Gated Recurrent Unit Hierarchical Architecture for Fundamental Stock Analysis and Forecast

Gabriel Adriano de Melo
Paulo Marcelo Tasinaffo

Summary

- Predicting stock prices using fundamental analysis;
- Data from more than two hundred companies over a twenty year period: trimester frequency;
- Using deep neural network architecture: Gated Recurrent Unit.

Common pitfalls in Forecasting

- Lack of a validation set: the test set evaluation becomes biased;
- Non-stationary data analysis using techniques that presume stationarity;
- Lack of a baseline estimator or using metrics that results in misleading conclusions.

... and the Cross-Section of Expected Returns

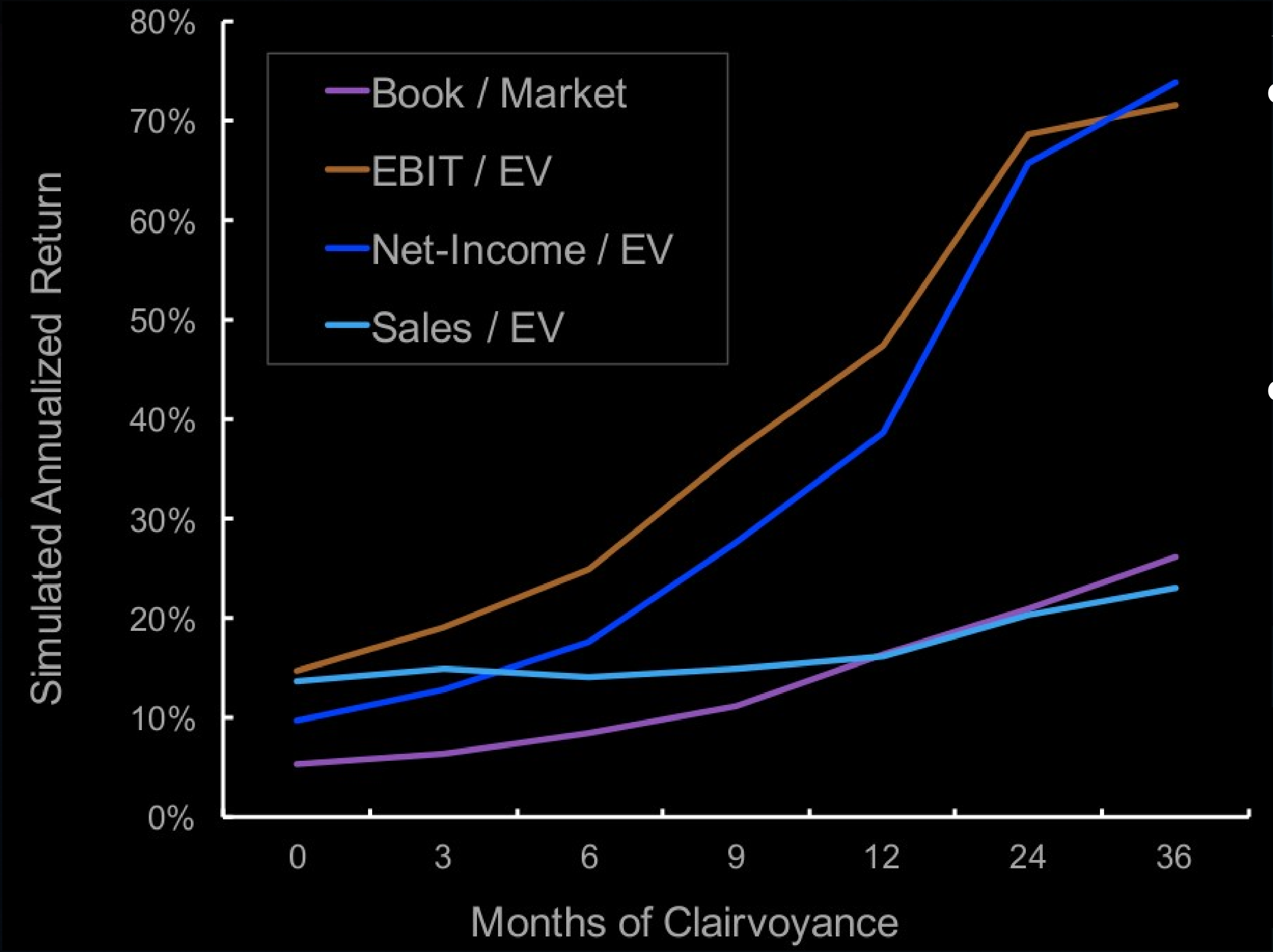
"Hundreds of papers and factors attempt to explain the cross-section of expected returns. Given this extensive data mining, it does not make sense to use the usual criteria for establishing significance. Which hurdle should be used for current research? Our paper introduces a new multiple testing framework and provides historical cutoffs from the first empirical tests in 1967 to today. A new factor needs to clear a much higher hurdle, with a t-statistic greater than 3.0. We argue that most claimed research findings in financial economics are likely false."

C. R. Harvey, Y. Liu, and H. Zhu, ". . . and the cross-section of expected returns," The Review of Financial Studies, vol. 29, no. 1, 2016.

Fundamental Analysis

- Different from technical analysis: not only the price is take in consideration;
- Financial and economic factors are examined;
- Estimates an intrinsic value which is compared to the actual price in order to issue a buy or sell recommendation.

Clairvoyance test



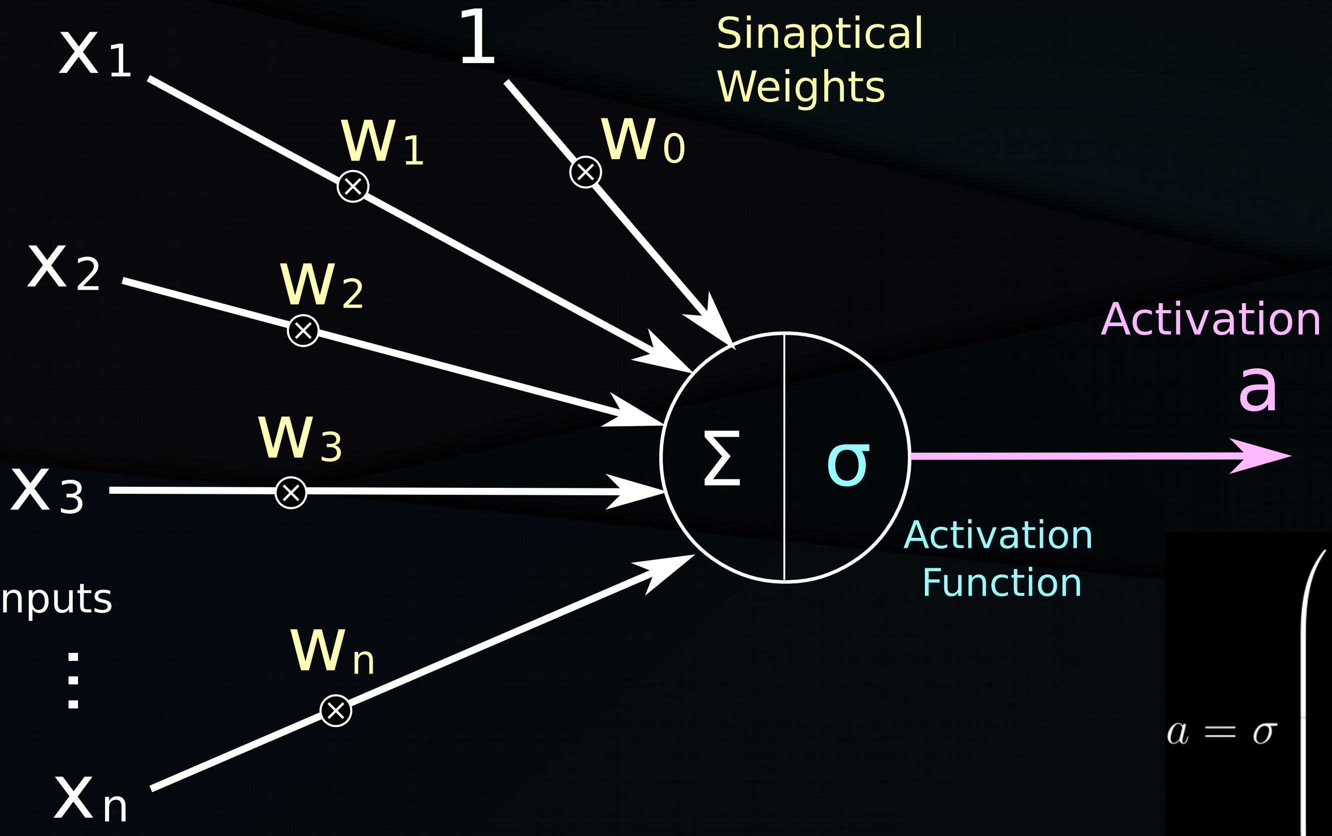
- Select stocks based on future fundamentals;
- Demonstrates the importance of financial data.

J. Alberg and Z. C. Lipton, “Improving factor-based quantitative investing by forecasting company fundamentals,” in 31st Conference on Neural Information Processing Systems (NIPS), 2017.

Deep Neural Networks

- Powerful technique widely used for language modelling, image and sound processing, time series analysis and other applications;
- Based on stacking more layers of neural networks and using clever strategies for training: layerwise greed optimization, unsupervised pre-training, ReLU activation functions.

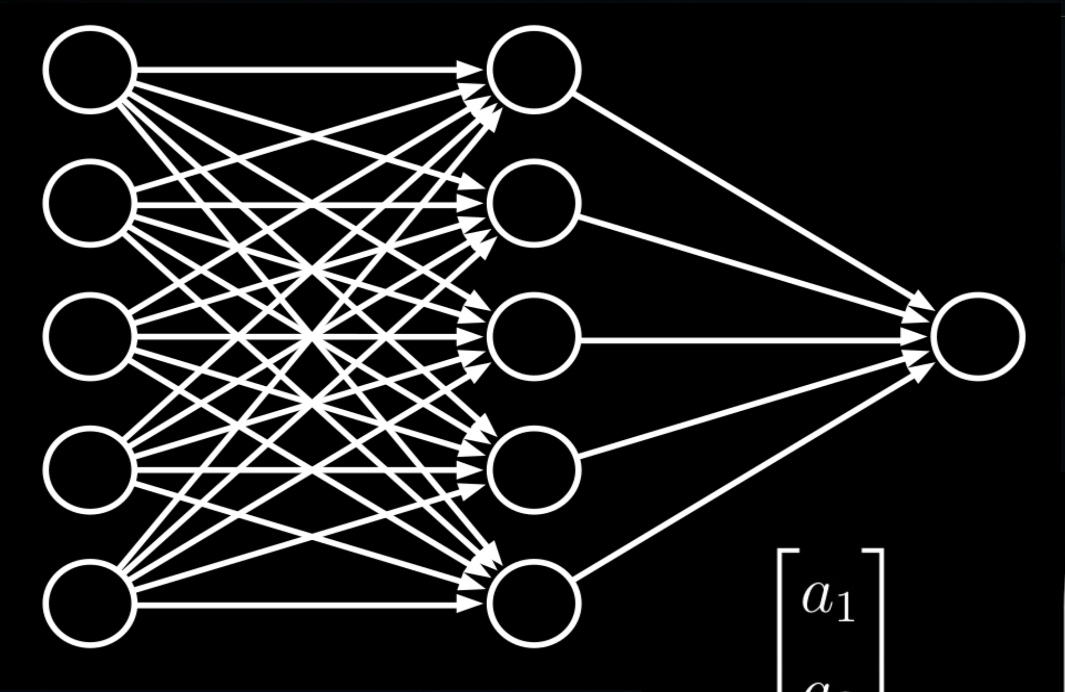
A primer on neural networks



$$a = \sigma(W\boldsymbol{x})$$

$$a = \sigma\left(\begin{bmatrix} w_0 & w_1 & w_2 & w_3 & \dots & w_n \end{bmatrix}\begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right)$$

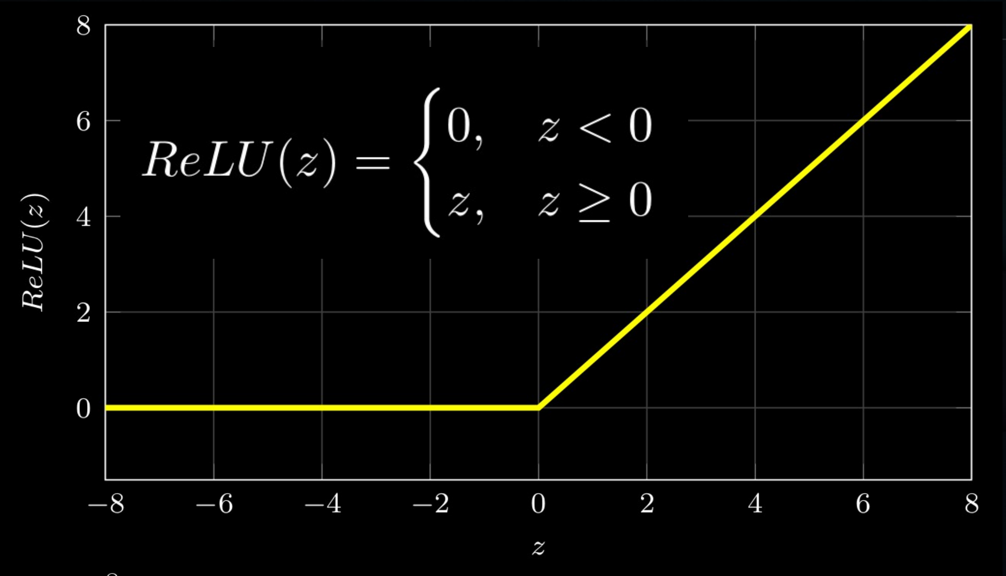
A primer on neural networks (2)



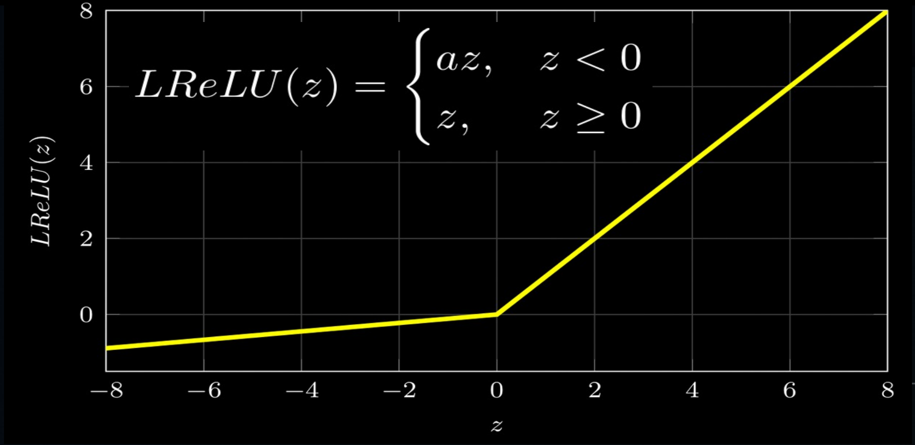
$$A = \sigma(WX)$$

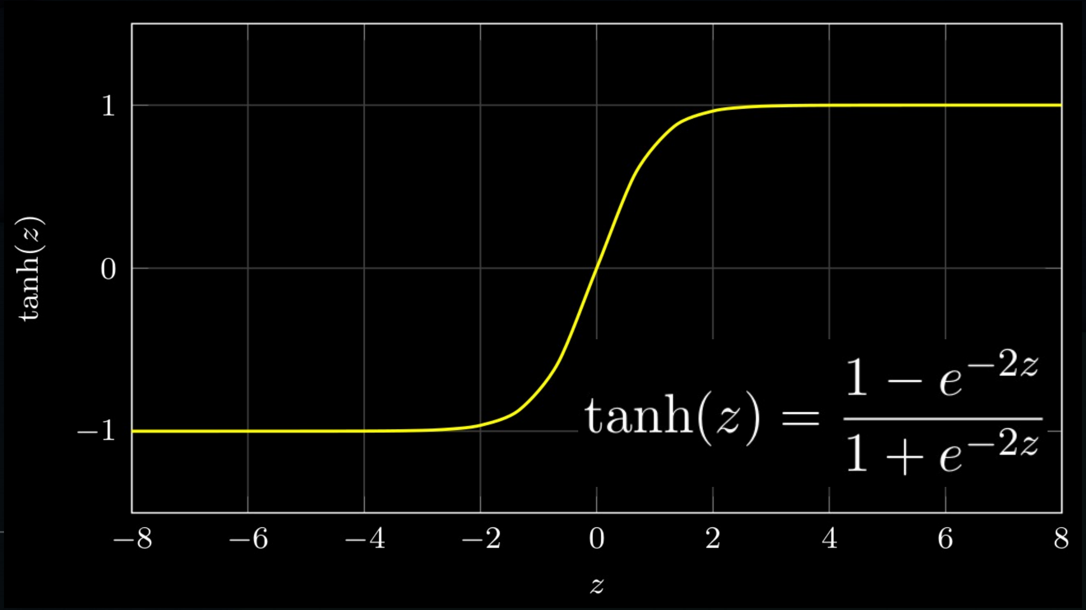
$$A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{bmatrix} = \sigma\left(\begin{bmatrix} w_{10} & w_{11} & w_{12} & \dots & w_{1n} \\ w_{20} & w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{d0} & w_{d1} & w_{d2} & \dots & w_{dn} \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right)$$

Activation Functions



The Rectified Linear Unit (ReLU), the Leaky Rectified Linear Unit (LReLU) and the Hyperbolic Tangent activation functions were used in the proposed neural network.





Backpropagation

- The gradient of the cost function is computed in order to update the weights.

$$\frac{\partial C}{\partial w_{ij}^{(L)}} = \frac{\partial C}{\partial z^{(L)}} \frac{\partial z^{(L)}}{\partial w_{ij}^{(L)}} = \frac{\partial C}{\partial a^{(L)}} \frac{da^{(L)}}{dz^{(L)}} \frac{\partial z^{(L)}}{\partial w_{ij}^{(L)}}$$

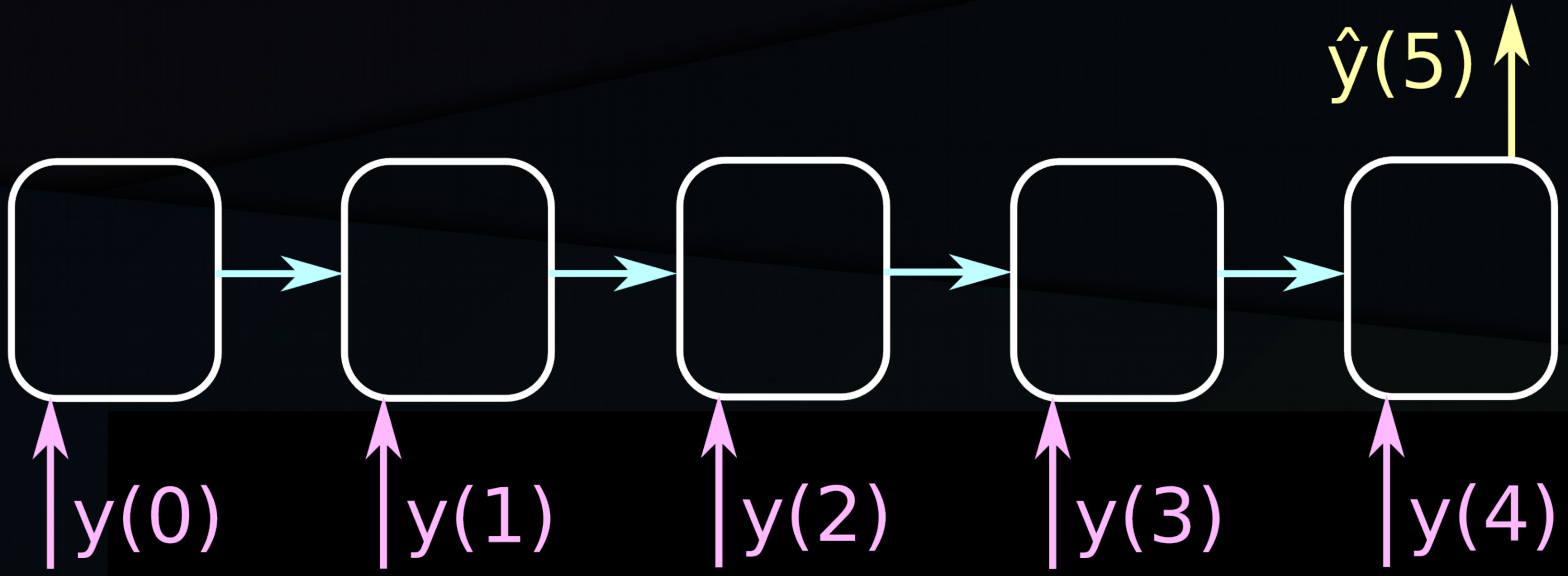
$$\frac{\partial C}{\partial z^{(L-1)}} = \frac{\partial C}{\partial z^{(L)}} \frac{\partial z^{(L)}}{\partial A^{(L-1)}} \frac{dA^{(L-1)}}{dZ^{(L-1)}}$$

Recurrent Neural Networks

- The neurons uses its past activation as an additional input: the network's output is also a function of its past inputs;
- It can be interpreted as a very deep neural network with weight sharing among the time dimension;
- Suitable for sequential input: time series of input features.

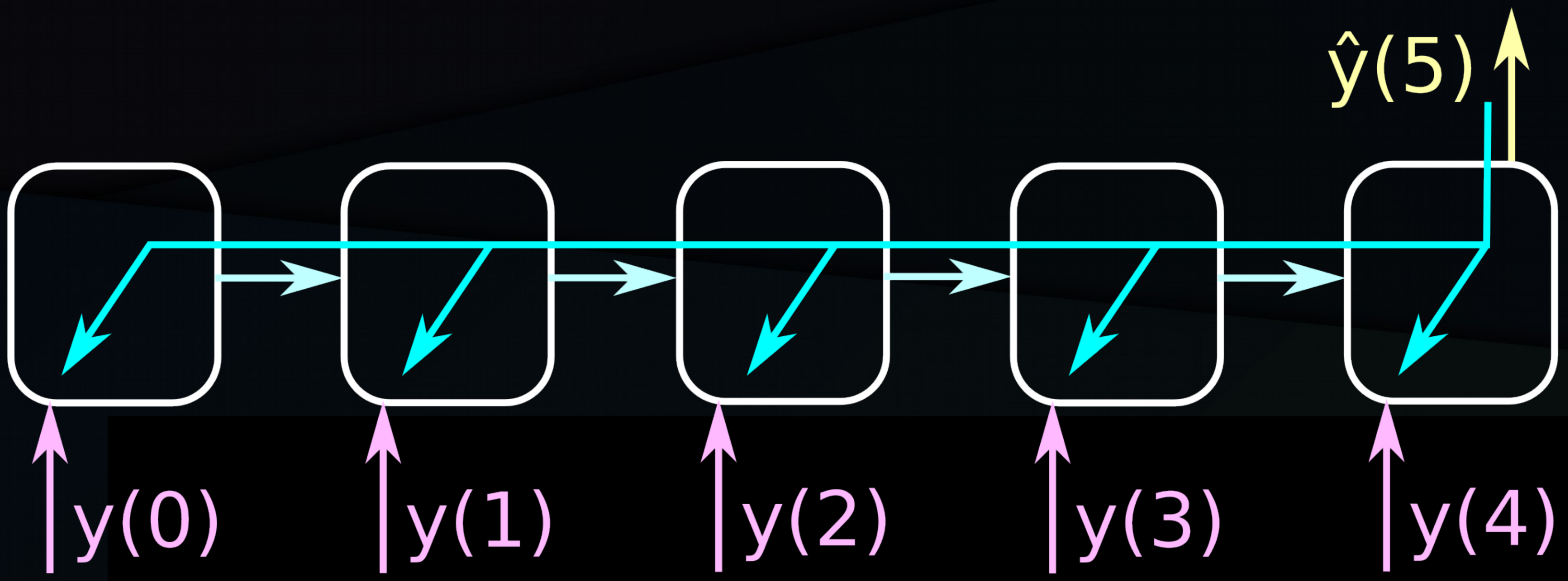
Backpropagation in RNN

- Unfold the RNN as a feedforward in time with weigh sharing.

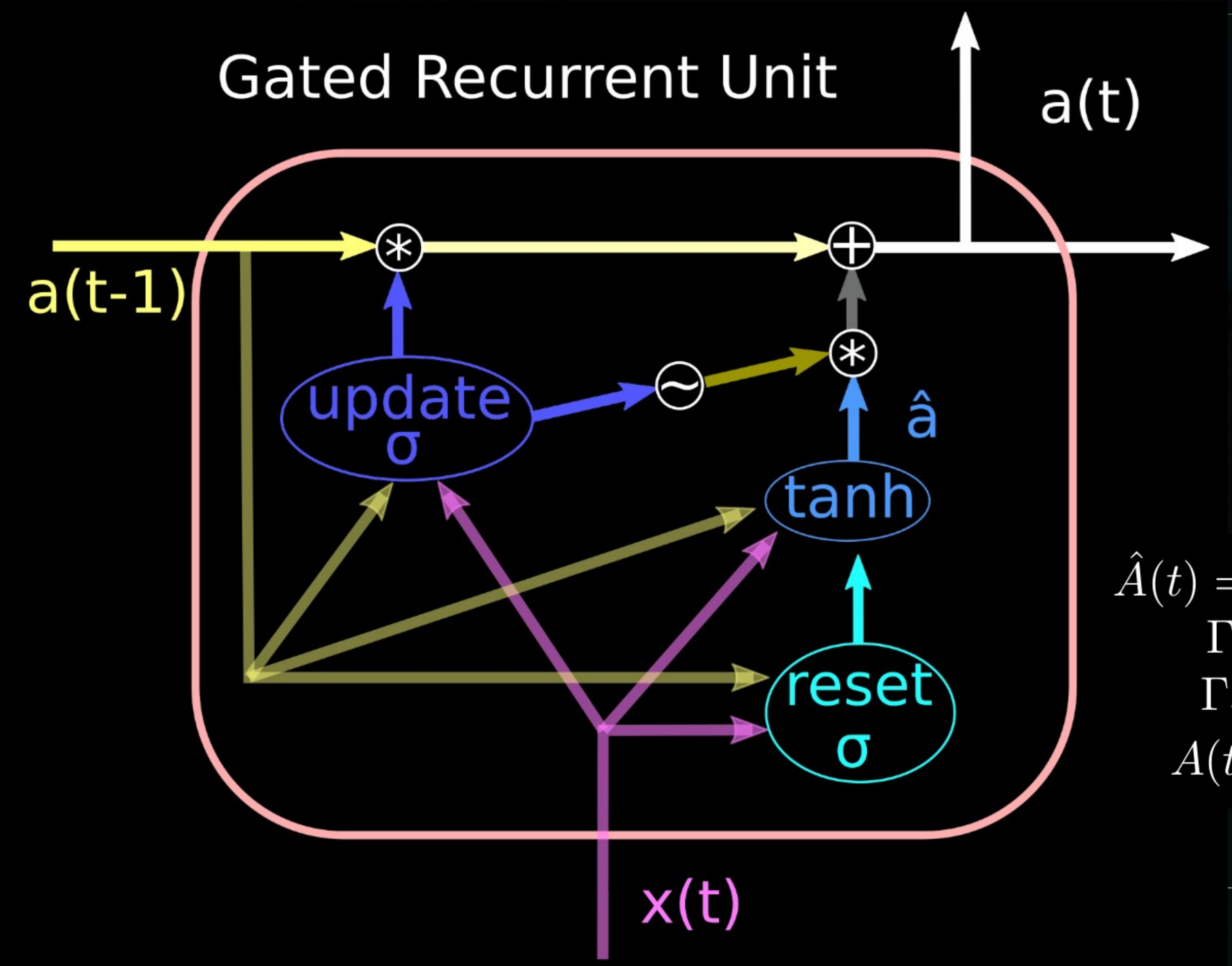


Backpropagation in RNN

- Unfold the RNN as a feedforward in time with weigh sharing.



Gated Recurrent Units (GRU)

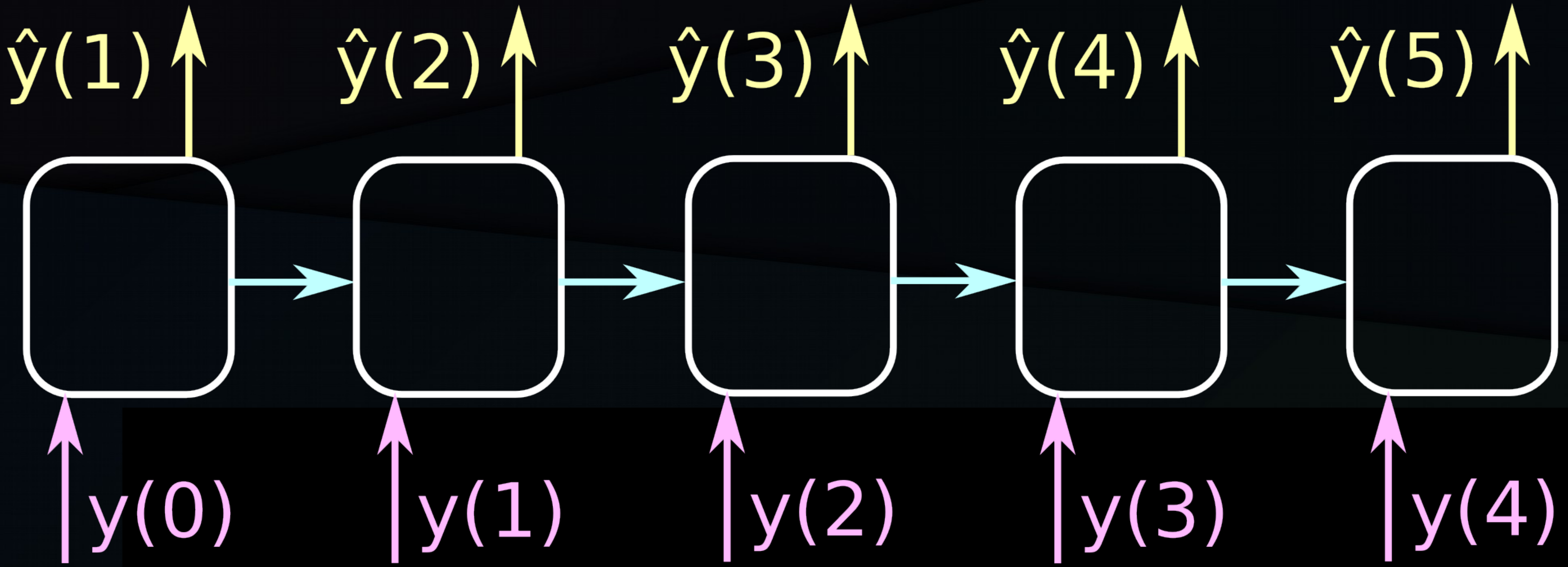


A modern approach to Long-Short Term Memory (LSTM) from 2014 with fewer trainable parameters and similar performance.

$$\hat{A}(t) = \tanh(W_{aa}(\Gamma_r * A(t-1)) + W_{ax}x(t))$$
$$\Gamma_r = \sigma(W_{ra}\Gamma_r A(t-1) + W_{rx}x(t))$$
$$\Gamma_u = \sigma(W_{ua}\Gamma_r A(t-1) + W_{ux}x(t))$$
$$A(t) = \Gamma_u * \hat{A}(t) + (1 - \Gamma_u) * A(t-1)$$

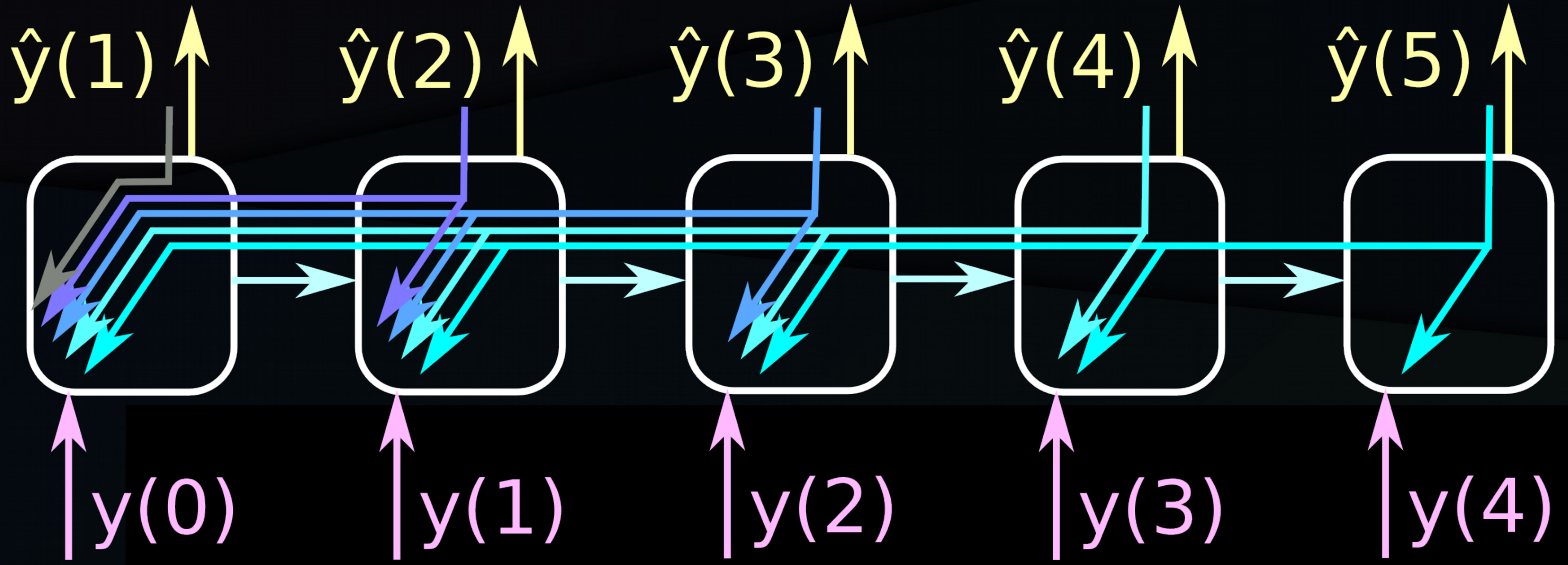
Sequence to Sequence training

- More propagated gradients acts have a regularizing effect.

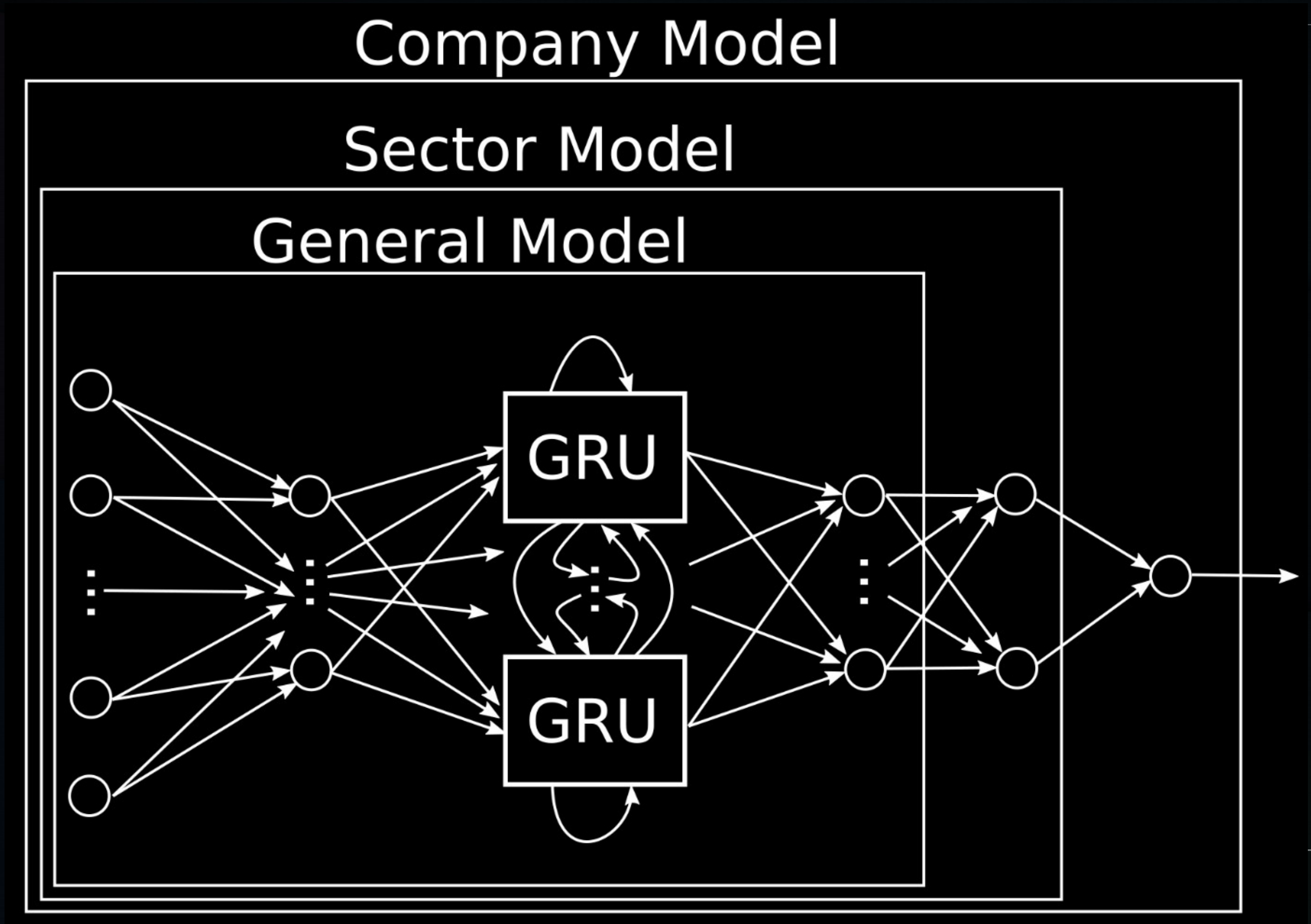


Sequence to Sequence training

- More propagated gradients acts have a regularizing effect.



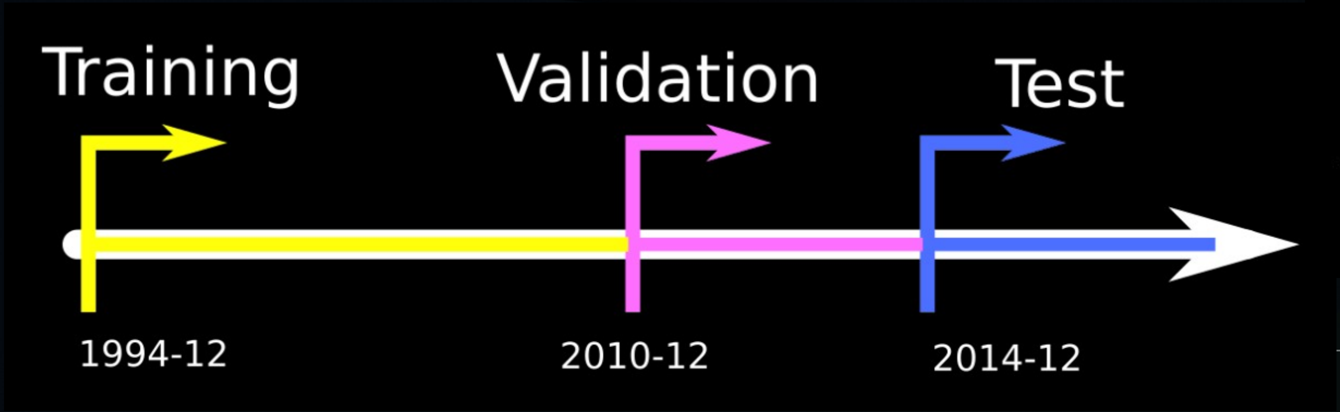
Proposed Model



- Hierarchical architecture based on the company sector;
- Weight sharing.

Data

- Sampled from every trimester, 209 US companies obtained from Stockpup.com;
- Normalized to have unitary standard deviation;
- Non-stationary data was differentiated.



$$R(t) = \frac{V(t)}{V(t-1) + \epsilon}$$

Companies: 209

AA, AAPL, ABT, ADBE, ADI, ADM, ADP, ADSK, AEP, AJG, ALK, AMAT, AMD, AMGN, AOS, APA, APC, APD, ARW, AVP, AVY, BA, BAX, BBY, BC, BCR, BDX, BEN, BF.B, BIG, BMS, BMY, CA, CAG, CAH, CAT, CBS, CCL, CDNS, CERN, CHD, CL, CLF, CLX, CMI, CMS, COG, COO, CPB, CR, CSCO, CSX, CTAS, CTL, CVS, CVX, D, DAL, DCI, DHR, DOV, DOW, ECL, EFX, EIX, EMR, EOG, EQT, ETN, ETR, F, FAST, FCX, FISV, FLS, FMC, GCI, GD, GIS, GLW, GPC, GPS, GT, GWW, HAL, HAS, HD, HES, HOG, HON, HP, HPQ, HRB, HRL, HRS, HSY, HUM, IBM, IFF, INTC, IP, IPG, ITW, JBHT, JCI, JEC, JNJ, JWN, K, KLAC, KMB, KO, KR, KSU, LEG, LH, LLY, LM, LMT, LNT, LUK, LUV, M, MAS, MAT, MCD, MDP, MDT, MGM, MKC, MMC, MMM, MO, MRK, MSFT, MSI, MU, MUR, MYL, NBL, NKE, NSC, NUE, NWL, NYT, ODP, OI, OMC, ORCL, OXY, PAYX, PBI, PEP, PFE, PG, PH, PKI, PNR, PNW, PPG, PVH, R, RAD, RDC, RHI, ROST, S, SCG, SHW, SJM, SNA, SO, SSP, STZ, SVU, SWK, SWKS, SWN, SYK, SYMC, SYY, T, TAP, TGT, TIF, TJX, TMO, TSN, TXN, TXT, UAL, UHS, UIS, UNH, UNP, UTX, VAR, VFC, VZ, WDC, WEC, WEN, WHR, WMB, WMT, WY, XEL, XOM, XRX

Sectors

1 Personal, 2 Brokers, 7 Chemicals, 5 Aerospace, 1 Engineering, 1 Employment, 7 Communication, 13 Medical, 6 Drug, 2 Asset, 8 Semiconductors, 16 Oil, 1 Biotechnology, 5 Travel, 3 Packaging, 2 Building, 1 Forest, 1 Tobacco, 2 Entertainment, 1 Farm, 2 Advertising, 5 Beverages, 2 Restaurants, 1 Consulting, 5 Manufacturing, 1 Steel, 3 Publishing, 5 Transportation, 3 Metals, 5 Computer, 17 Retail, 3 Autos, 11 Utilities, 4 Health, 19 Industrial, 11 Application, 4 Airlines, 18 Consumer, 5 Business

Stationary Data

EPS basic, EPS diluted, Dividend per share, ROE (Return on equity), ROA (Return on assets), P/B ratio (the ratio of Price to Book value of equity per share), P/E ratio (the ratio of Price to EPS diluted TTM as of the previous quarter), Dividend payout ratio, Long-term debt to equity ratio, Equity to assets ratio, Net margin (the ratio of Earnings TTM to Revenue TTM), Asset turnover (the ratio of Revenue TTM to TTM average Assets), Free cash flow per share, Current ratio.

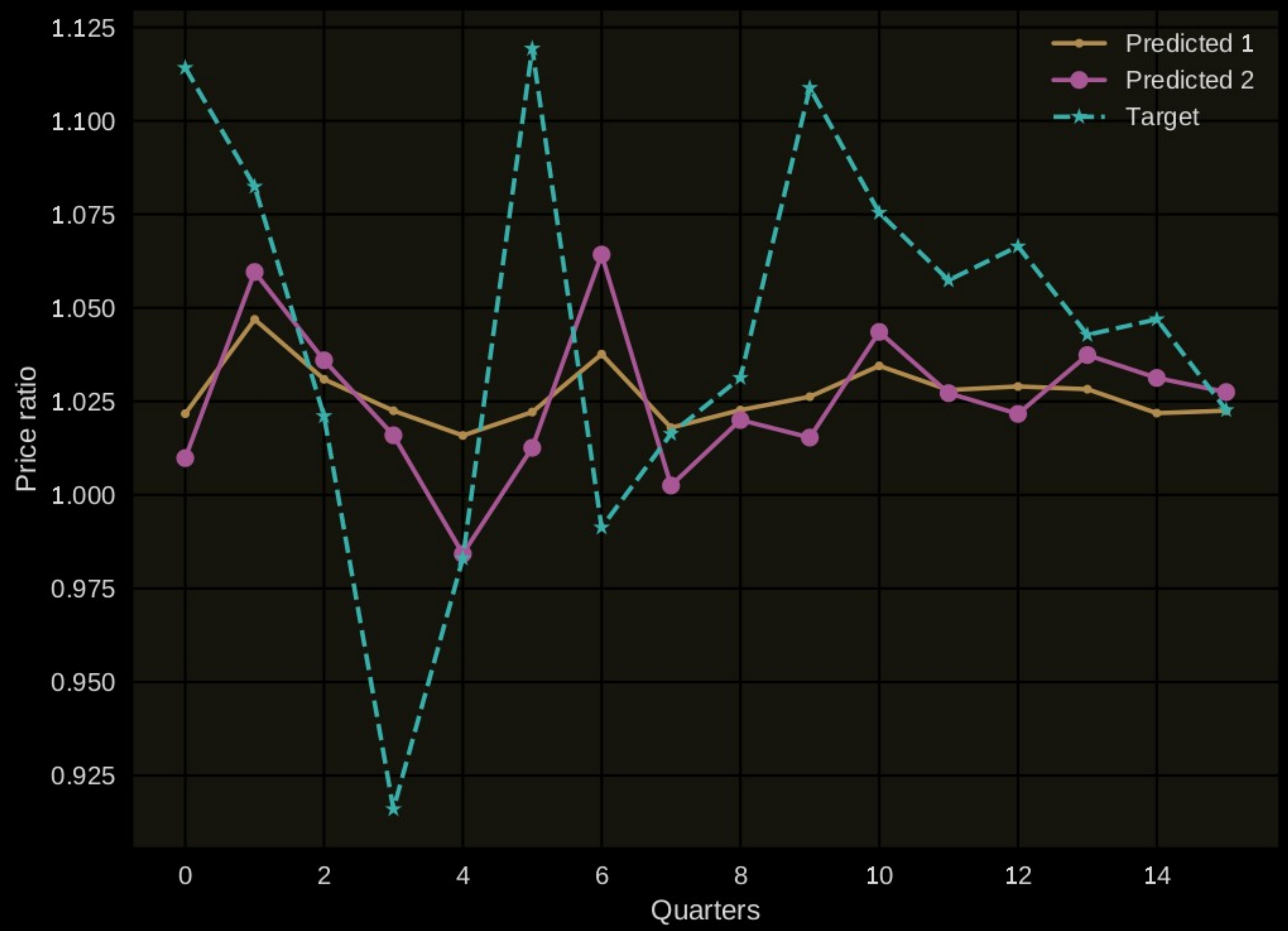
Differentiated data

Assets, Current Assets, Liabilities, Current Liabilities (at the end of a quarter), Shareholders equity (includes both common and preferred stockholders), Goodwill & intangibles, Long-term debt, Revenue, Earnings, Earnings available for common stockholders, Cash from operating activities, Cash from investing activities, Cash from financing activities, Cash change during period, Cash at end of period, Capital expenditures, Price (the medium price per share of the company common stock during a given quarter as reported, not adjusted for subsequent dividends), Book value of equity per share, Cumulative dividends per share, Non-controlling interest, Preferred equity.

Training

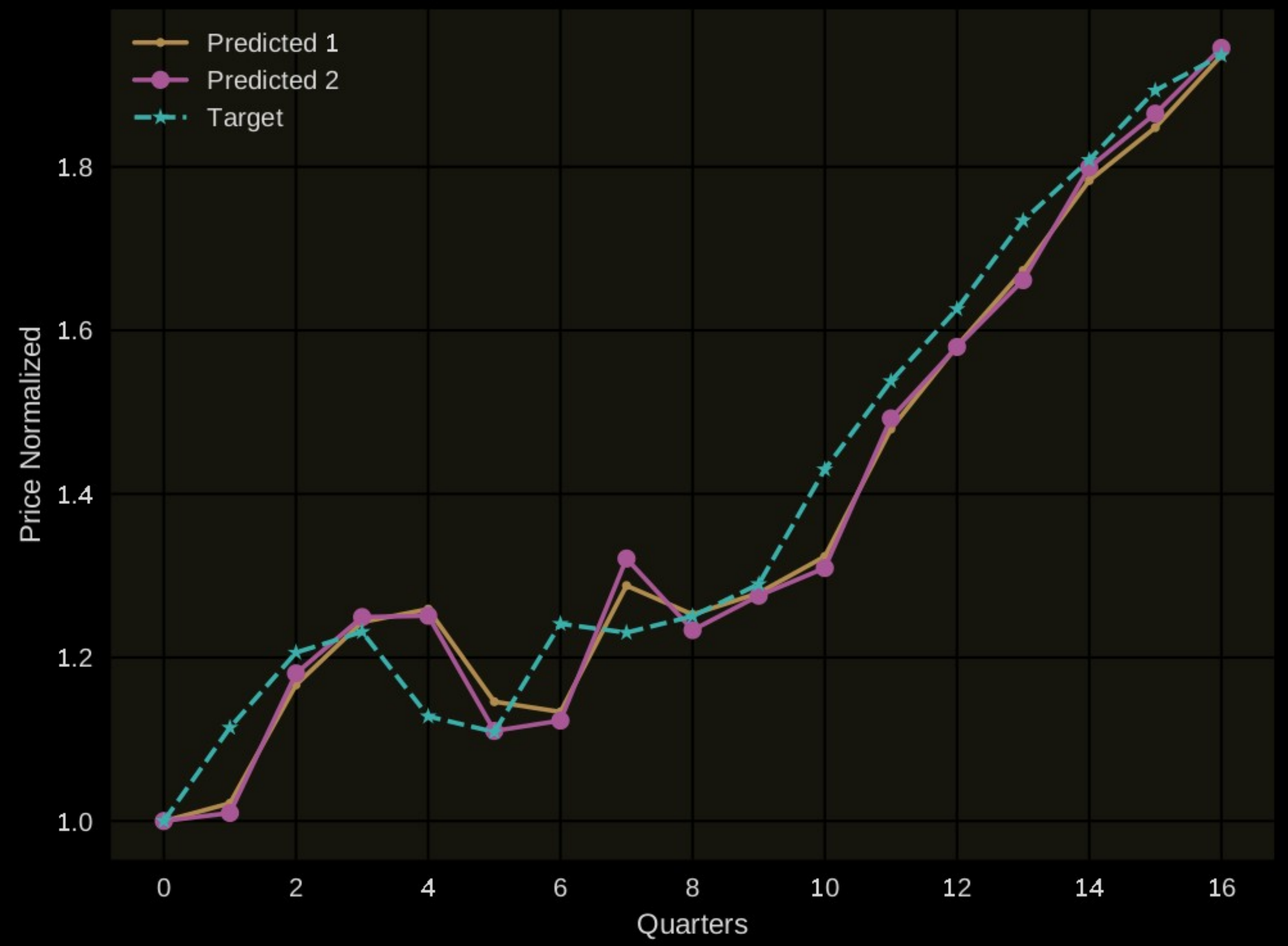
- Metric: squared error for price ratios;
- Optimizer: Adaptive moment estimation (Adam);
- Early Stopping;
- Steps: global pre-training, sector training and company fine-tuning.

Results



- Two selected models:
  - Model 1 has more dropout layers;
  - Model 2 has more free parameters

Results aggregated



- Product of the ratios gives the absolute normalized value;
- One quarter ahead prediction;
- Not impressive at all.

Conclusions

- The model has learned a persistent function: good companies tend to continue good companies;
- It didn’t outperform that baseline estimator;
- Future work should focus in other data sources such as reports, news and comments using natural language processing techniques.

References

- [1] C. R. Harvey, Y. Liu, and H. Zhu, ". . . and the cross-section of expected returns," The Review of Financial Studies, vol. 29, no. 1, 2016.
- [2] D. A. Hara, M. A. Botelho, A. Panariello, and C. H. C. Ribeiro, "Algorithmic trading using artificial intelligence tools," in Workshop of Artificial Intelligence Applied to Finance (WAIAF), 2018.
- [3] E. Jabbur, R. Oliveira, and A. Pereira, "Proposal and implementation of machine learning and deep learning models for stock markets," in Workshop of Artificial Intelligence Applied to Finance (WAIAF), 2018.
- [4] E. F. Fama, "Market efficiency, long-term returns, and behavioral finance," Journal of Financial Economics, vol. 49, pp. 283–306, 1998.
- [5] J. Zheng, C. Xu, Z. Zhang, and X. Li, "Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network," in 2017 51st Annual Conference on Information Sciences and Systems (CISS), March 2017, pp. 1–6.
- [6] Y. Song, "Stock trend prediction: Based on machine learning methods," Master's thesis, University of California Los Angeles, 2018.
- [7] E. Chong, C. Han, and F. C. Park, "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies," Expert Systems with Applications, vol. 83, pp. 187–205, 2017.
- [8] M. F. Dixon, N. G. Polson, and V. O. Sokolov, "Deep learning for spatio-temporal modeling: Dynamic traffic flows and high frequency trading," Appl Stochastic Models Bus Ind., pp. 1–20, 2018.
- [9] J. B. Heaton, N. G. Polson, and J. H. Witte, "Deep learning for finance: deep portfolios," Applied Stochastic Models in Business and Industry, vol. 33, pp. 3–12, 2016.

References (2)

- [10] Z. Xiong, X.-Y. Liu, S. Zhong, H. B. Yang, , and A. Walid, “Practical deep reinforcement learning approach for stock trading,” in NIPS 2018 Workshop on Challenges and Opportunities for AI in Financial Services: the Impact of Fairness, Explainability, Accuracy, and Privacy, Montréal, Canada., 2018.
- [11] H. M, G. E. A., V. K. Menon, and S. K. P, “Nse stock market prediction using deep-learning models,” Procedia Computer Science, vol. 132, pp. 1351–1362, 2018.
- [12] F. Zhou, H. min Zhou, Z. Yang, and L. Yang, “Emd2fnn: A strategy combining empirical mode decomposition and factorization machine based neural network for stock market trend prediction,” Expert Systems With Applications, vol. 115, pp. 136–151, 2019.
- [13] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997. [Online]. Available: http://dx.doi.org/10.1162/neco.1997.9.8.1735
- [14] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” CoRR, vol.abs/1412.3555, 2014. [Online]. Available: http://arxiv.org/abs/1412.3555
- [15] F. Chollet et al., “Keras,” https://keras.io, 2015.
- [16] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in Proceedings of the 3rd International Conference on Learning Representations (ICLR), 2015.
- [17] L. Tilden Group, “Corporate fundamental data,” http://www.stockpup.com/data/, 2018.
- [18] S. UG, “Simplifying finance: Data finder,” https://simfin.com, 2018.
- [19] J. Alberg and Z. C. Lipton, “Improving factor-based quantitative investing by forecasting company fundamentals,” in 31st Conference on Neural Information Processing Systems (NIPS), 2017.