

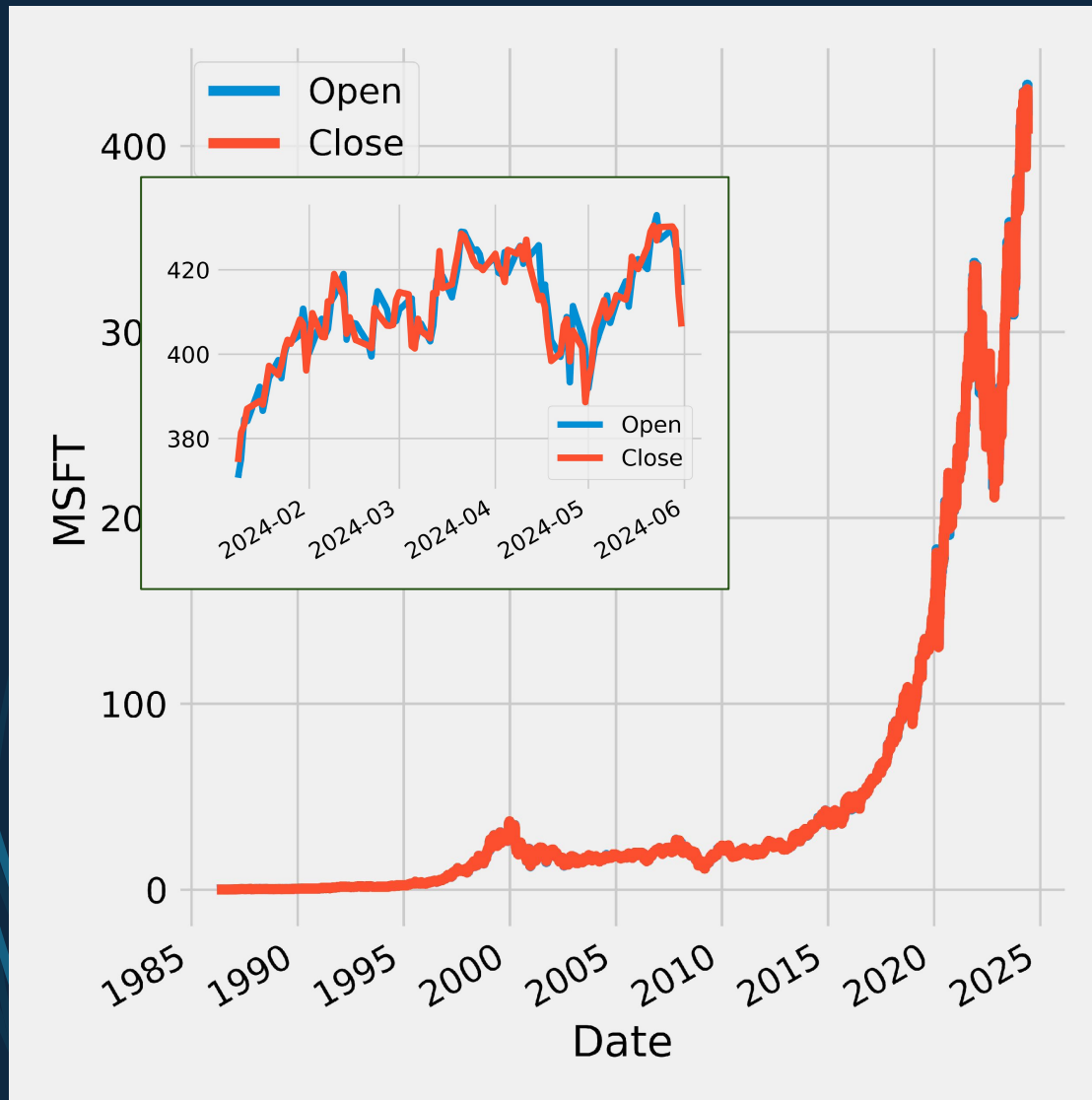


Stock market modeling and forecasting

Suman Aich, Nafis Fuad, Siu Cheung Lam, Xiaoyu Wang

ErDOS Institute Data Science Boot Camp Summer 2024

Contents



- Stock market analysis
 - Trends at long timescales
 - Volatility at short timescales
- Modeling and forecasting
 - Baseline model, rolling average, exponential smoothing, ARIMA
 - RNN with LSTM cell, what do we do that's new?
 - Quantile regression
 - Does stock open higher or lower than prior day?
 - Exploratory:
 - Multiday forecasting
 - Concatenating stocks
- Summary and outlook

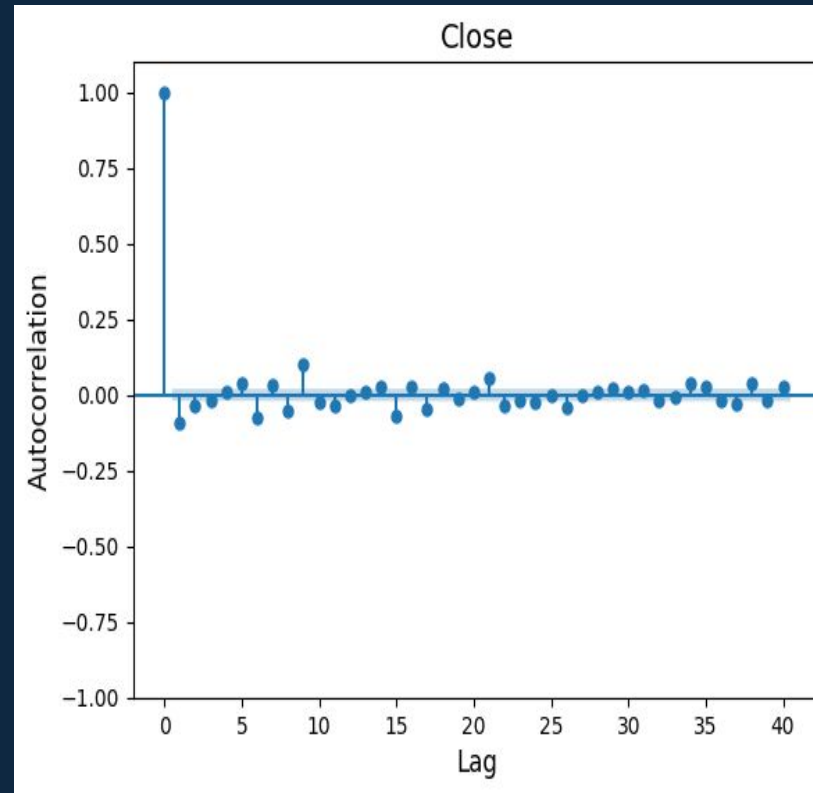
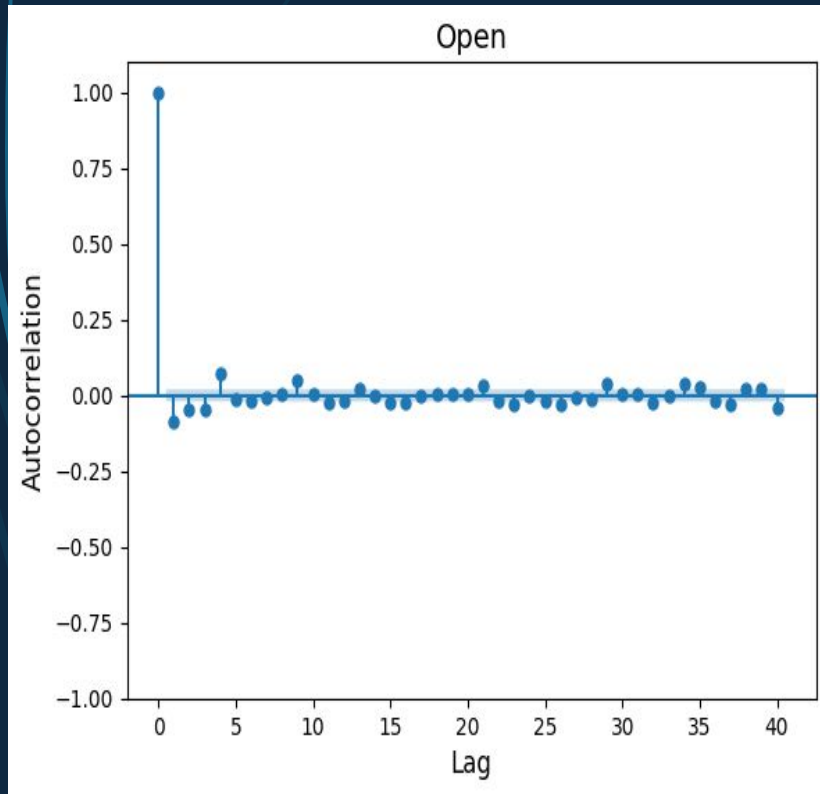
Stakeholders

- Day trading services such as Fidelity Investments, Interactive Brokers, Charles Schwab, etc.
- Long-term investors



Exploratory data analysis

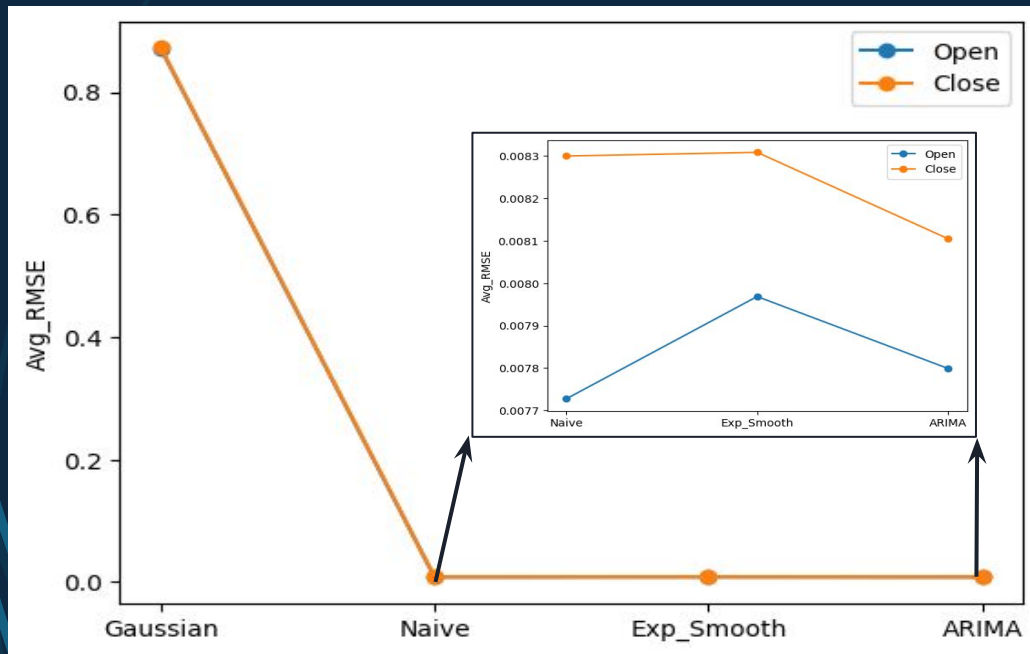
- We plot the autocorrelation of the first difference for the opening and closing Microsoft stock prices at various lags.



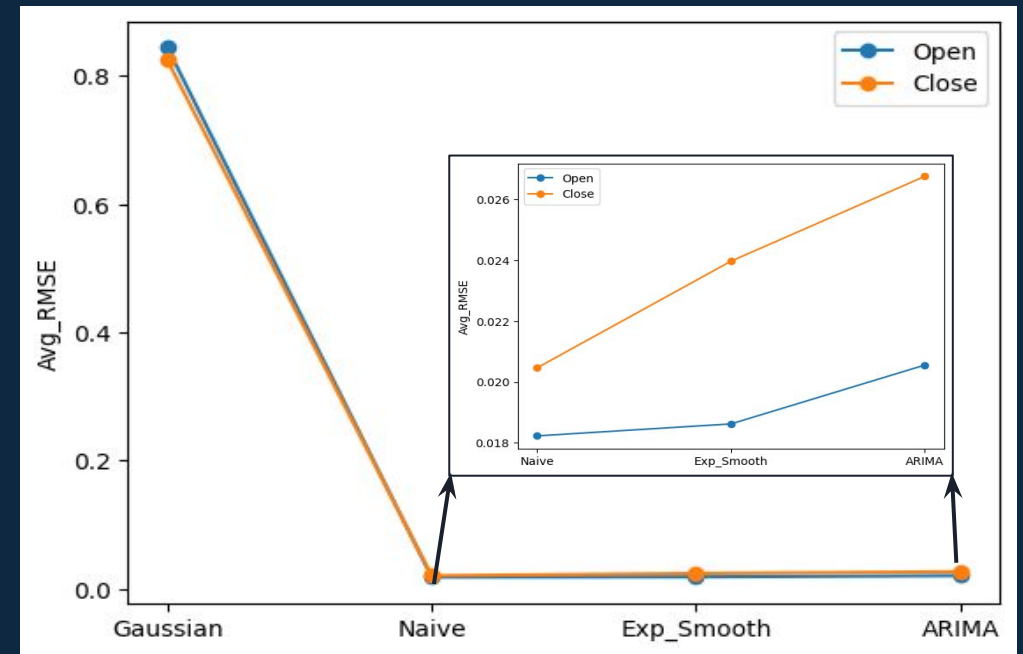
- We clearly see that there is no trend or seasonality and the data is truly stationary

Statistical Modeling

- We implement Gaussian rolling average, naive forecast, simple exponential smoothing and ARIMA models.
- Our Key Performance Indicator is Root Mean Squared Error.

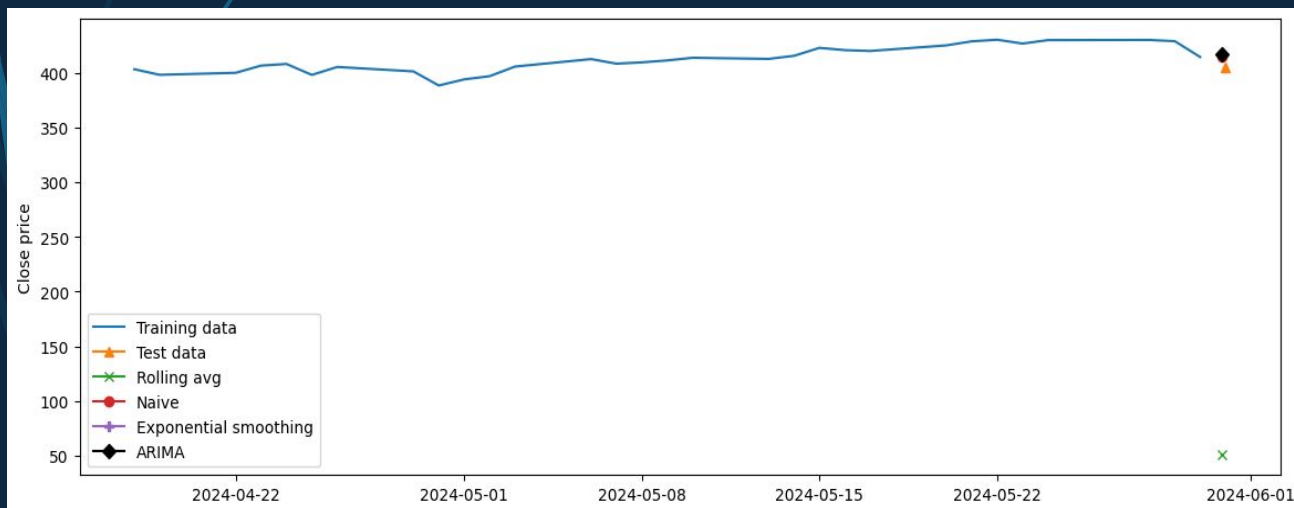
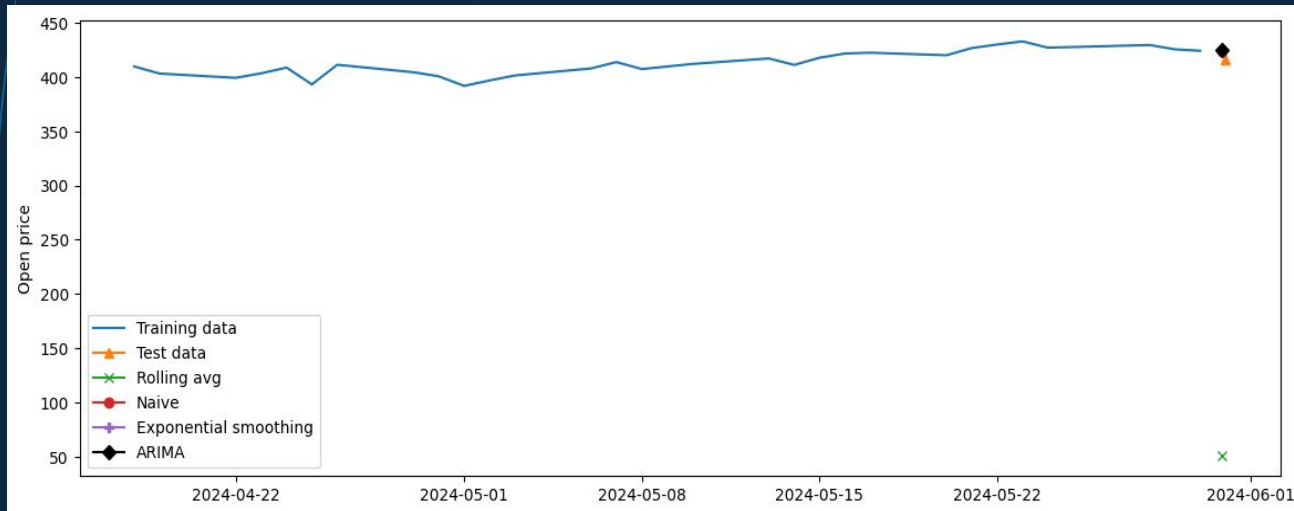


Training data



Test data

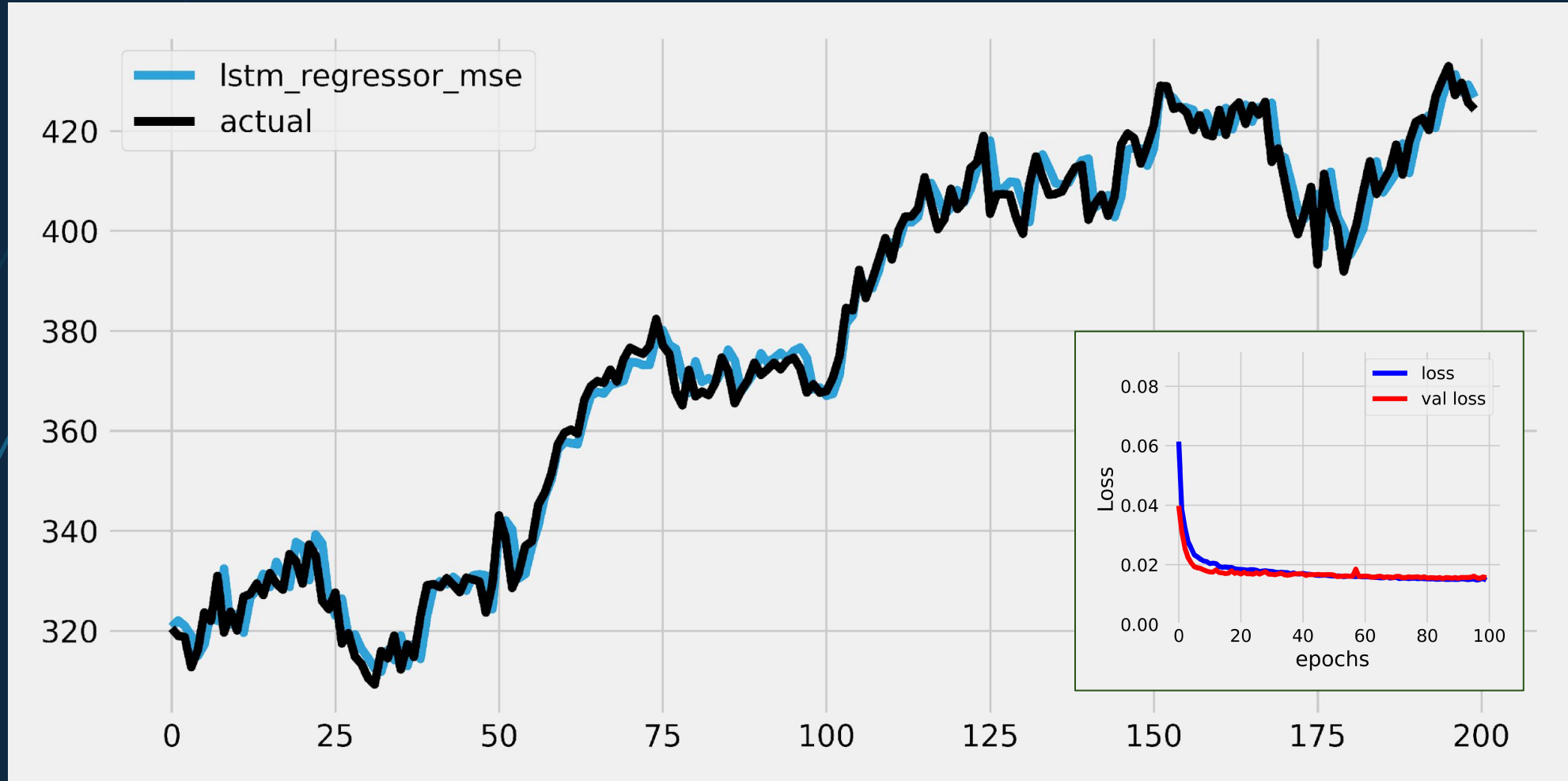
One day forecasts



- We can clearly see that the naive forecast, exponential smoothing and ARIMA models perform equally good and much better than the Gaussian rolling average model.
- This is also evident from the RMSE.

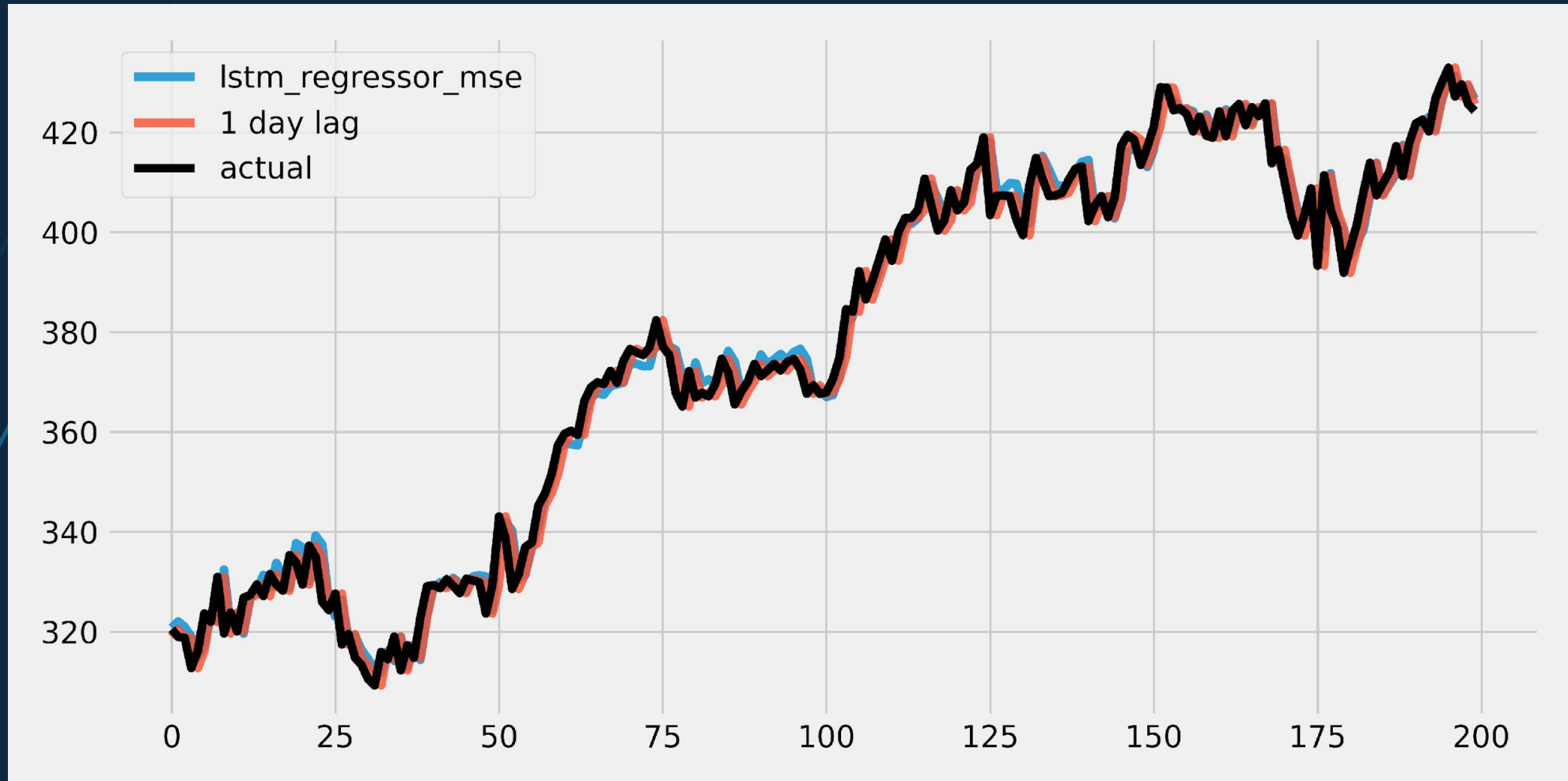
RNN regressor with LSTM cell

- Task: use 40 days of MSFT open price to predict the opening price of 41st day



Is LSTM regressor useful?

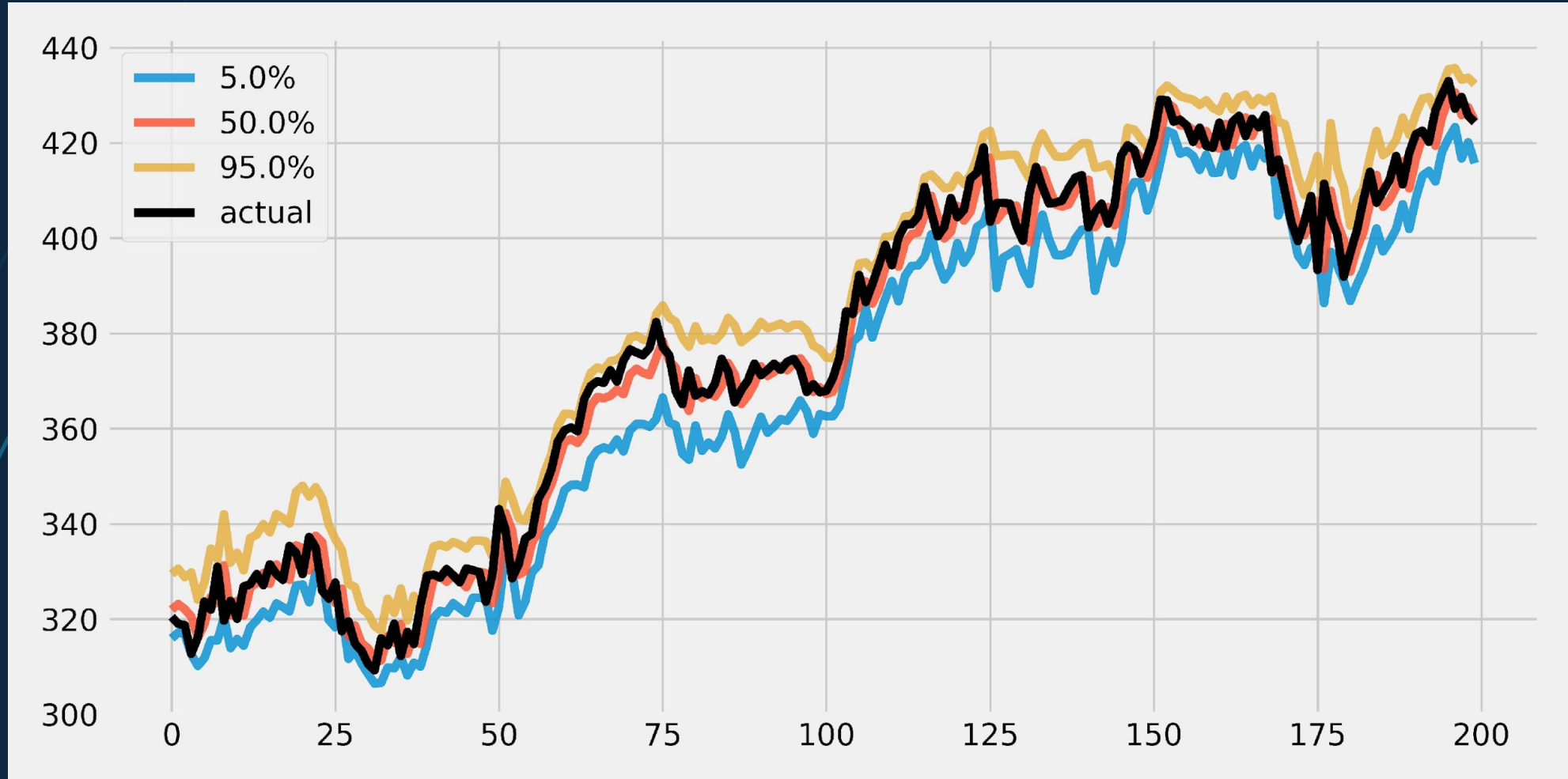
- LSTM regressor appears to give excellent modeling of the market movement
- However, the predicted price appears to be a lagged curve of actual price!



LSTM regressor quantile loss

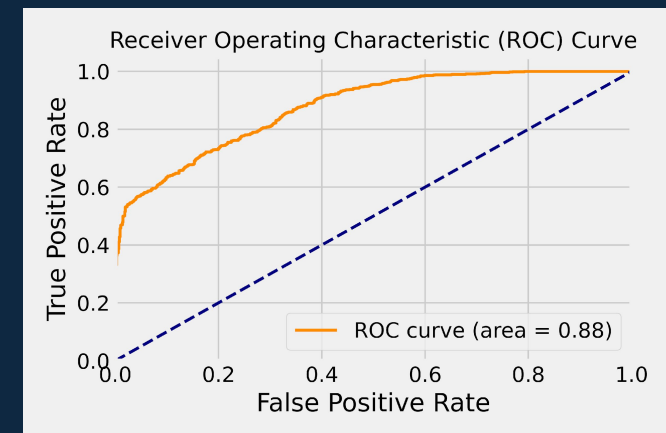
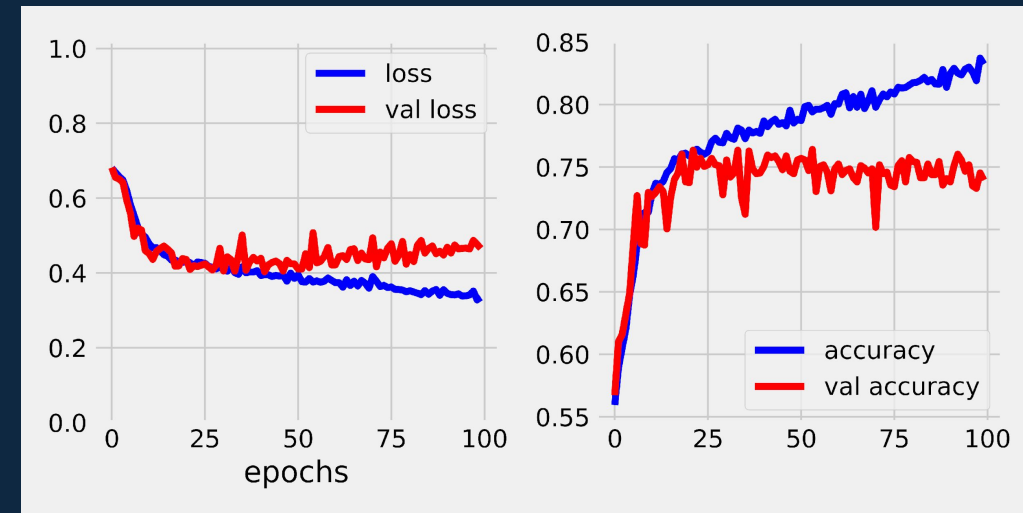
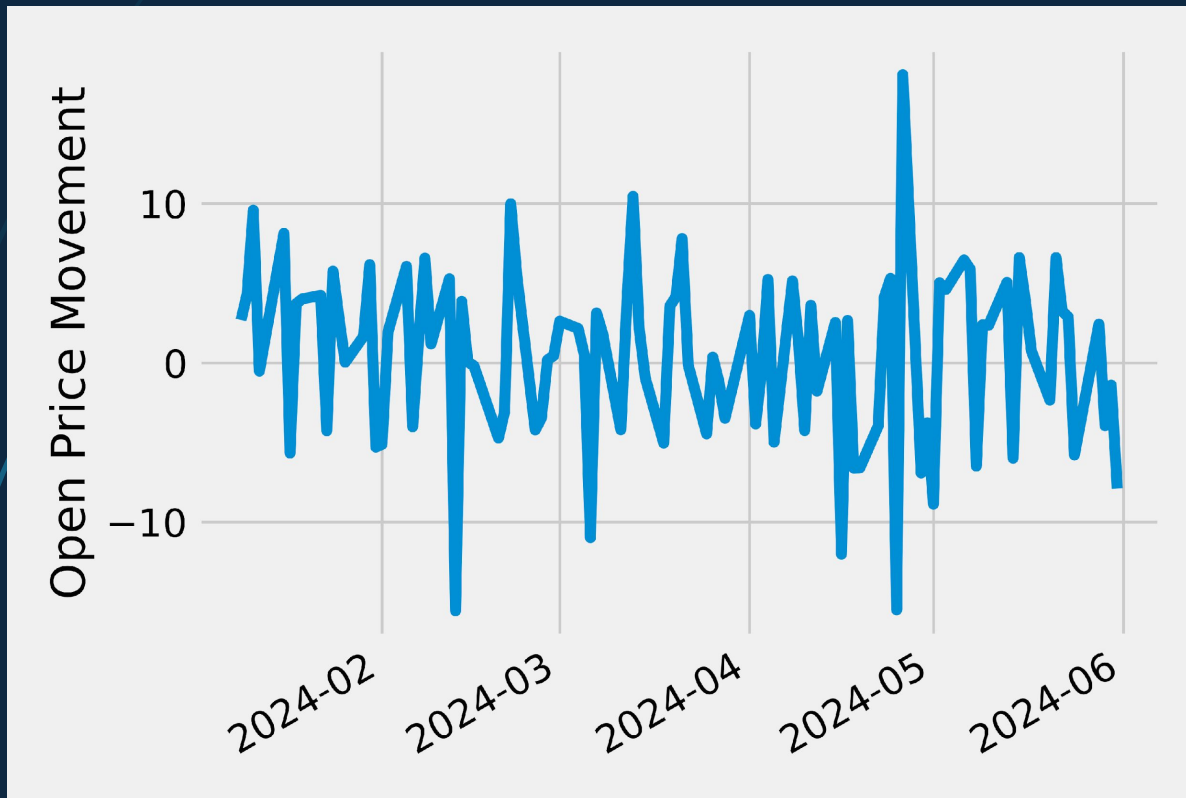
Reference for quantile loss:
<https://www.evergreeninnovations.co/blog-quantile-loss-function-for-machine-learning/>

- Task: use 40 days to predict a confidence interval of opening price of 41st day



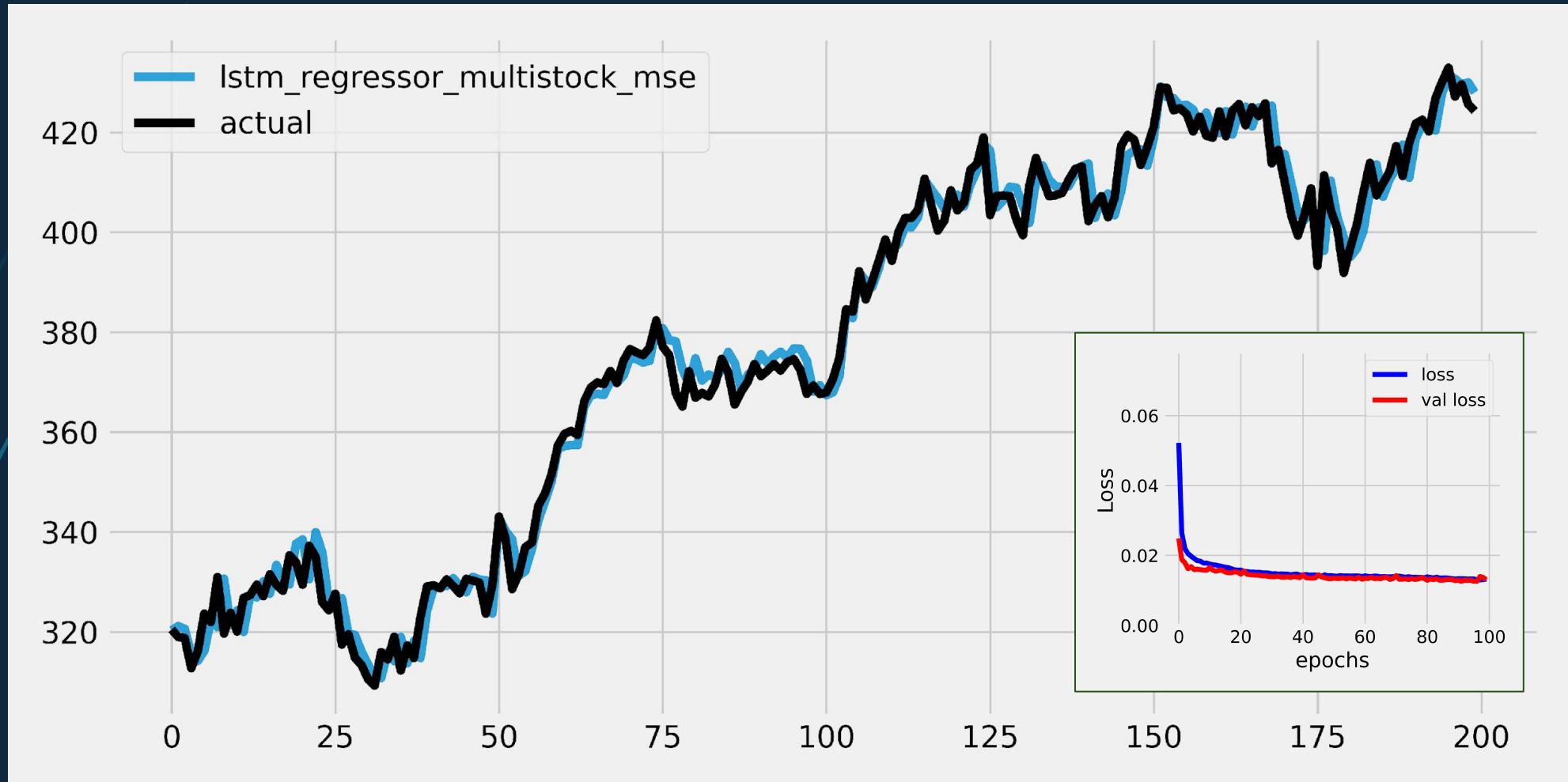
LSTM classifier

- Task: predict if opening price on day $n+1$ is higher or lower than on day n
- Data: ~ 51% chance of price moving up
- LSTM classifier does much better than random guess!

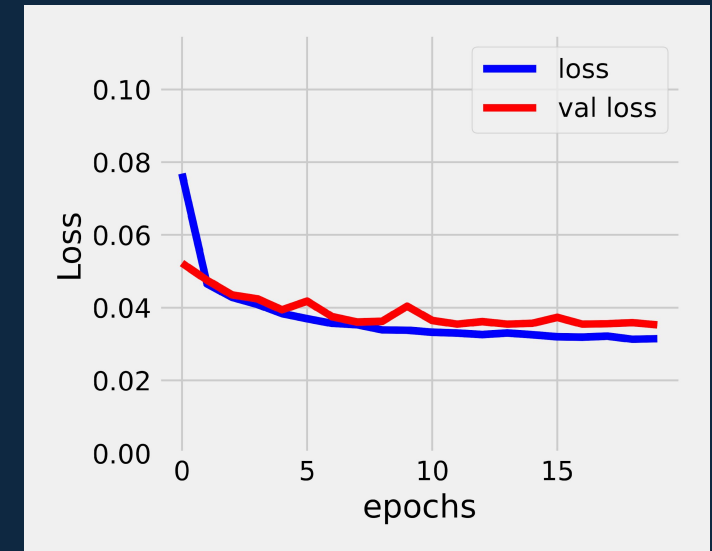
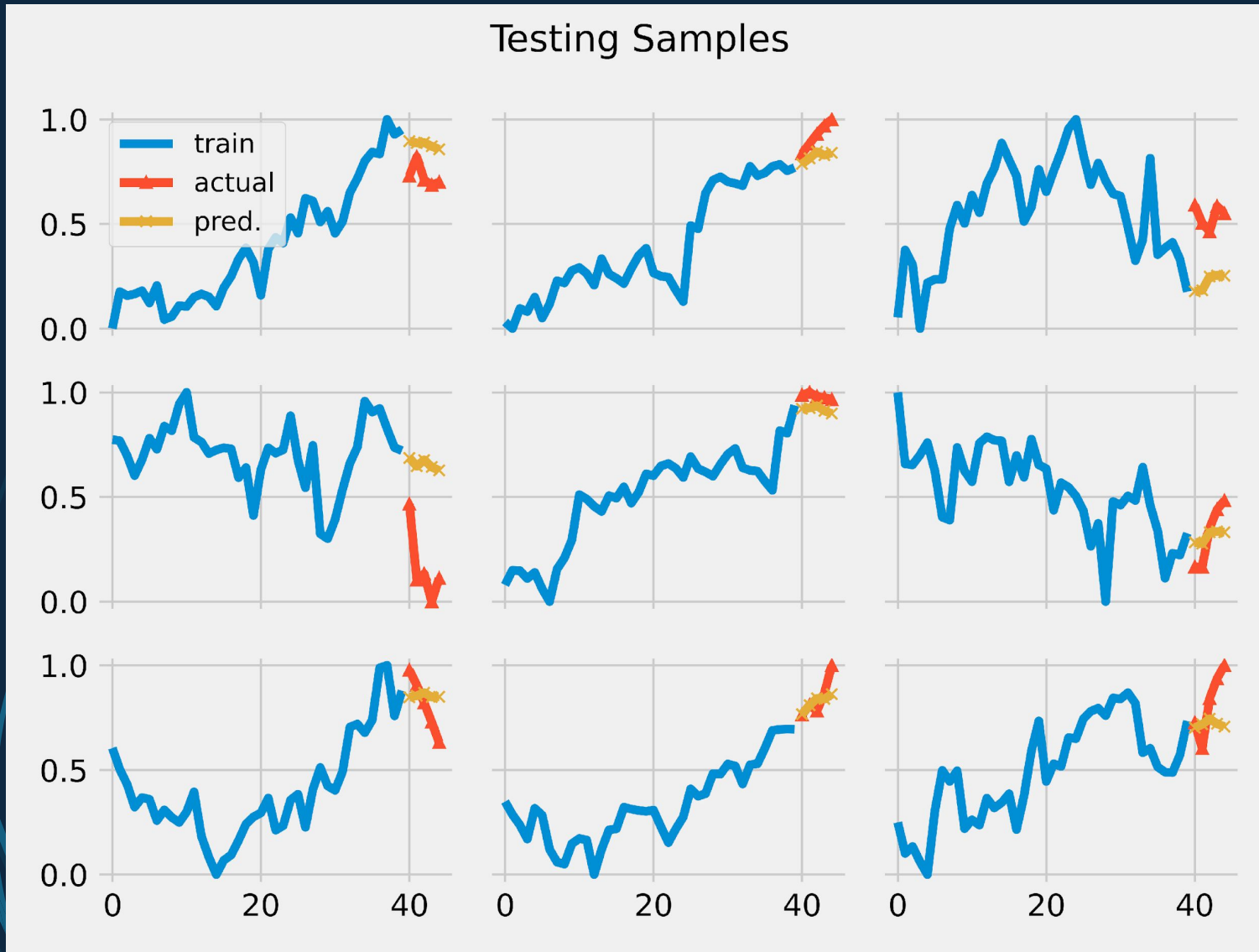


LSTM regressor combining stocks

- Task: combine opening prices of various stocks as features to train generic regressor



LSTM regressor for multiday forecasting



- Regressor severely underestimates market volatility

Summary and Outlook

- Statistical modeling useful for checking trends and quantifying metrics.
- LSTM useful for predicting market movement.
- Can we train a multiclass classifier?
- Does adding market indicators help with classification accuracy (ex. sentiment analysis or technical indicators such as relative strength index)?

Acknowledgement

We would like to thank *Dr. Steven Gubkin* and *Dr. Alec Clott* for the wonderful lectures and guidance.

We would also like to thank *Erdos Institute* and in particular *Prof. Roman Holowsky* for giving us the opportunity to apply our knowledge and work on a project.

And, special thanks to *Aditya Chander* for mentoring us throughout the project and providing useful insights and feedback.