CLASSIFICATION BASED FINANCIAL MARKET PREDICTIONS

Deep Neural Networks

Diego Klabjan

Professor, Industrial Engineering and Management Sciences

Founding Director, Master of Science in Analytics

Northwestern | McCORMICK SCHOOL OF ENGINEERING

Team

- Research in
  - Deep learning (models, algorithms, analysis)
  - Natural language processing (ontologies, knowledge management)
  - Machine learning (robust PCA, large-scale optimization)
- Director, Master of Science in Analytics
  - Full-time single cohort program
  - All aspects of analytics, data science, artificial intelligence
- Joint work with Mark Harmon
  - Ph.D candidate
  - Lida Zhang, Research Assistant

MS in Analytics at Northwestern

- Full-time
  - On-campus in Evanston
- Fifteen months to complete
  - Sept 2016 to Dec 2017
- Small cohort
  - Maximum 40 students
- From SQL to machine learning to business

- Practical
  - Two company sponsored project
  - Summer internship
- Rigorous
  - Design of algorithms in models
  - Implementations
  - Artificial Intelligence

Motivation

Long-term Goal

- The trader bot
  - Automatically trade
- Precursor
  - Predict movement/direction of prices
- Is the stock price
  - Going up/down by more than
    - 10% or one standard deviation
    - In the next short period of time
- Challenge
  - Correlation of assets

http://rubybot.blogspot.com

Strategy Engineering

How can we engineer a strategy producing buy / sell decisions ?

Are securities going

- up by 10%
- down by 10%
- unchanged

in the next few minutes?

Strategy Engineering

Are securities going

- one and two standard deviations
- more than two standard deviations
- unchanged

in the next few minutes?

Traditional Pipeline

Raw features — 5 minute mid-prices

Label

- Lagged price differences from 1 to 100
- Moving price averages with window size from 5 to 100
- Pair-wise correlation of returns

Labels
- Positive
- Neutral
- Negative market returns

Engineered features → Normalized features → Feature selection → Labeled feature set

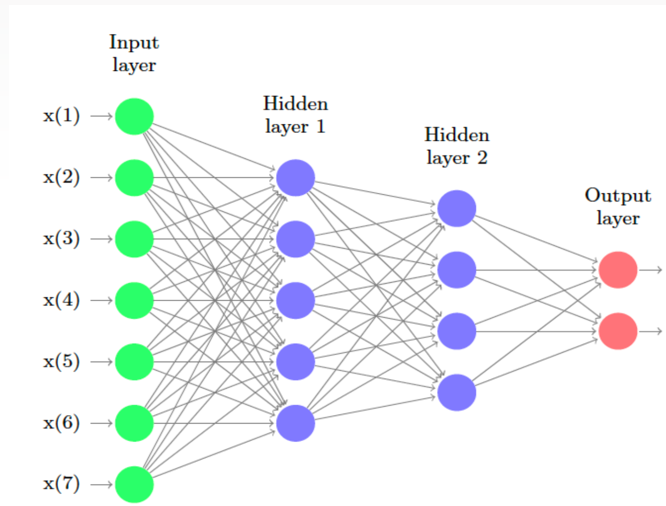
Feature set consists of 9,895 features

Select 100 features

Old vs New

- Traditional
  - 100 features

- Deep learning
  - Use all features



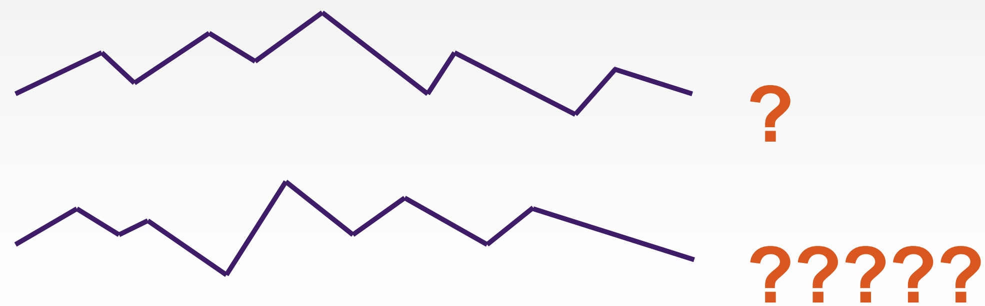
Concepts

- Feature vector at time timestamp t
  - Prices
- Model has to capture temporal dimension
  - Interactions
    - In time and among assets modeled
- Temporal aspects
  - Recurrent neural networks
- Feature selection
  - 1 dimensional convolutional neural networks

10.5
45.1
52.1
46.3
15.3

Goal

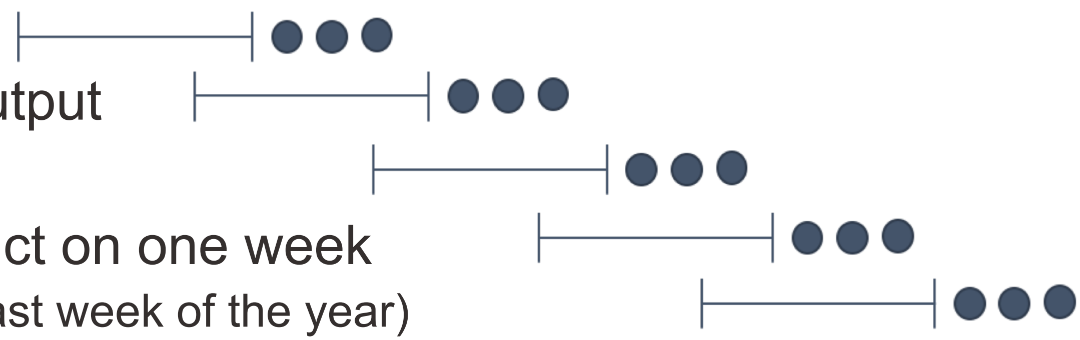
- Predict security price given
  - Past prices
  - Prices of other securities
- Five labels for each security
  - No difference in price
  - Small increase/decrease (with a standard deviation)
  - Large increase/decrease (more than a standard deviation)
- Varying prediction time horizon



Models

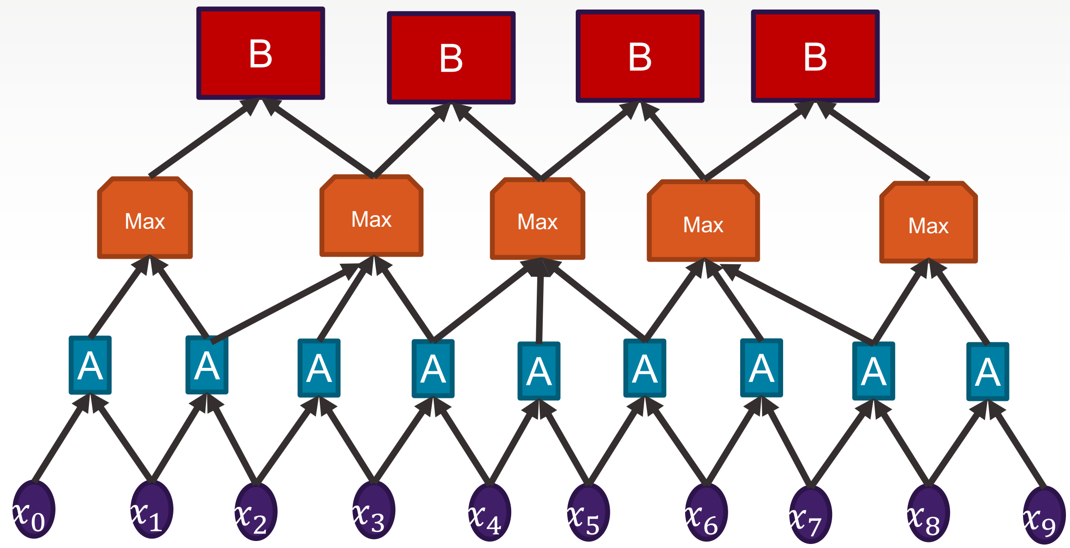
Characteristics

- Difficulties of financial time series
  - Correlations
  - Drift within the series
  - Infinite time series
    - Standard approaches assume finite sequences
  - Sequence length undetermined
- Our approach
  - Vary sequence length on output
  - Overlap sequences
  - Train on one year and predict on one week
    - Walk forward (validation is last week of the year)



Convolutional Neural Networks

- Learns filters
  - Short sliding windows
  - Weighted sums
    - One for each filter
- Advantageous for
  - Observing groups of data close in time
    - Model captures moving averages, etc
  - Observes similarly to humans on images

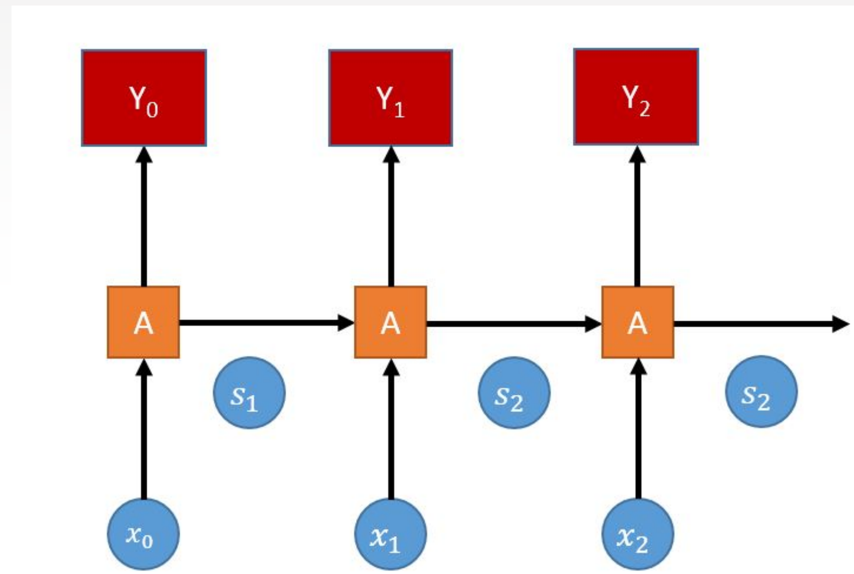


Input Features

Stock	δ_1	δ_2	δ_3	δ_4	δ_5
Stock 1	0	-0.27	0	0.25	0.25
Stock 2	0.25	0	0	-0.52	0
Stock 3	0	0.25	0.5	0	-0.25
Stock 4	0	0	0	0.25	0
Stock 5	-0.25	-0.28	0.5	0	0.25

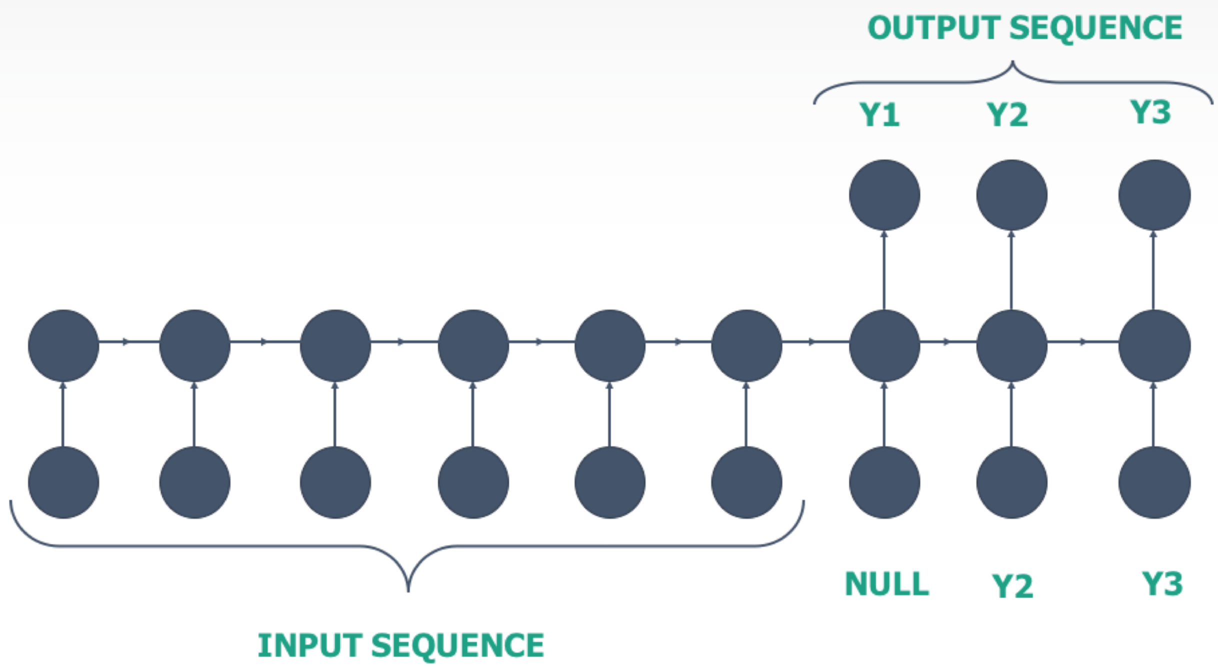
Recurrent Neural Network

- Created specifically for time series
- Model flow
  - Input to output
  - Previous cell passes encoding to next time step
  - Chooses information to retain

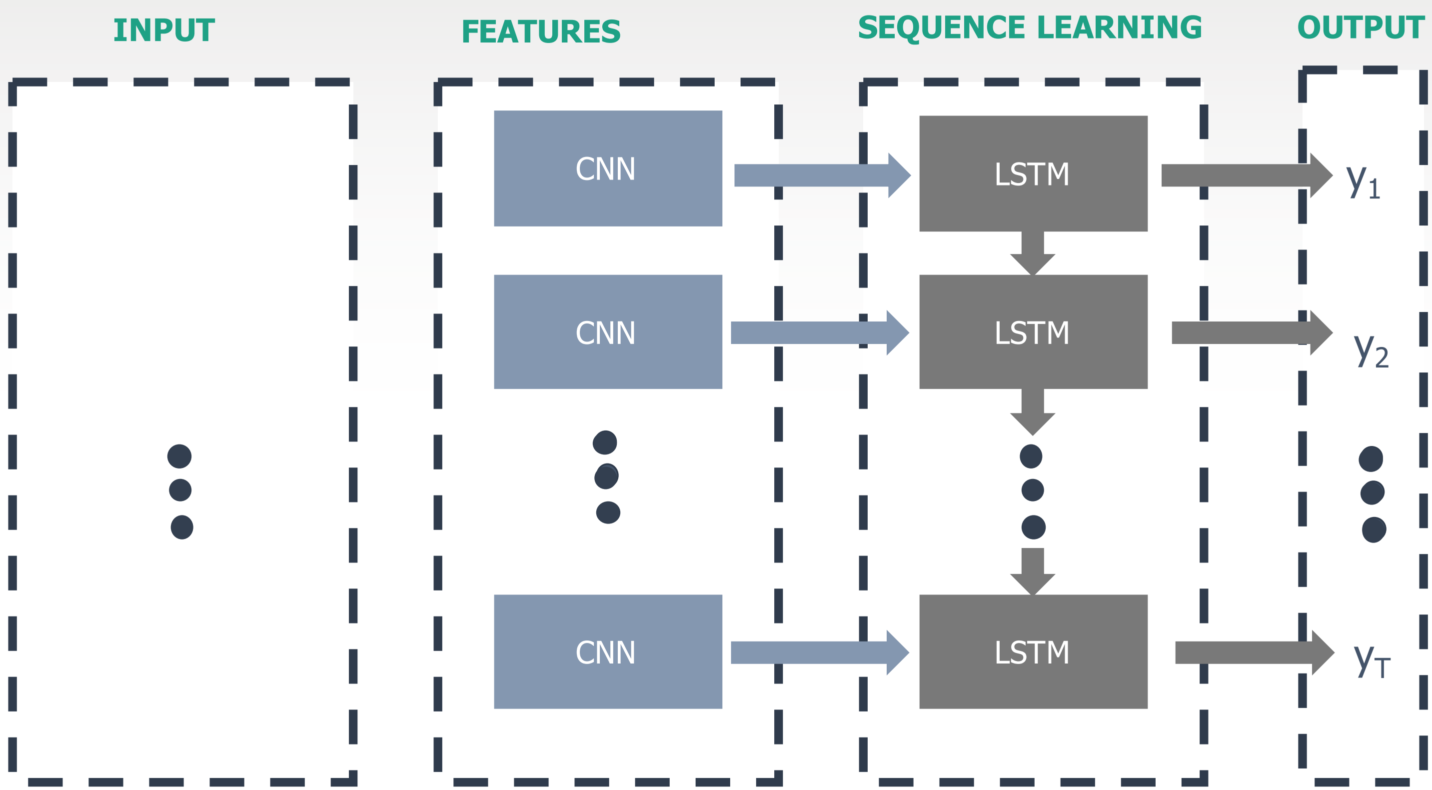


Sequence to Sequence Network

- Variable size input and output
  - Output prediction length flexible
- Two parts
  - Encoder
  - Decoder



Combined Model



Adaptive Prediction Length

- Function that captures 'reliability' of prediction
  - Max probability value $g_t = \max_i y_{i,t}$
  - Entropy $g_{i,t} = \sum_i y_{it} \log y_{it}$
  - Total Variation $g_t = \max_i |y_{i,t} - y_{i,t-1}|$
  - Wasserstein

$$\min_{w \geq 0} \sum_j w_j (y_{jt} - y_{jt-1})$$

$$|w_p - w_q| \leq |p - q| \; \forall p, q$$

Loss Functions

- Contains two parts
  - Accuracy loss (log loss)
  - Time prediction penalty
- Balance loss to
  - Increase F1 performance
  - Predict as far in time as possible

$$\sum_{g_{t(X,\theta)} \geq \tau} KL(Y_t || p_t(X,\theta)) + \lambda \sum_{t} \max(\tau - g_t(X,\theta), 0)$$

Improving Robustness

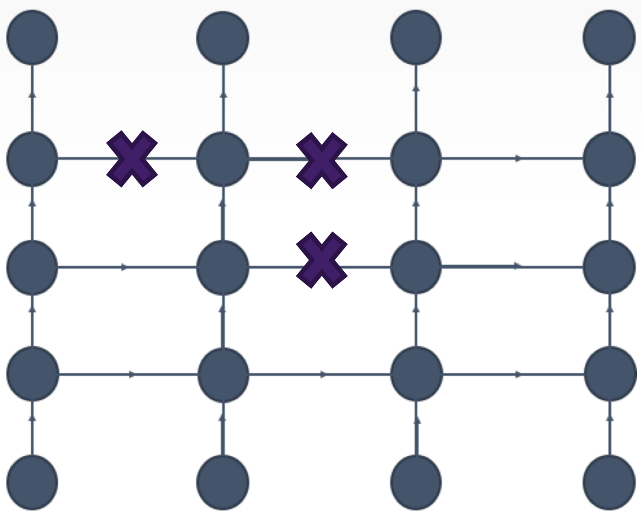
- Normal prior on weights
- Standard
  - Compute expected value of each weight
- Idea
  - Compute expected value
  - Compute standard deviation for each weight

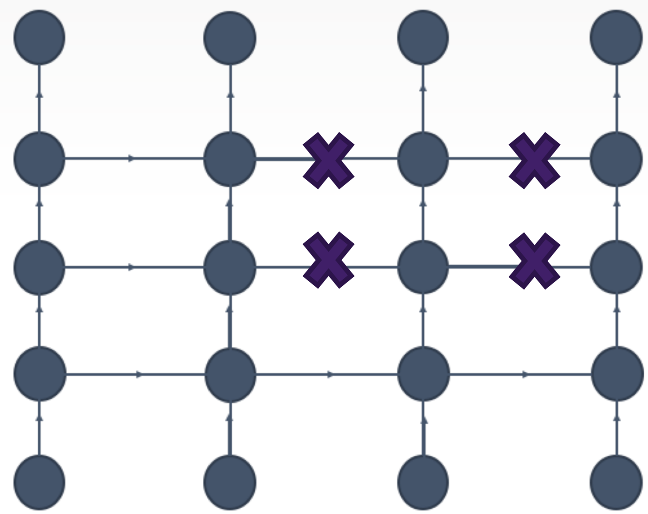
$$\min_{\overline{w},\overline{\sigma}} E_{X,Y} E_{z\sim N(\overline{w},\overline{\sigma})} KL(Y||p(X,z))$$

$$\min_{\overline{w},\overline{\sigma}} E_{X,Y} E_{z\sim N(0,1)} KL(Y||p(X,\overline{w}+z\overline{\sigma}))$$

Adaptive Computational Time

- Alex Graves (2017): Adaptive Computation Time for Recurrent Neural Networks



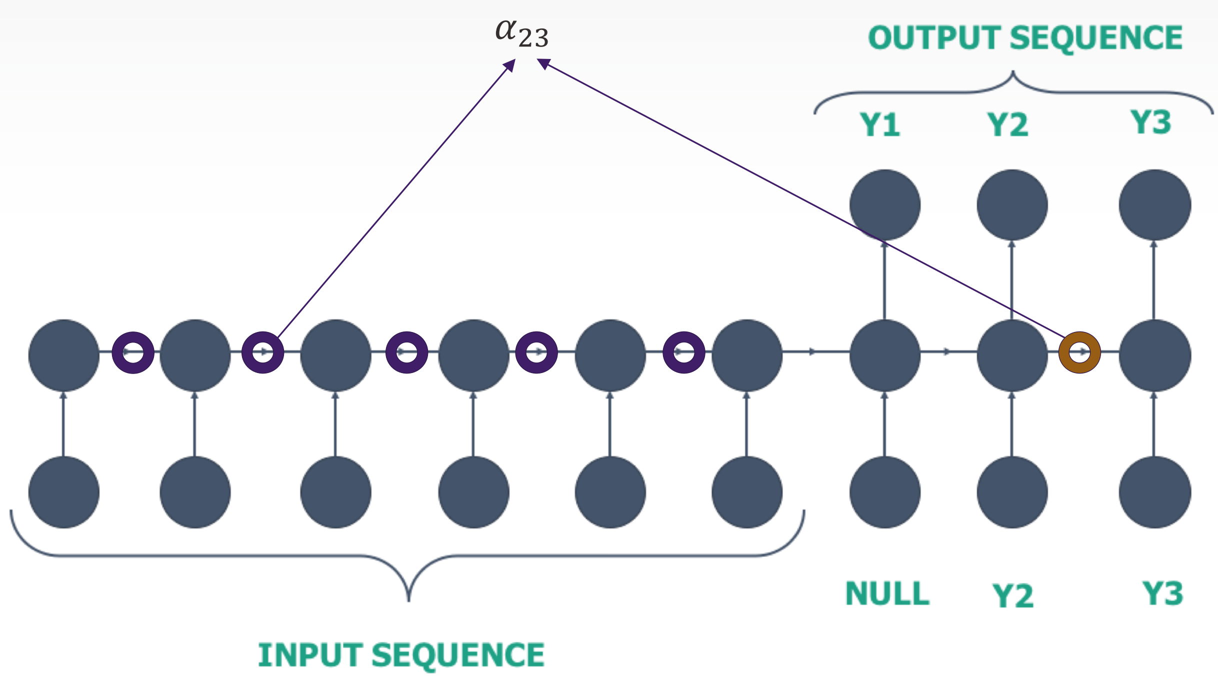


ACT

- Developed for RNN
  - Extended to sequence-to-sequence
- Embed attention
  - In decoder add

$$att_t = \sum_i \alpha_{it} hidden_i^{enc}$$

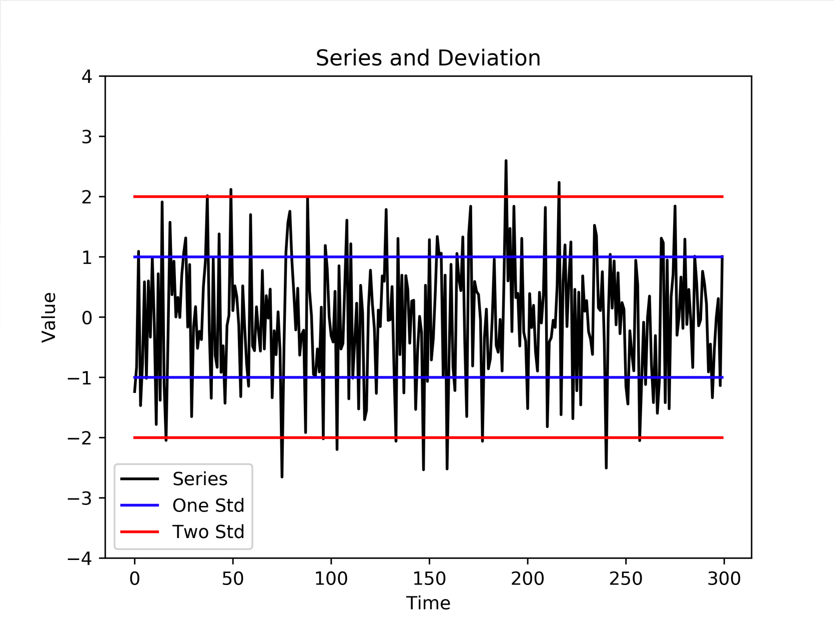
$$\alpha_{it} = f(hidden_{t-1}^{dec}, hidden_i^{enc})$$



Evaluation

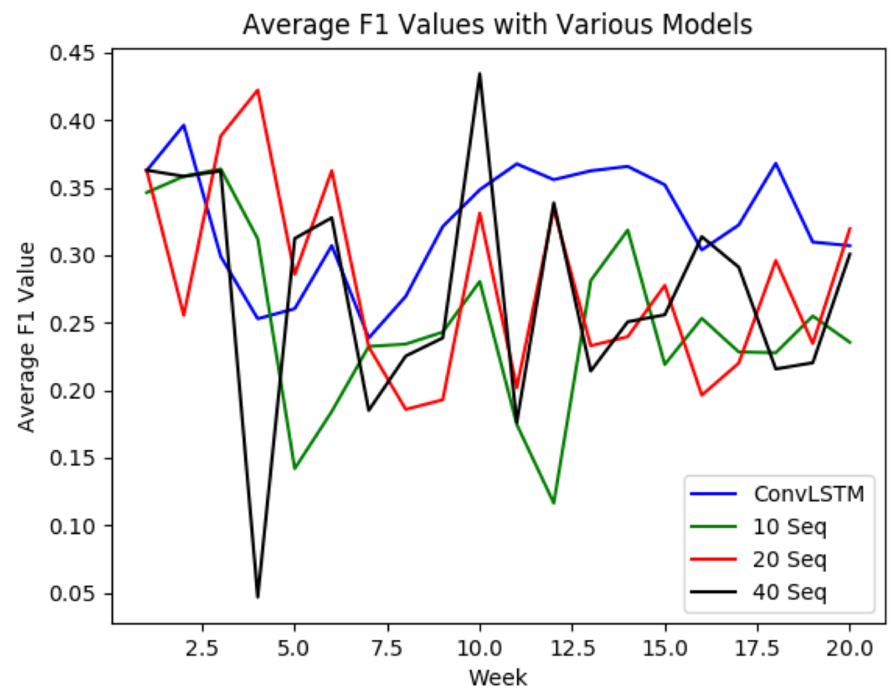
Data

- Two datasets
  - Commodity securities
  - ETF's
- Data contains
  - 5 minute tick data
  - Roughly 14 years
  - No volume information
- Data imbalanced
  - Few samples outside of two standard deviations



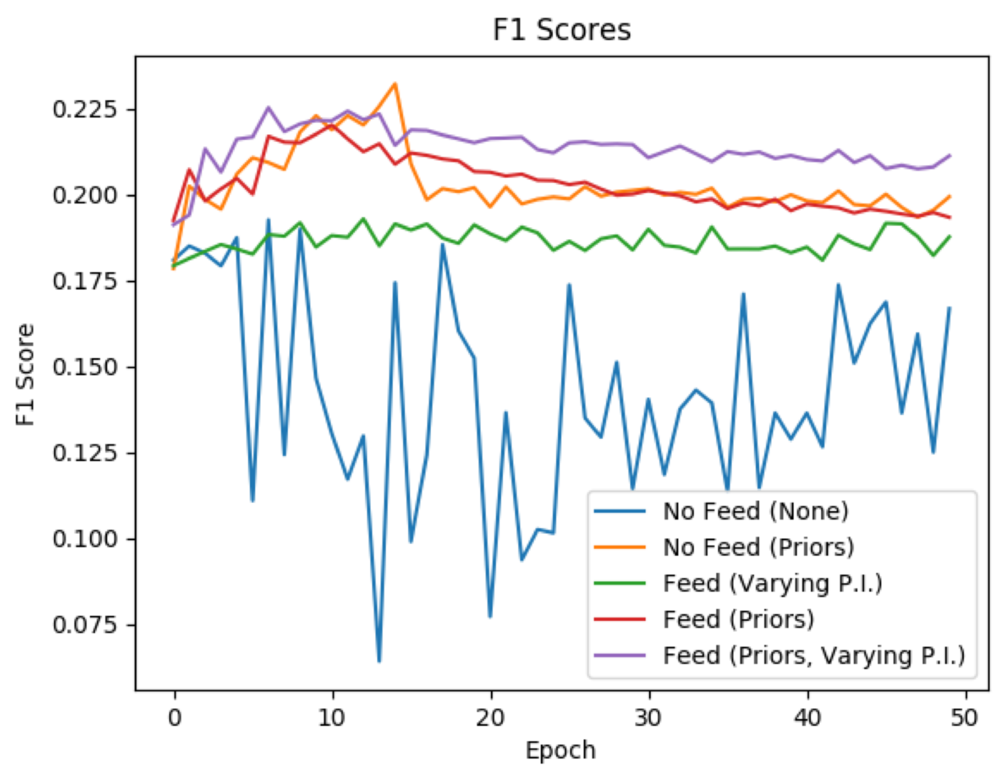
Performance

Model	Average F1 Value
ConvLSTM	0.324
10 ticks	0.250
20 ticks	0.279
40 ticks	0.271



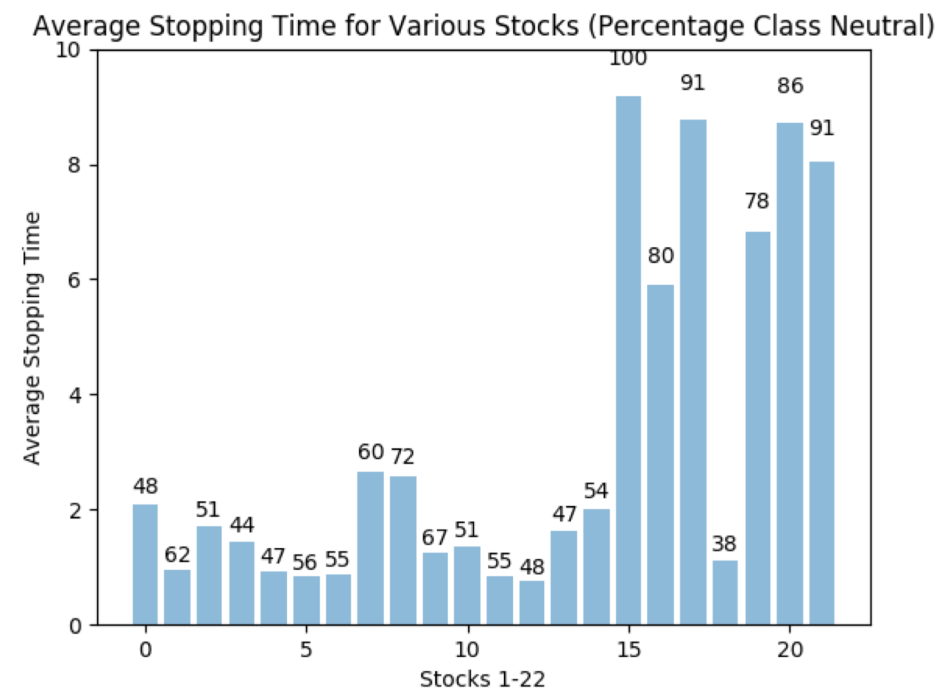
After first few weeks of training convolutional LSTM works much better.

Enhancements



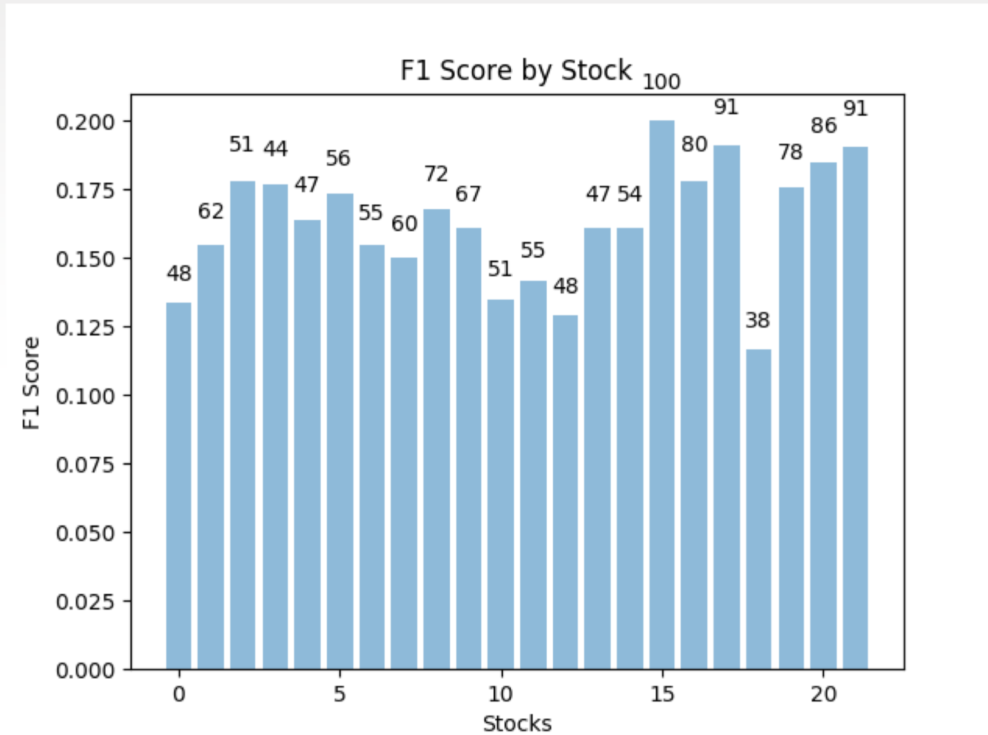
Prediction Reliability

- Maximum prediction probability above 0.35
  - Challenge how to set up training set
  - Specific loss function
- Stock price does not change
  - Reliable 'longer term predictions'

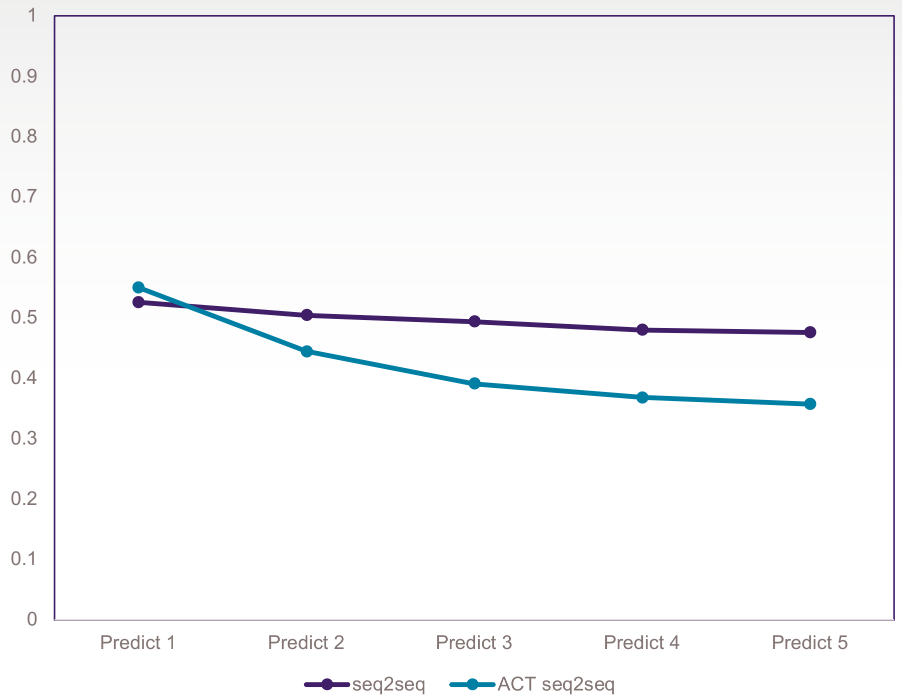


F-score by Stock

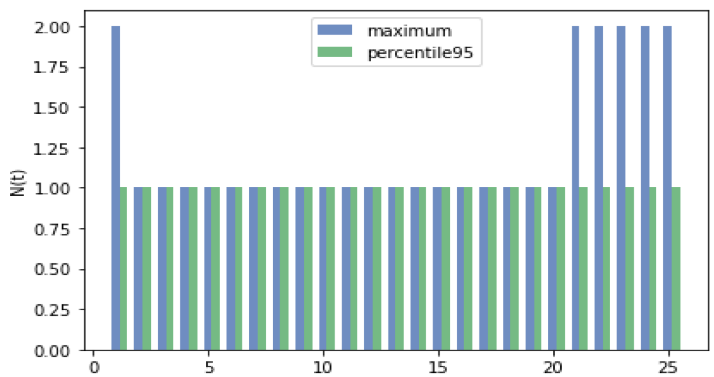
- 'Easier' to predict stocks with no large swings



ACT



	Predict 1
Seq2seq	0.53
ACT	0.25
Seq2seq + ACT	0.54



Conclusions on Deep Networks

- Pros
  - Works for complex models
  - No need for feature selection
  - Improved accuracy
- Cons
  - Long time to train
  - Hyperparameter nightmare
  - Expertise
    - Models
    - Implementation
    - Tricks of trade

- Convolution with sequence to sequence works best
- Challenge with unbalanced data

Ongoing Work

- Further enhancement to ACT
  - Use for varying prediction sequences
- Drift detection
  - Autoencoders
  - ACT
- Deep reinforcement learning
  - Place orders
  - Or perhaps contextual bandit?

Thank you very much!

@dklabjan
d-klabjan@northwestern.edu

Northwestern | McCORMICK SCHOOL OF ENGINEERING