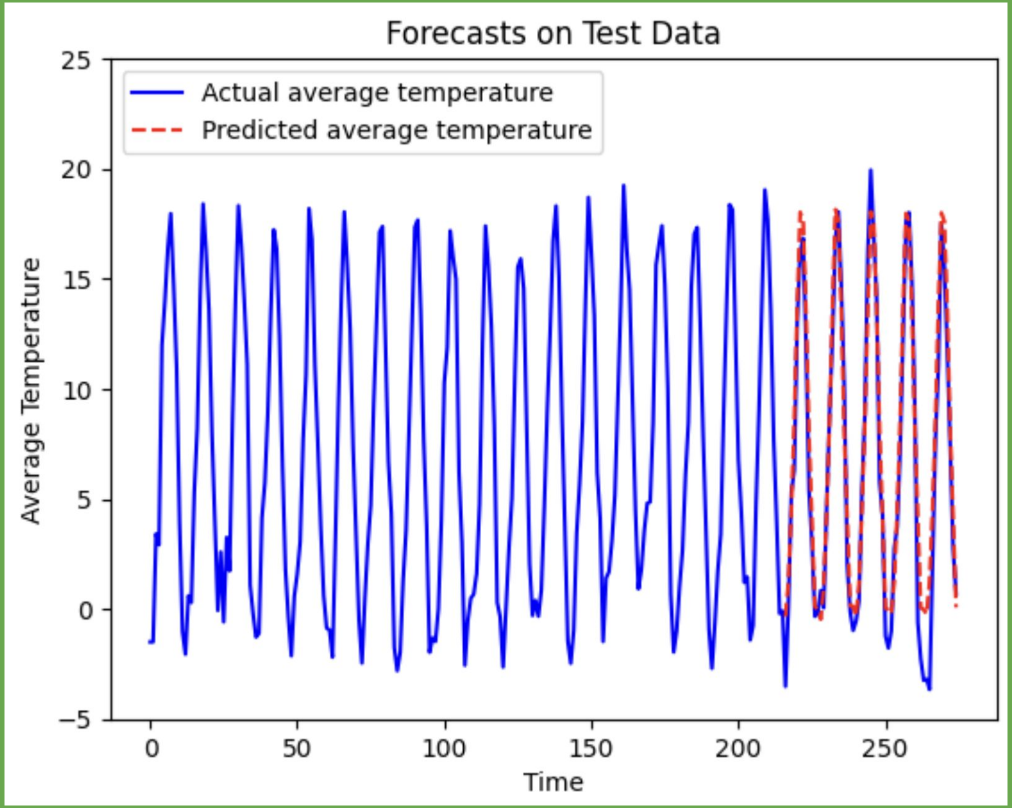
I

Integrated: the number of differencing required to make time series stationary (where the time series has constant variance and mean)

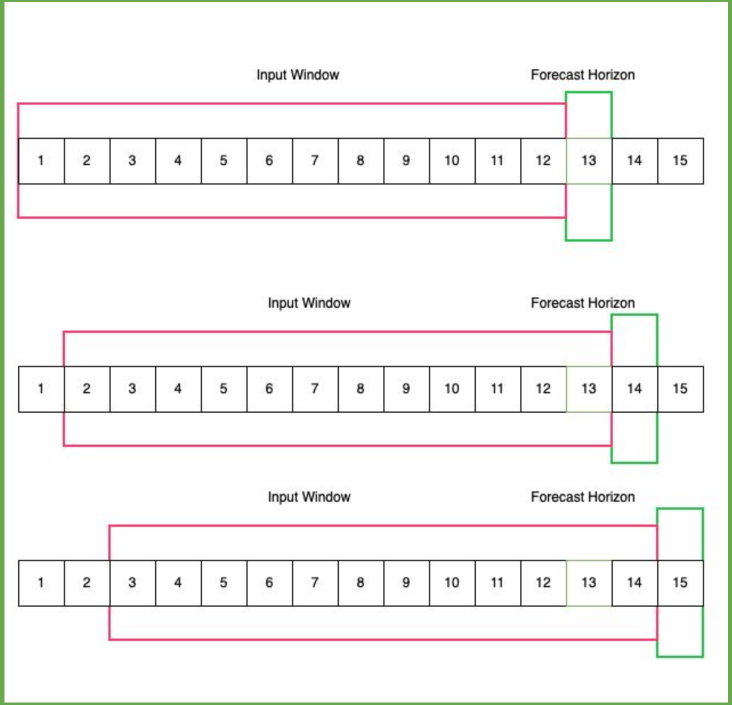
MA

Moving Average: forecasting using past errors instead of actual observed values

Time Series using ARIMA *Loss 1.720*



Time Series using Neural Networks



date	tavg	tmin	tmax	tc_loss_ha_average_scaled
2001-01-01	-1.50	-7.28	4.28	-1.797473
2001-02-01	-1.50	-6.60	3.60	-1.797473
2001-03-01	3.41	-2.42	9.25	-1.797473
2001-04-01	2.91	-3.38	9.21	-1.797473
2001-05-01	11.98	3.67	20.29	-1.797473
2001-06-01	13.97	5.22	22.72	-1.797473
2001-07-01	16.35	7.69	25.02	-1.797473
2001-08-01	17.96	8.41	27.52	-1.797473
2001-09-01	14.68	6.11	23.25	-1.797473
2001-10-01	10.94	2.72	19.15	-1.797473
2001-11-01	3.93	-1.73	9.60	-1.797473
2001-12-01	-1.02	-5.68	3.64	-1.797473

Features (X)

2002-01-01	-2.06	-7.93	3.80	-0.699343

Label (y) == TAVG

Sliding with LSTM

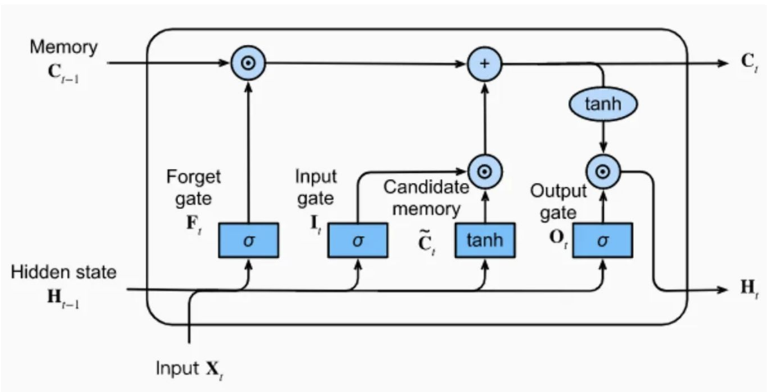


Figure 1: Architecture of a LSTM Unit (Credits: https://d2l.ai/chapter_recurrent-modern/lstm.html)

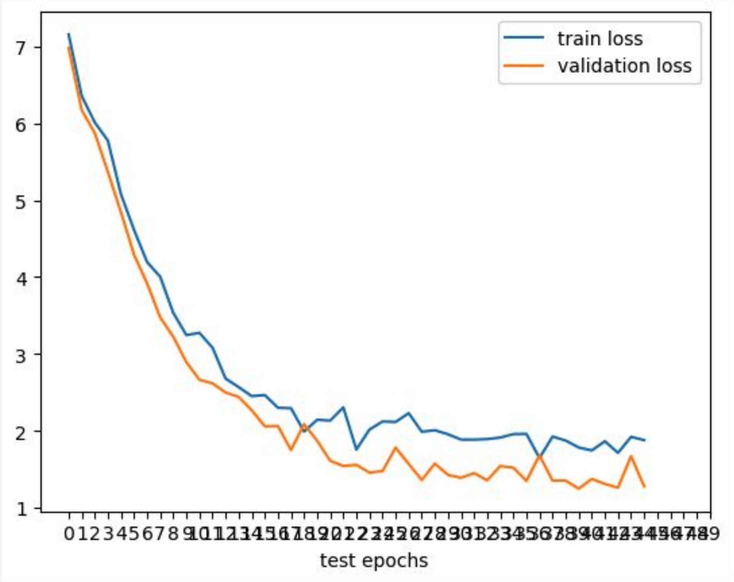
Long Term Trends from LSTM Memory

date	tavg	tmin	tmax	tc_loss_ha_average_scaled
2001-01-01	-1.50	-7.28	4.28	-1.797473
2001-02-01	-1.50	-6.60	3.60	-1.797473
2001-03-01	3.41	-2.42	9.25	-1.797473
2001-04-01	2.91	-3.38	9.21	-1.797473
2001-05-01	11.98	3.67	20.29	-1.797473
2001-06-01	13.97	5.22	22.72	-1.797473
2001-07-01	16.35	7.69	25.02	-1.797473
2001-08-01	17.96	8.41	27.52	-1.797473
2001-09-01	14.68	6.11	23.25	-1.797473
2001-10-01	10.94	2.72	19.15	-1.797473
2001-11-01	3.93	-1.73	9.60	-1.797473
2001-12-01	-1.02	-5.68	3.64	-1.797473
2002-01-01	-2.06	-7.93	3.80	-0.699343

Short Term Trends from Hidden State

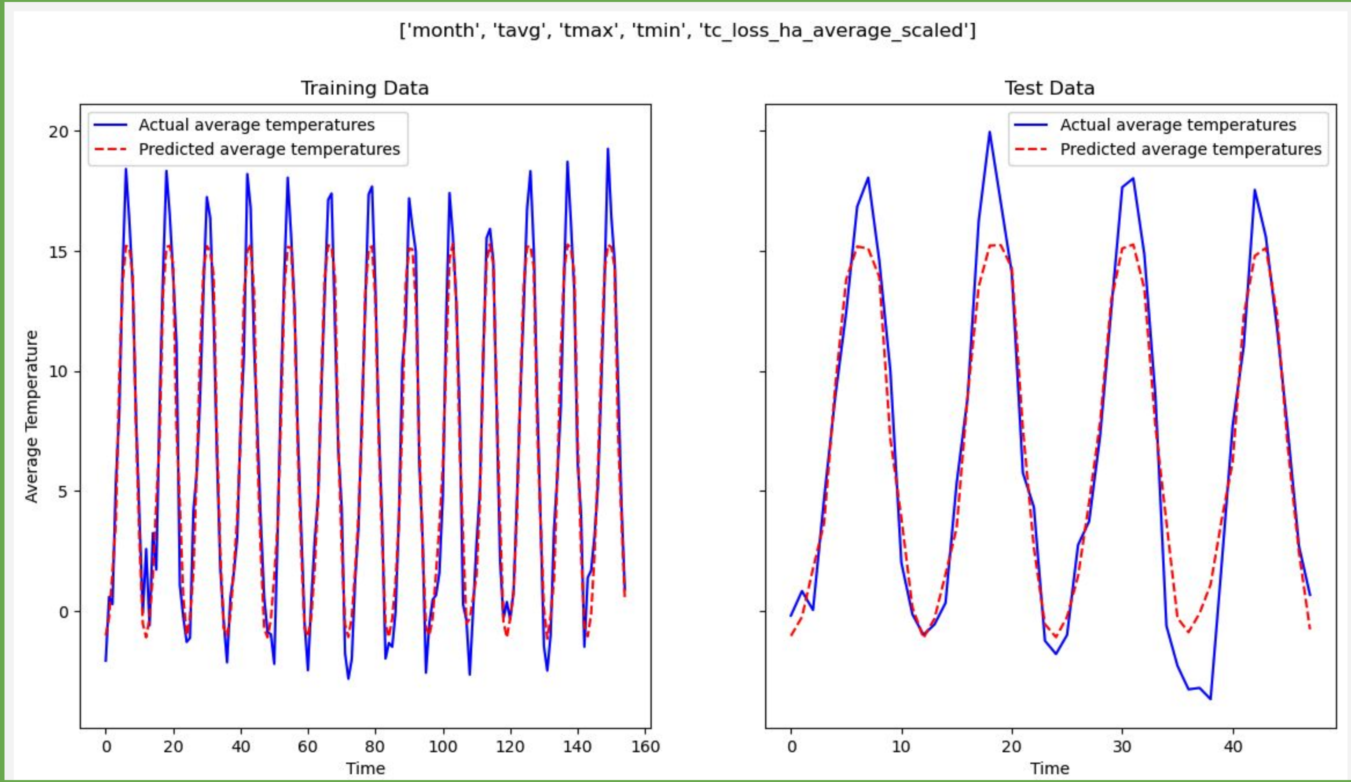
Neural Network using LSTM

```
Model: "sequential_6"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm_12 (LSTM)              (None, 12, 8)             448

 dropout_12 (Dropout)        (None, 12, 8)             0

 lstm_13 (LSTM)              (None, 8)                 544

 dropout_13 (Dropout)        (None, 8)                 0

 dense_11 (Dense)            (None, 1)                 9

=================================================================
Total params: 1,001
Trainable params: 1,001
Non-trainable params: 0
_________________________________________________________________
```




```
At last epoch, training loss: 1.8788
At last epoch, Validation loss: 1.2780
5/5 [==============================] - 1s 2ms/step
2/2 [==============================] - 0s 3ms/step
```

Training and Evaluation




```
5/5 [==============================]
Train Loss: 1.1989307403564453
2/2 [==============================]
Test Loss: 1.5361968278884888
```

Comparing All Models

Model Type	Loss
Linear Regression (Base)	RSME: 4.748
Neural net regression (Base)	RSME: 3.992
XGBoost regressor (Base)	RSME: 3.059
ARIMA	RMSE: 1.720
LSTM	RMSE: 1.536

Conclusion and Future Steps

- Limited NCEI data (only about 240 training examples).
- The Global Forest Data had only a yearly forest cover loss granularity.
- Predicting weather is difficult and depends on various factors far beyond the features and data we used in our modeling
- Model works only with one location: Tahoe City