FIM2021

Kitakyushu, 26-28 December 2021

Black-Scholes Option Pricing using Machine Learning

S.Sood[1], N.Batra[1], T.Jain[1], Prof. H.C. Taneja (Mentor)
[1] Delhi Technological University, New Delhi, India
shreyansood_2k18mc110@dtu.ac.in, nishantbatra_2k18mc073@dtu.ac.in,
tanmayjain_2k18mc116@dtu.ac.in, hctaneja@dce.ac.in

Keywords—Black-Scholes Model, Option Greeks, Option Pricing, LSTM, Machine Learning, Moneyness, Volatility

Abstract:

(i) **aim and contributions:** The main objective of this paper is to explore the effectiveness of machine learning models in predicting stock option prices benchmarked by the Black-Scholes Model. We employ the following 4 machine learning models - Support Vector Machine (SVM), Extreme Gradient Boosting (XGB), Multilayer Perceptron (MLP) and Long Short Term Memory (LSTM), trained using 2 different set of input features, to predict option premiums based on the S&P 500 Apple Stock (APPL) option chain historical data from 2018 and 2019. Statistical analysis of the results show that LSTM is the best out of the chosen model for pricing both Call and Put options. Further analysis of the predictions based on Moneyness and Maturity show consistency of results with the expected behaviour that validates the effectiveness of the prediction model.

(ii) **background:** The Black-Scholes Model is one of the most fundamental and widely used financial models for pricing stock option premiums. However, due to the standard limitations and assumptions of the model, it is considered to be just a useful approximation tool or a robust framework for other models to build upon. Most research studies that attempt to discern the relevance of the Black-Scholes Model in real world scenarios conclude that the assumption of constant underlying volatility over the life of the derivative is the biggest contributing factor for the empirical inaccuracy of the model. Modifications based on the concepts of Stochastic Volatility and Jump Diffusion are widely implemented in the field of Financial Mathematics to correct the shortcomings of the Black-Scoles Model. The high dimensionality and flexibility of factors upon which option premiums depend makes the task of accurately predicting them extremely complex. Recently, the concept of Machine Learning, specifically, time-series forecasting using predictive models is finding much application in the field of Finance. In our approach to provide a solution for predicting option premiums accurately, we implement certain Machine Learning Models designed with the intent to effectively build upon and outperform the Black-Scholes Model while using the same set of input parameters and subsequently calculated option greeks. This approach of using option greeks as training input for option pricing prediction models is relatively unexplored and our research contributes to the scarce amount of existing literature which utilizes it. We compare and explore the behaviours and performance of different models with the benchmark predictions obtained from the Black-Scholes Model. We perform a comprehensive comparative statistical analysis of the best obtained results separately for call and put based on the moneyness and maturity values.

(iii) **methods:** We use the S&P 500 Apple (AAPL) stock option chain historical data from 2018 and 2019. After the necessary data cleaning and preprocessing, we compile 2 separate chronological datasets for Call and Put

Options with the Call Option data containing about 2,77,000 rows and the Put Option data containing about 2,42,000 rows. Each row of data contains the closing values of the Option Premium (C for Call, P for Put), Underlying Stock Price (S), Strike Price (K), Implied Volatility (σI) and Time to Expiration in years (t), as well as the values of the following Option Greeks - Delta (δ), Gamma (γ), Theta (θ), Vega (ν) and Rho (ρ). The Option Premiums serve as our output ground truth values whereas the rest serve as our input features. The 4 machine learning models are trained using 2 different sets of input features:

Set 1 which excludes option greeks i.e. it contains only 4 features - S, K, t and σI.

Set 2 that includes option greeks i.e. it contains all 9 features - S, K, t, σ_I, δ, γ, θ, ν and ρ.

Since the corresponding information regarding the Risk-Free Interest Rates (r) is unavailable, it is not used as an input feature. For the purpose of calculating the Black-Scholes Option Premiums, the average value of the U.S. 1 Year Treasury Rate across 2018 and 2019 is taken as R. Both Call and Put datasets are split into a 70:30 ratio chronologically for the purpose of generating training and testing datasets.

The 4 machine learning models we implement in addition to the Black Scholes Model: 1. SVM 2. XGB 3. MLP 4. LSTM

We analyze the predictions of the BS Model on both Call and Put Option datasets and compare it with our 4 regression models, each trained on the 2 sets of input parameters, which gives a total of 9 models for comparison. Out of these 9 models, we pick the model with the best results and further breakdown its performance based on the following 2 parameters: 1. Moneyness 2, Maturity

(iv) **results and concluding remarks**: The following 2 regression metrics are used in the evaluation of models: 1. Mean Absolute Error (MAE) 2. Root Mean Squared Error (RMSE). A combination of the metrics is taken as the main loss function for training and testing purposes. The following 2 tables represent the results for the metrics obtained from the testing dataset predictions of call options and put options respectively. The results show that LSTM is the best model for option pricing for both Call and Put Options. SVM and XGB perform worse than BS i.e. these models undefit and are unable to sufficiently capture the underlying relation for option premium predictions. Results also show that Option Greeks when included as additional input features actually produce worse final test metrics in every single model even though they improved the training metrics in some cases. This is a new finding as most of the literature about using option greeks as input features for option pricing mostly report an improvement in performance. Along with LSTM, MLP also outperforms BS. The metrics show similar distribution and trends for both types of options which implies that the training of the models is impartial to Option Type. The results from our statistical analysis show that when grouped by Moneyness, Out-of-the-Money options give significantly better metrics followed by At-the-Money and then In-the-Money samples. Similarly, the further results show that when grouped by Maturity, Short-Term options give better metrics followed by Medium-Term and then Long Term options. The analysis of the option predictions on the basis of Moneyness and Maturity validate that the LSTM prediction model is able to efficiently and accurately predict option prices in a real world market scenario.

(v) **key references:**

[1] Drucker, Harris, Christopher J. C. Burges, Linda Kaufman, Alex Smola and Vladimir Naumovich Vapnik. “Support Vector Regression Machines.” NIPS. (1996).

[2] Hochreiter, Sepp and Jürgen Schmidhuber. “Long Short-Term Memory.” Neural Computation 9, 1735-1780. (1997).

[3] Chen, Tianqi and Carlos Guestrin. “XGBoost: A Scalable Tree Boosting System.” Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2016).

[4] J. M. Hutchinson, A. W. Lo, and T. Poggio. A nonparametric approach to pricing and hedging derivative securities via learning networks. The Journal of Finance, 49(3):851–889. (1994).

[5] Ramazan Gencay and Aslihan Salih, “Degree of mispricing with the Black-Scholes model and nonparametric cures”, Annals of Economics andFinance, Vol. 4, pp. 73-101. (2003).

[6] Nikola Gradojevic, Ramazan Gencay, and Dragan Kukolj. Option pricing with modular neural networks. IEEE transactions on neural networks, 20(4):626–637. (2009).

[7] Palmer, Samuel. “Evolutionary algorithms and computational methods for derivatives pricing.” (2019).

[8] Robert Culkin and Sanjiv R Das. “Machine learning in finance: the case of deep learning for option pricing”. Journal of Investment Management, 15(4):92–100. (2017).

[9] Liu, Shuaiqiang, Cornelis W. Oosterlee and Sander M. Bohté. “Pricing options and computing implied volatilities using neural networks.” (2019).

[10] Ruf, Johannes and Weiguan Wang. “Neural networks for option pricing and hedging: a literature review.” Journal of Computational Finance (2020).