Chaos, overfitting and equilibrium:

To what extent can machine learning beat the financial market?

Yaohao Peng
João Gabriel de Moraes Souza

Machine Learning Laboratory in Finance and Organizations - LAMFO/UnB

March 26, 2022

Summary

Abstract

- **Main Objective:** In this paper, we applied 10 technical analysis indicators to predict stock price movement directions using support vector machines, investigating the effects of hyperparameter variations on the out-of-sample classification performance and the profitability of the resulting trading strategies.

Abstract

I **Main Objective:** In this paper, we applied 10 technical analysis indicators to predict stock price movement directions using support vector machines, investigating the effects of hyperparameter variations on the out-of-sample classification performance and the profitability of the resulting trading strategies.

II **Main Contribution:** This study contributes to filling in the gap of the combined classic results in the finance literature (such as the Efficient Market Hypothesis) and empirical setbacks commonly seen in machine learning experiments, notably the occurrence of overfitting under the incorporation of high-dimensional non-linear interactions.

Abstract

I **Main Objective:** In this paper, we applied 10 technical analysis indicators to predict stock price movement directions using support vector machines, investigating the effects of hyperparameter variations on the out-of-sample classification performance and the profitability of the resulting trading strategies.

II **Main Contribution:** This study contributes to filling in the gap of the combined classic results in the finance literature (such as the Efficient Market Hypothesis) and empirical setbacks commonly seen in machine learning experiments, notably the occurrence of overfitting under the incorporation of high-dimensional non-linear interactions.

III **Data:** We collected daily data between April 1st, 2020 and November 19th, 2021 for the 30 firms that compose the Dow Jones Industrial Average (DJIA).

Abstract

Main Results

- Our results indicated that while the out-of-sample accuracy converged to 50%, 14.27% of the hyperparameter combinations yielded gains above the buy-and-hold strategy; on the other hand, no clear patterns about the best-performing hyperparameter combinations emerged, as the behavior of the out-of-sample performance was much more chaotic than its in-sample counterpart.

Summary

The Bias-Variance Dilemma

- One critical aspect of empirical applications of machine learning models is to maximize their generalization ability, by finding the ideal balance between in-sample fitness and complexity – often known as the bias-variance dilemma. The essence of generalization is to learn patterns from a sample of data and make inferences for out-of-sample data, and a good model should be complex enough to be able to learn the relevant patterns, but also simple enough to not incorporate the effect of noisy and irrelevant information on the final predictions.

The Bias-Variance Dilemma

- This result can be observed in Hoeffding's inequality [Hoeffding, 1963]:

$$\mathbb{P}(|E_{in}(h) - E_{out}(h)| > \varepsilon) \leq 2 \cdot \mathbb{M} \cdot e^{-2 \cdot \varepsilon^2 \cdot n} \tag{1}$$

where $E_{in}(h)$ and $E_{out}(h)$ are, respectively, the in-sample error and the out-of-sample error associated with a predicting model h; ε is user-specified margin, n is the sample size and $\mathbb{M}$ is a measure of the model's complexity, thus representing the overall variance level of the model and, consequently, the potential sensitivity of the out-of-sample predictions to small variations in the observed data.

The Bias-Variance Dilemma

- Expression **1** can be rearranged as:

$$\begin{aligned}\underbrace{E_{out}(h)}_{\text{out-of-sample generalization error}} &\leq E_{in}(h) + \sqrt{\frac{1}{2\cdot n} log\left(\frac{2\cdot \mathbb{M}}{\delta}\right)} \\ &\leq \underbrace{E_{in}(h)}_{\text{in-sample error}} + \underbrace{\Omega}_{\text{model complexity}} \qquad (2)\end{aligned}$$

where $\delta = 2 \cdot \mathbb{M} \cdot e^{-2\cdot\varepsilon^2\cdot n}$ is a constant and Ω represents the penalization for the complexity of the model. The bias-variance dilemma is concisely stated in expression **2**, showing that the generalization error E_{out} assumes higher values with both a high in-sample bias or a high out-of-sample variance (complexity).

Technical Analysis Indicators

Independent variable	References
Simple Moving Average	[Chiang et al., 2016, Gorenc Novak and Velušček, 2016, Kumar et al., 2016] [Chen and Hao, 2017, Gunduz et al., 2017, Li and Tam, 2017] [Shynkevich et al., 2017, Weng et al., 2017, Alhashel et al., 2018] [Henrique et al., 2018, Merello et al., 2019, Sezer and Ozbayoglu, 2018]
Exponential Moving Average	[Gorenc Novak and Velušček, 2016, Kumar et al., 2016, Chen et al., 2017] [Chen and Hao, 2017, Gunduz et al., 2017, Li and Tam, 2017] [Shynkevich et al., 2017, Weng et al., 2017, Alhashel et al., 2018] [Nakano et al., 2018, Sezer and Ozbayoglu, 2018]
Moving Average Convergence-Divergence	[Chiang et al., 2016, Kumar et al., 2016, Chen and Hao, 2017] [Gunduz et al., 2017, Li and Tam, 2017, Alhashel et al., 2018] [Nakano et al., 2018, Sezer and Ozbayoglu, 2018]
Momentum	[Chiang et al., 2016, Gorenc Novak and Velušček, 2016, Kumar et al., 2016] [Gunduz et al., 2017, Weng et al., 2017, Merello et al., 2019]
Accumulation/Distribution Oscillator	[Patel et al., 2015a, Patel et al., 2015b, Alhashel et al., 2018] [Henrique et al., 2018]
Rate of Change	[Gorenc Novak and Velušček, 2016, Kumar et al., 2016, Gunduz et al., 2017] [Li and Tam, 2017, Shynkevich et al., 2017, Weng et al., 2017] [Alhashel et al., 2018]
On Balance Volume	[Creamer, 2012, de Oliveira et al., 2013, Chen and Hao, 2017] [Nakano et al., 2018, Sezer and Ozbayoglu, 2018]
Relative Strength Index	[Gorenc Novak and Velušček, 2016, Chen and Hao, 2017, Gunduz et al., 2017] [Weng et al., 2017, Shynkevich et al., 2017, Li and Tam, 2017] [Alhashel et al., 2018, Henrique et al., 2018, Nakano et al., 2018] [Sezer and Ozbayoglu, 2018]

Technical Analysis Indicators

Independent variable	References
Stochastic K%	[Kumar et al., 2016, Gunduz et al., 2017, Shynkevich et al., 2017] [Alhashel et al., 2018, Nakano et al., 2018]
William's R%	[Kumar et al., 2016, Gunduz et al., 2017, Li and Tam, 2017] [Shynkevich et al., 2017, Alhashel et al., 2018, Sezer and Ozbayoglu, 2018]

Summary

Methods and Empirical Analysis

- SVM's formulation is defined as the following convex quadratic programming problem:

Methods and Empirical Analysis

- SVM's formulation is defined as the following convex quadratic programming problem:

$$\begin{aligned} \text{Minimize :} \quad & \frac{1}{2}\boldsymbol{w^T w} + C\boldsymbol{\xi^T 1} \\ \text{Subject to :} \quad & \boldsymbol{D}(\boldsymbol{\Phi w} - b\boldsymbol{1}) \geq \boldsymbol{1} - \boldsymbol{\xi} \\ & b \in \mathbb{R}, \boldsymbol{w} \in \mathbb{R}^q, \boldsymbol{\xi} \geq 0 \end{aligned} \tag{3}$$

where $C \in \mathbb{R}^+$ is a user-defined hyperparameter that represents the penalization for misclassified observations, ξ is a vector of slack variables, $\boldsymbol{D}$ is the diagonal square matrix of the class labels for each observation, Φ is the matrix of non-linear mappings applied to each pair of observations $\boldsymbol{x_i}$ and $\boldsymbol{x_j}, i, j = 1, 2, 3 \ldots, n$, and n is the sample size.

Methods and Empirical Analysis

- This optimization problem has a global optimal solution, and the decision function of the SVM classifier is given by:

Methods and Empirical Analysis

- This optimization problem has a global optimal solution, and the decision function of the SVM classifier is given by:

$$f(\boldsymbol{x_i}) = sgn\left(\sum_{j=1}^{n} \kappa(\boldsymbol{x_i}, \boldsymbol{x_j}) y_j \lambda_j - b\right) \tag{4}$$

where $sgn(.)$ is the sign function, $\kappa(\boldsymbol{x_i}, \boldsymbol{x_j}) = \varphi(\boldsymbol{x_i}) \cdot \varphi(\boldsymbol{x_j}) \in \mathbb{R}, i, j = 1, 2, 3 \ldots, n$ is the Kernel function that generalizes the inner product of the mappings φ for each pair of observations, $y_i \in \{-1, +1\}$ is the label of the i-th observation, λ_i is the i-th Lagrange multiplier of problem **4**, and $b \in \mathbb{R}$ is the bias term (intercept) of the decision function.

Methods and Empirical Analysis

- The data was split into two sequential and mutually exclusive subsets: the training set being composed by observations between April 1st, 2020 and July 30th, 2021, while the observations between August 1st, 2021 and November 19th, 2021 composed the test set. The dependent variable was the price direction movement between periods t and $t+1$, and the set of independent variables was composed by 10 technical analysis indicators commonly utilized in the recent literature of machine learning models applied to financial forecasting.

- For each model, we applied 10-fold cross-validation for the respective training set and applied the respective decision functions to the observations from the test set.

Methods and Empirical Analysis

- The grid of hyperparameters that we used is displayed in table **1** below:

Hyperparameter	Grid-search interval
C	$[e^{10^0}, e^{10^{0.25}} \ldots, e^{10^{9.75}}, e^{10^{10}}]$
σ	$[0, 0.25, \ldots, 9.75, 10]$

Table: Hyperparameters intervals used for the models' training

Summary

Results

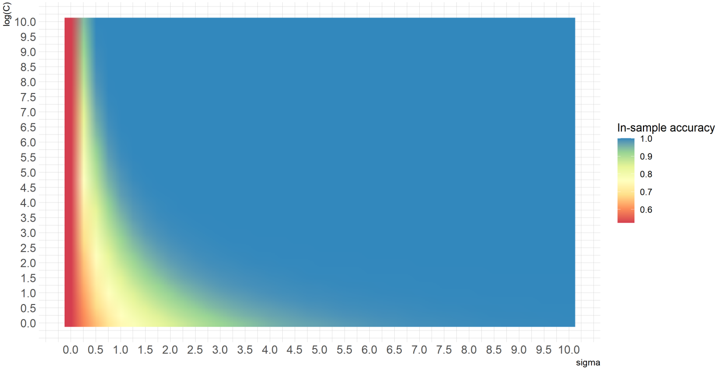


Figure: Heatmap of the models' in-sample accuracy

Results

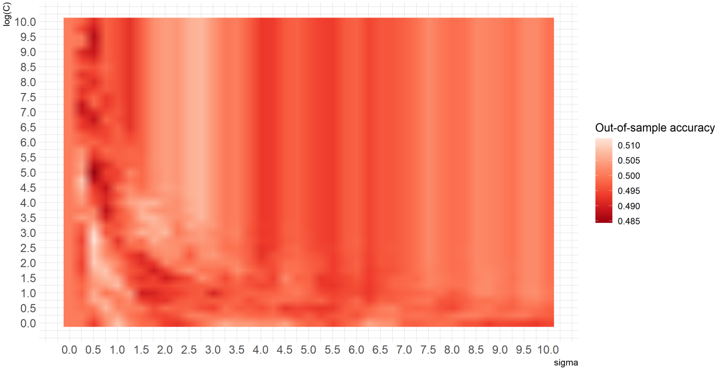


Figure: Heatmap of the models' out-of-sample accuracy

Results

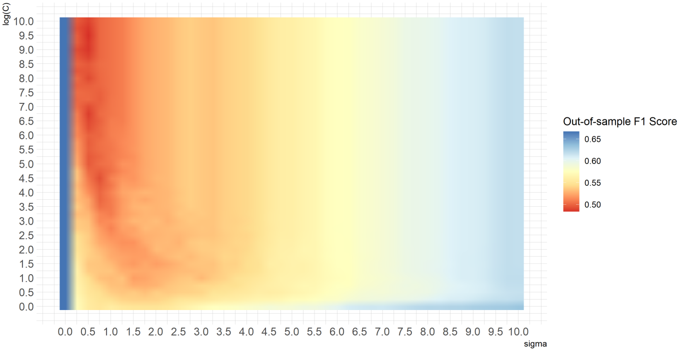


Figure: Heatmap of the model's out-of-sample F1 Score

Results

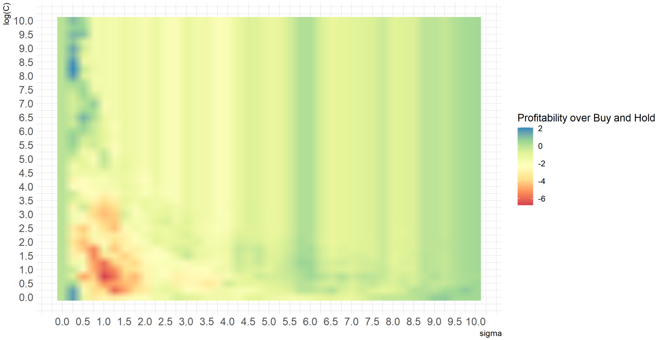
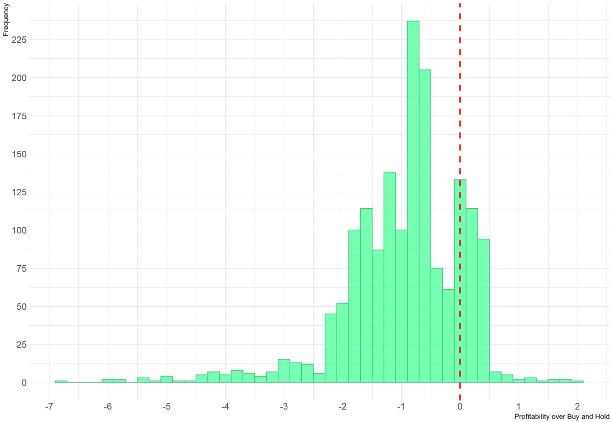


Figure: Heatmap of the model's out-of-sample profitability over the buy-and-hold strategy

Results


Frequency
225
200
175
150
125
100
75
50
25
0
-7
-6
-5
-4
-3
-2
-1
0
1
2
Profitability over Buy and Hold

Results

Metric	Strategy profitability	TC_0	TC_{BH}
Maximum	5.9422	1.9437	0.0765
Third quartile	3.6302	0.1262	−0.0094
Median	3.0899	0.1004	−0.0275
First quartile	2.3960	0.0690	−0.0435
Minimum	−2.8630	−0.0875	−0.2062
Mean	2.9321	0.1439	−0.0288
Standard deviation	0.9976	0.2879	0.0299
Skewness	−1.3164	5.9481	−1.1628
Kurtosis	6.8896	37.2863	6.9468
Buy-and-hold strategy mean profitability over the out-of-sample period: 3.38			

Table: Descriptive statistics of the models' profitability and maximum transaction costs

Summary

Final Remarks

- The results indicated that most models had a high in-sample accuracy but only around 50% when applied to out-of-sample data.

Final Remarks

- The results indicated that most models had a high in-sample accuracy but only around 50% when applied to out-of-sample data.
- Specifically, a number of models had an out-of-sample F1 Score close to 65%, however the models that yielded the best profitabilities followed a chaotic behavior, as the out-of-sample performance of the models showed significant oscillations for small variations on the hyperparameters.

Final Remarks

- The results indicated that most models had a high in-sample accuracy but only around 50% when applied to out-of-sample data.
- Specifically, a number of models had an out-of-sample F1 Score close to 65%, however the models that yielded the best profitabilities followed a chaotic behavior, as the out-of-sample performance of the models showed significant oscillations for small variations on the hyperparameters.
- The findings of this research provide an assessment of empirical challenges of machine learning applications to quantitative finance, while also adding to the literature of financial theory by exploring the connections between well-established concepts in finance and novel evidence brought by data science methods.

Final Remarks

- The results indicated that most models had a high in-sample accuracy but only around 50% when applied to out-of-sample data.
- Specifically, a number of models had an out-of-sample F1 Score close to 65%, however the models that yielded the best profitabilities followed a chaotic behavior, as the out-of-sample performance of the models showed significant oscillations for small variations on the hyperparameters.
- The findings of this research provide an assessment of empirical challenges of machine learning applications to quantitative finance, while also adding to the literature of financial theory by exploring the connections between well-established concepts in finance and novel evidence brought by data science methods.
- As topics for further research, we suggest improvements and additional cases to the combination of hyperparameters. For instance, the only predictive model that we used was the SVM with Gaussian Kernel, there is a range of Kernels that can be used in this same methodology that may lead to different results. In addition to combinations of other predictive methodologies such as ensemble-based models that can improve the predictive capacity of our strategies.

Bibliography I

Alhashel, B. S., Almudhaf, F. W., and Hansz, J. A. (2018).
Can technical analysis generate superior returns in securitized property markets? evidence from east asia markets.
Pacific-Basin Finance Journal, 47:92–108.
11, 12

Chen, H., Xiao, K., Sun, J., and Wu, S. (2017).
A double-layer neural network framework for high-frequency forecasting.
ACM Transactions on Management Information Systems (TMIS), 7(4):11.
11

Bibliography II

Chen, Y. and Hao, Y. (2017).
A feature weighted support vector machine and k-nearest neighbor algorithm for stock market indices prediction.
Expert Systems with Applications, 80:340–355.
11

Chiang, W.-C., Enke, D., Wu, T., and Wang, R. (2016).
An adaptive stock index trading decision support system.
Expert Systems with Applications, 59:195–207.
11

Bibliography III

Creamer, G. (2012).
Model calibration and automated trading agent for euro futures.
Quantitative Finance, 12(4):531–545.
11

de Oliveira, F. A., Nobre, C. N., and Zárate, L. E. (2013).
Applying artificial neural networks to prediction of stock price and improvement of the directional prediction index–case study of petr4, petrobras, brazil.
Expert Systems with Applications, 40(18):7596–7606.
11

Bibliography IV

Gorenc Novak, M. and Velušček, D. (2016).
Prediction of stock price movement based on daily high prices.
Quantitative Finance, 16(5):793–826.
11

Gunduz, H., Yaslan, Y., and Cataltepe, Z. (2017).
Intraday prediction of borsa istanbul using convolutional neural networks and feature correlations.
Knowledge-Based Systems, 137:138–148.
11, 12

Bibliography V

Hansen, P. R. and Lunde, A. (2005).
A forecast comparison of volatility models: does anything beat a garch (1, 1)?
Journal of applied econometrics, 20(7):873–889.

Henrique, B. M., Sobreiro, V. A., and Kimura, H. (2018).
Stock price prediction using support vector regression on daily and up to the minute prices.
The Journal of Finance and Data Science, 4(3):183–201.
11

Bibliography VI

Hoeffding, W. (1963).
Probability inequalities for sums of bounded random variables.
Journal of the American statistical association, 58(301):13–30.

9

Kumar, D., Meghwani, S. S., and Thakur, M. (2016).
Proximal support vector machine based hybrid prediction models for trend forecasting in financial markets.
Journal of Computational Science, 17:1–13.
11, 12

Bibliography VII

Li, Z. and Tam, V. (2017).
Combining the real-time wavelet denoising and long-short-term-memory neural network for predicting stock indexes.
In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–8. IEEE.
11, 12

Maronna, R. A. and Zamar, R. H. (2002).
Robust estimates of location and dispersion for high-dimensional datasets.
Technometrics, 44(4):307–317.

Bibliography VIII

Merello, S., Ratto, A. P., Oneto, L., and Cambria, E. (2019).
Ensemble application of transfer learning and sample weighting for stock market prediction.
In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.
11

Montenegro, M. R. and Albuquerque, P. H. M. (2017).
Wealth management: Modeling the nonlinear dependence.
Algorithmic Finance, 6(1-2):51–65.

Bibliography IX

Nakano, M., Takahashi, A., and Takahashi, S. (2018).
Bitcoin technical trading with artificial neural network.
Physica A: Statistical Mechanics and its Applications, 510:587 – 609.
11, 12

Patel, J., Shah, S., Thakkar, P., and Kotecha, K. (2015a).
Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques.
Expert Systems with Applications, 42(1):259–268.
11

Bibliography X

Patel, J., Shah, S., Thakkar, P., and Kotecha, K. (2015b).
Predicting stock market index using fusion of machine learning techniques.
Expert Systems with Applications, 42(4):2162–2172.
11

Rousseeuw, P. J. (1984).
Least median of squares regression.
Journal of the American statistical association, 79(388):871–880.

Bibliography XI

Sezer, O. B. and Ozbayoglu, A. M. (2018).
Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach.
Applied Soft Computing, 70:525–538.
11, 12

Shynkevich, Y., McGinnity, T. M., Coleman, S. A., Belatreche, A., and Li, Y. (2017).
Forecasting price movements using technical indicators: Investigating the impact of varying input window length.
Neurocomputing, 264:71–88.
11, 12

Bibliography XII

Weng, B., Ahmed, M. A., and Megahed, F. M. (2017).
Stock market one-day ahead movement prediction using disparate data sources.
Expert Systems with Applications, 79:153–163.
11