Targeted Manipulation: Slope-Based Attacks on Financial Time Series Data

Dominik Luszczynski, Department of Computer Science, University of Toronto

Introduction

Problem: The increased use of machine learning (ML) and artificial intelligence (AI) models in high-risk sectors, such as healthcare and finance, make it imperative that these predictive models are robust. Most contributions to adversarial attack research has been focused on image and text classification, while research about attacks on time series data, especially financial data, is still in its early stages [1].

Contribution: This research aims to build upon previous studies on adversarial attacks by introducing two slope-based targeted attacks on financial time series data, aimed to alter the temporal characteristics of a model's predictions. Furthermore, Generative Adversarial Networks (GANs) were experimented with to generate adversarial examples, based on a slope-based objective.

Robustness Criteria: The attack methods were tested on the novel N-HiTS architecture. In addition, the stealthiness of the attacks were measured against a 4-layered CNN was trained on varying sets of adversarial methods.

Method

Two novel slope attacks were implemented, the General Slope attack (GSA) and the Least Squares Slope Attack (LSSA), where each varies on how they compute the slope of the forecasted time series.

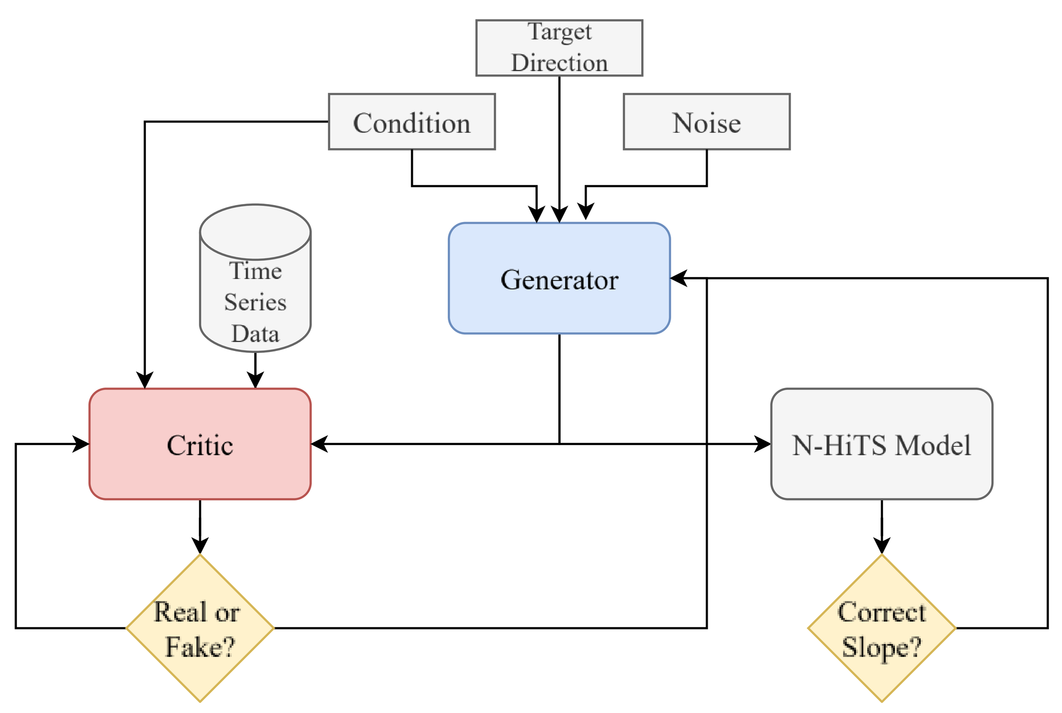
For hyperparameter scalars *c, d*, direction $t \in \{-1, 0, 1\}$, and a forecasted time series $(x_1, y_1), \ldots, (x_n, y_n)$, we compute the slopes *m* and loss function as follows:

$$\text{GSA} = \frac{y_n - y_1}{x_n - x_1}$$

$$LSSA = \frac{\sum_{i=0}^{N} (x - \bar{x})(y - \bar{y})}{(x - \bar{x})^2}$$

$$loss(m) = \begin{cases} ce^{-tdm}, & t \in \{-1, 1\} \\ cm^2, & t = 0 \end{cases}$$

An Adversarial GAN (A-GAN) was trained with stock data extracted similar to the N-HiTS model, specifically using the stock with the ticker *A*, where random continuous intervals of 99 days of log returns were selected for each sample of training data. A Conditional Wasserstein GAN was used for the A-GAN, conditioned with the corresponding 99 days of log returns. The condition is then concatenated with the noise vector in the feature dimension for both the generator and the critic.



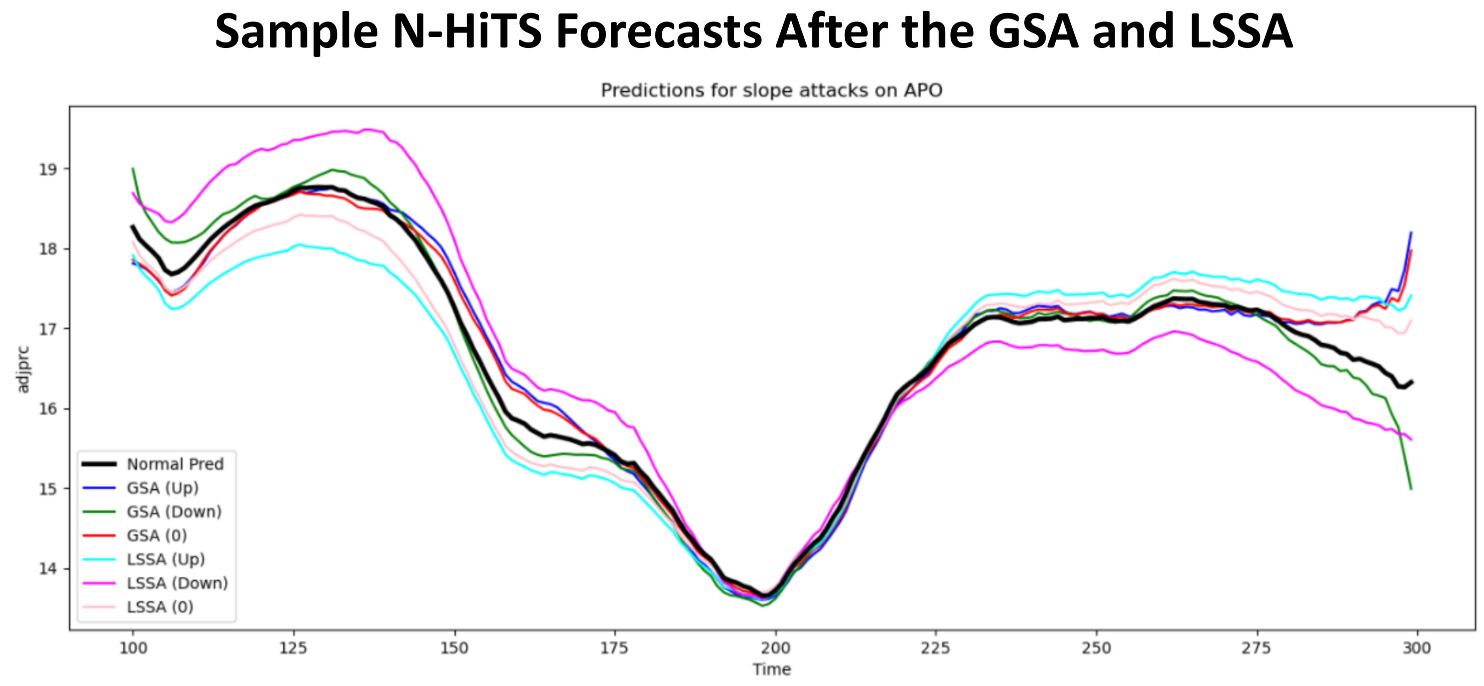
Results

GSA and LSSA Compared to Baseline Adversarial Attacks

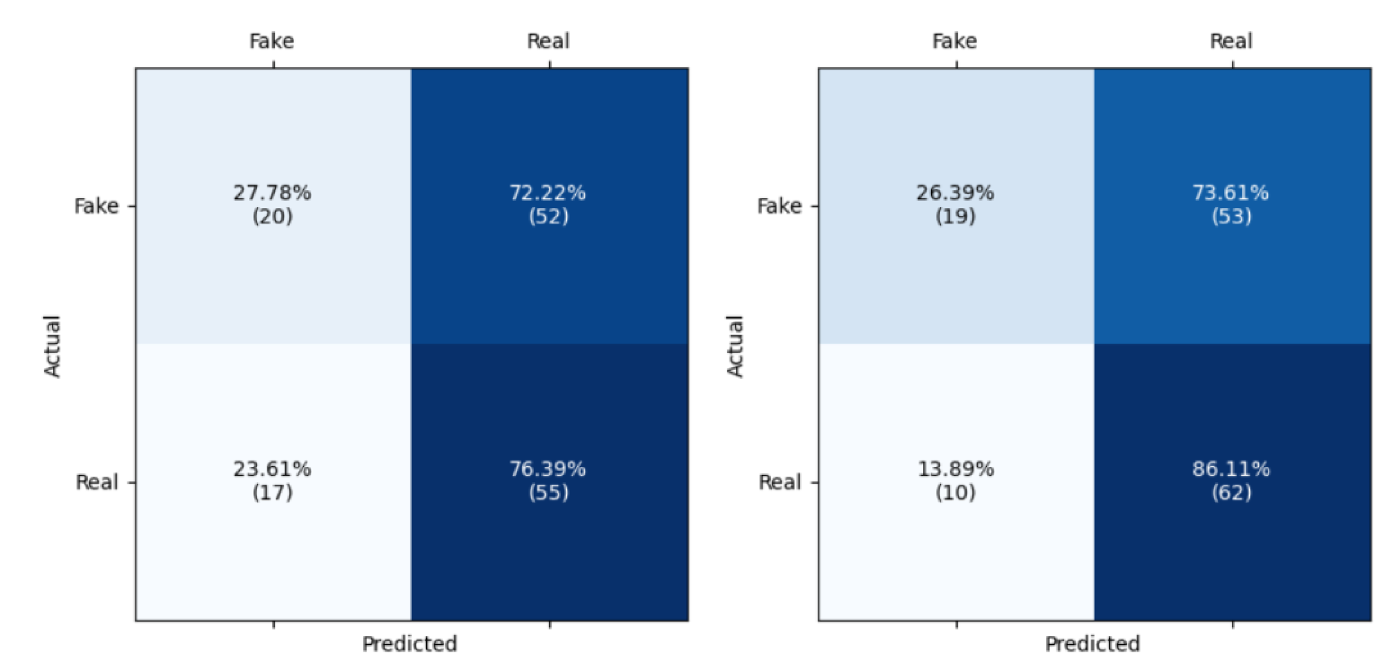
Table 1: Average metrics for different attack methods performed on the first 300 days of each recording, with $\epsilon = 2\% \cdot median(adjprc)$. The best metrics are bolded.

Attack	*MAE*	*RMSE*	*MAPE*	*Gen. Slope*	*LS Slope*
Normal	2.15	2.72	3.82×10^{-2}	3.37×10^{-2}	2.22×10^{-2}
FGSM	2.57	3.21	4.51×10^{-2}	3.22×10^{-2}	2.34×10^{-2}
BIM	**3.38**	**3.99**	$\mathbf{5.68\times10^{-2}}$	3.48×10^{-2}	2.39×10^{-2}
MI-FGSM	3.37	**3.99**	5.67×10^{-2}	3.44×10^{-2}	2.39×10^{-2}
SIM	2.57	3.08	4.29×10^{-2}	3.37×10^{-2}	2.23×10^{-2}
TIM (Up)	2.49	3.21	4.52×10^{-2}	3.72×10^{-2}	2.00×10^{-2}
TIM (Down)	2.74	3.26	4.44×10^{-2}	3.32×10^{-2}	2.51×10^{-2}
GSA (Up)	2.26	2.88	4.03×10^{-2}	$\mathbf{6.76\times10^{-2}}$	2.77×10^{-2}
GSA (Down)	2.23	2.83	3.89×10^{-2}	$\mathbf{-1.68\times10^{-4}}$	1.75×10^{-2}
GSA (0)	2.30	2.93	4.01×10^{-2}	1.80×10^{-2}	2.00×10^{-2}
LSSA (Up)	2.49	3.10	4.26×10^{-2}	5.38×10^{-2}	$\mathbf{4.96\times10^{-2}}$
LSSA (Down)	2.71	3.33	4.63×10^{-2}	1.56×10^{-2}	$\mathbf{-5.04\times10^{-3}}$
LSSA (0)	2.68	3.31	4.55×10^{-2}	2.82×10^{-2}	1.29×10^{-2}

Sample N-HiTS Forecasts After the GSA and LSSA

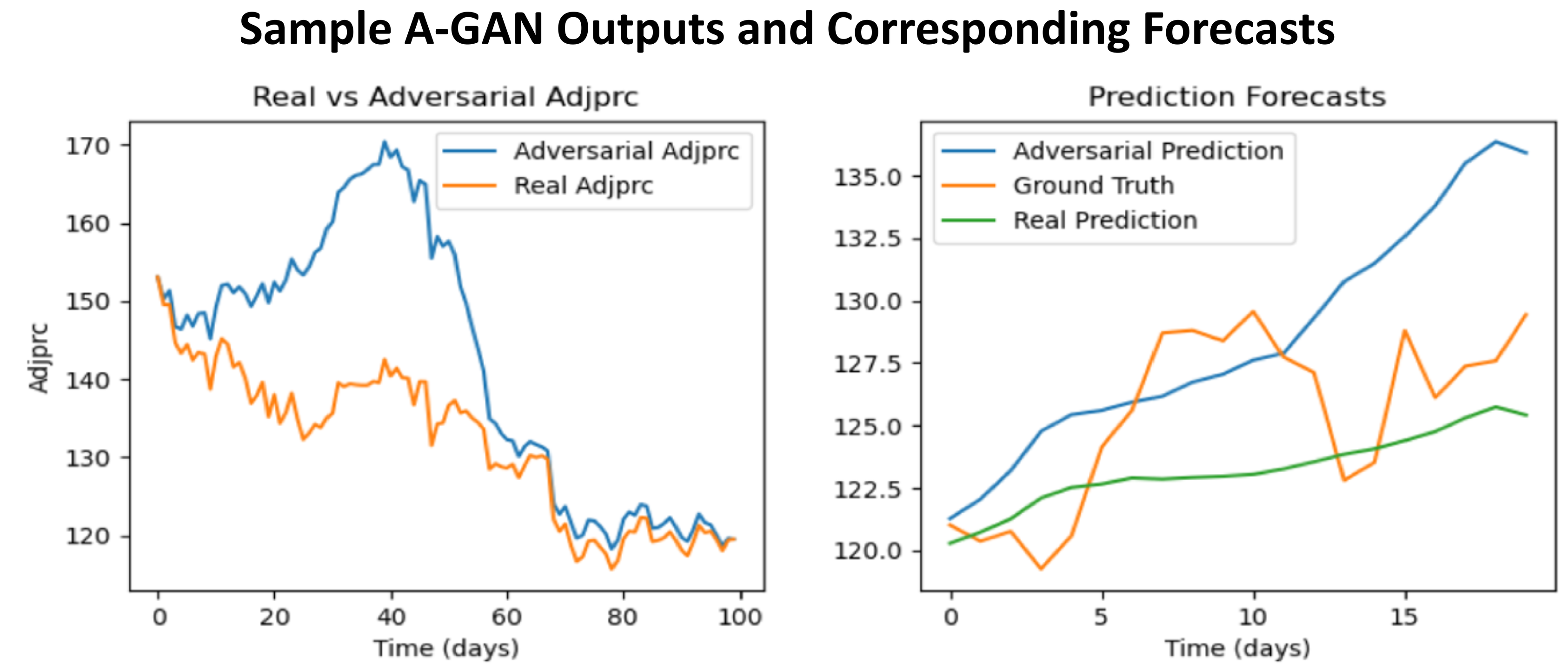


Stealthiness of the GSA and LSSA



(a) GSA (b) LSSA

Sample A-GAN Outputs and Corresponding Forecasts



Key Findings and Conclusions

- Not only do the GSA and LSSA show a drastic effect on a state-of-the-art N-HiTS architecture trained for financial forecasting, they also have the ability to covertly bypass standard security measures.
- Although the A-GAN has shown to successfully increase the slope of the victim model's forecast, the A-GAN suffers from mode collapse, which occurs when the generator creates data with limited diversity [2].

References

[1] Z. Shen and Y. Li, "Temporal characteristics-based adversarial attacks on time series forecasting," *Expert systems with applications*, vol. 264, Art. no. 125950, 2025, doi: 10.1016/j.eswa.2024.125950.

[2] Z. Dai, L. Zhao, K. Wang, and Y. Zhou, "Mode standardization: A practical countermeasure against mode collapse of GAN-based signal synthesis," *Applied soft computing*, vol. 150, Art. no. 111089, 2024, doi: 10.1016/j.asoc.2023.111089.

Acknowledgements

Thank you to Prof. Irene Huang, at the University of Toronto Scarborough, for supervising me during this CSCD94 project.

Further Information

For more information:

- GitHub: https://github.com/Dominik-luszcz/TargetedAdvGAN
- Email: dominik.Luszczynski@mail.utoronto.ca