**Machine Learning in Option Market:**

**Training Neural Networks for Equity Option Pricing**

Rongdong Huang, Mengqi Chen, Ziwei Guo, Haoxing Liu

GR5265 Stochastic Methods in Finance

April 12, 2018

**Abstract**

In the world of Mathematical Finance, pricing of an American style option has always be challenging. The tradition Black and Scholes Model(1973) provided an elegant solution to general European options but failed to incorporate various issues such as dynamics of volatility and early exercises of options. In this project, instead of looking for a way to amend the existing model, we want to explore the possibility of Machine Learning in option pricing. First we will test its ability to learn closed-form mathematical expression, the Black-Scholes formula, by generating random inputs and matching the results to the Black-Scholes price. Then we will utilize large amount of market data from options based on equities of American companies, which are primarily American style options, to directly train the neural network to come up with its on pricing formula. In the end we will discuss the accuracy of our model and how input parameters affect the final results.

**1. Background Introduction**

**1.1 Neural Networks**

Neural Networks, a sub-branch of machine learning, is designed to simulate how human brains receive and construct perceptions. In simple mathematical terms, it’s similar to a function that takes in a set of values and returns an output. Details about the function, usually involving multiple layers of perceptrons, are entirely developed by the machine in the process of minimizing a given loss function that measures the differences between the outputs and the target values. Using backpropagation algorithm, the program will calculate the “gradient” of the loss function at each node and adjust the weights according to it in the next trail, or “batch”. Ideally, after filtering through the sample recursively, the program will arrive at an optimal set of parameters that generate stable results and errors.

Generally, Neural Networks have thousands or even millions of parameters and can learn any nonlinear function. The advancement in computational hardware like GPUs and the availability of big data make the powerful application possible in object detection, image captioning, auto-driving, etc. Besides, Neural Networks has also been applied to the finance world, for example, using recurrent neural networks to analyze large volume of news to conclude market sentiment towards some specific stock, building fully connected neural networks to predict SP500 minute index with its components’ share prices. More relatedly, Robert Culkin & Sanjiv R. Das (2017) and Hutchinson et al. (1994) have tried to price financial derivatives with neural networks and achieve good results. Our purpose in this part is to train our own neural networks to establish a pricing formula for American style options.

**1.2 American Equity Option Market**

The American option allows the option holders to exercise the option at any time prior to or until its maturity date. The underlying assets could be equity, bond, future, index, commodity and currency. Dow Jones Industrial Average (DJIA) is one of the most trusted and historical index which includes one third of largest and famous companies in American markets. In this paper, we chose 30 large public owned companies comprising of DJIA (needs a index to show 30 companies and date) equity options to analyze.

**2. Data Collection**

**2.1 Simulated Data**

First, simulate 1,000,000 European option data, where call options and put options take a half respectively, including risk free rate, dividend yield, volatility, maturity, spot price and strike prices. Then, plug these variables into the Black Scholes formula to obtain the option price.

**2.2 Market Data**

As of April 8, 2018, collect all the unexpired put/call equity option data of the Dow Jones components from Bloomberg, including risk free rate, dividend yield, implied volatility, days to expiration, strike price and option mid price. Noticeably, all of them are of American style. For further exploration, collect the 30 companies’ historical stock adjusted prices from April 8, 2017 to April 8, 2018 from Yahoo Finance.

**3. Implementation**

The implementation is divided into four scenarios. In Scenario I, train the neural network to explore whether it is possible to learn the Black Scholes formula from the simulated option variables and prices. Then dived into real market data. Since American options have no closed-form pricing formula, it is really exciting to challenge the power of neural networks. In Scenario II, put market data with implied volatility included into the neural network. In Scenario III, exclude the implied volatility. In Scenario IV, add the underlying historical stock prices. The training process is aimed at minimizing the mean squared error of model predicted option prices and BS formula prices/real market prices.

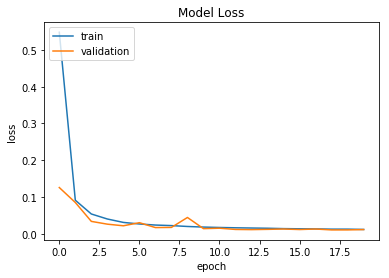
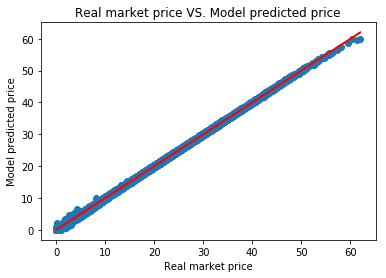
Our neural networks are built with Keras package. And model specifications are listed below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Input | Hidden Layers | Hidden States | Output | Unknown Parameters | Dataset Size |
| Scenario I | Simulated data: call/put type, risk free rate, dividend yield, volatility, maturity, spot price, strike price | 2 | 128 | Predicted option price | 17,665 | 100,000,000 |
| Scenario II | Market data: call/put type, risk free rate, dividend yield, implied volatility, maturity, spot price, strike price | 2 | 128 | Predicted option price | 17,665 | 12,668 |
| Scenario III | Market data: call/put type, risk free rate, dividend yield, spot price, strike price | 2 | 128 | Predicted option price | 17,665 | 14,190 |
| Scenario IV | Market data: call/put type, risk free rate, dividend yield, maturity, spot price, strike price, historical stock prices | 3 | 256 | Predicted option price | 197,337 | 14,190 |

*Note: For each scenario, we spare 20% of the whole dataset as test set and further split the remaining into training set and validation set with a ratio of 0.2.*

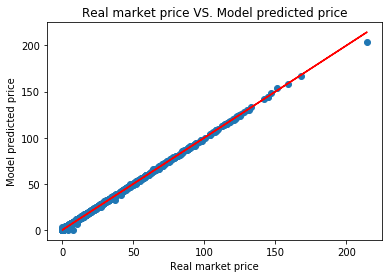
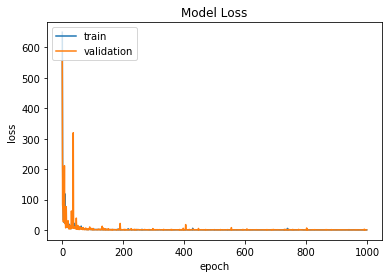
**4. Results and Analysis**

**Scenario I: Learning Black Scholes formula with simulated data**

As the training graph shows, both the training and validation losses drop quickly to almost zero. When performing the trained model on the test set, the mean square error is 0.011. Moreover, the price comparison is provided in test set which matches the diagonal line closely. Hence, neural network can learn the Black Scholes formula pretty fast and well.

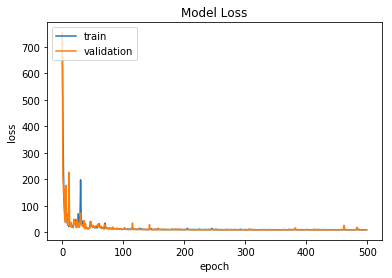
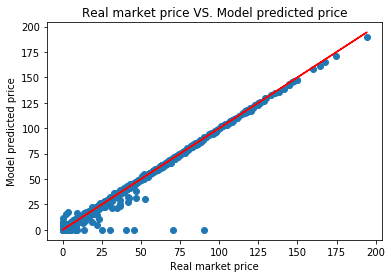
**Scenarios II: Learning American option pricing with market data (Implied volatility included)**

When it comes to American option with market data, the neural network still has excellent performance with both training and validation losses converging to zero, though the training process takes a little longer. On the test set, we obtain a mean square error of 0.445. The result is still satisfactory. But considering that the training is based on implied volatility, that is, there may be some predefined relationship imposed on the option variables and prices, we step further to exclude implied volatility.



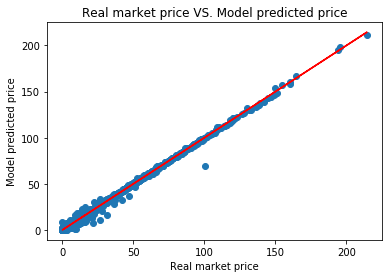
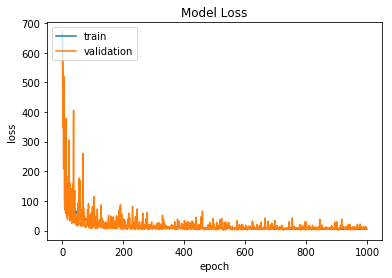
**Scenario III: Learning American option pricing with market data (Implied volatility excluded)**

When we excluded the implied volatility, the model performance deteriorated as expected. Both training and validation losses stop improving at level of 8. The mean square error for test set is 8.429, which is much larger compared to the previous one. In the test-set price comparison, we can observe a lot of mismatches. This is because the volatility information during training is totally abandoned. It proves that volatility plays an important role in option prices. To make up for that, historical stock prices is included into the model and then explore if the network itself can learn the volatility information from those prices.



**Scenario IV: Learning American option pricing with market data (Historical prices included)**

When historical prices are included into the model, the input variable size increases and hence enlarge the neural network. But meanwhile the unknown parameters increase by more than 11 times. Surprisingly, the result shows the neural network has the ability to infer volatility information from historical prices and hence more accurately price the American options. When applied to the test set, the model obtains a mean square error of 1.964 which is much lower than Scenario III though still higher than Scenario II. The predicted prices can approximately matches real prices. We can expect that with a larger dataset and even better selected input, the neural network can price the American options well without the implied volatility.



**5. Conclusion**

Neural networks have the ability to learn closed-form European option pricing formula and the final model generated by Neural Networks is quite fast and accurate.

When it comes to American option without an explicit pricing formula, neural networks can still have impressive performance with implied volatility as input or more well selected information, like the underlying historical prices, if implied volatility is excluded. Given enough market data, we could employ the deeper neural networks and expect to achieve state-of-art result.

**Reference**

*Black, F. and M. Scholes (1973), The pricing of options and corporate liabilities, Journal of Political Economy 81, 637-659.*

*Hutchinson, J. M., A. W. Lo, and T. Poggio (1994), A nonparametric approach to pricing and hedging derivative securities via learning networks. Journal of Finance 49 (3), 851-889.*

*Robert Culkin & Sanjiv R. Das(2017), Machine Learning in Finance: The Case of Deep Learning for Option Pricing, Santa Clara University*

*Ian Goodfellow and Yoshua Bengion and Aaron Courville(2016), Deep Learning, MIT Press*

*Lijuan Cao and Francis E.H. Tay(2001), Financial Forecasting Using Support Vector Machines, National University of Singapore*

**Appendix**

**Part 1 Top 5 and Worst 5 prices comparison in test set for Scenario I- IV.**

Scenario I: Learning Black Scholes formula with simulated data

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Model Predicted | BS Formula Price | Absolute Difference |  | Model Predicted | BS Formula Price | Absolute Difference |
| Top 1 | 1.0898 | 1.0902 | 0.0004 | Worst 1 | 59.8056 | 62.0365 | 2.2309 |
| Top 2 | 13.2221 | 13.2217 | 0.0004 | Worst 2 | 6.4104 | 4.2259 | 2.1845 |
| Top 3 | 6.9200 | 6.9204 | 0.0004 | Worst 3 | 10.0785 | 8.1626 | 1.9159 |
| Top 4 | 4.7562 | 4.7558 | 0.0004 | Worst 4 | 4.7620 | 2.8647 | 1.8973 |
| Top 5 | 0.5503 | 0.5499 | 0.0004 | Worst 5 | 2.0350 | 0.1507 | 1.8843 |

Scenarios II: Learning American option pricing with market data (Implied volatility included)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Model Predicted | Real Price | Absolute Difference |  | Model Predicted | Real Price | Absolute Difference |
| Top 1 | 42.5997 | 42.6000 | 0.0003 | Worst 1 | 203.9817 | 214.3750 | 10.3933 |
| Top 2 | 0.7456 | 0.7450 | 0.0006 | Worst 2 | 0 | 7.5000 | 7.500 |
| Top 3 | 57.4987 | 57.500 | 0.0013 | Worst 3 | 32.6378 | 37.2250 | 4.5872 |
| Top 4 | 0.3734 | 0.3750 | 0.0016 | Worst 4 | 0 | 4.0250 | 4.0250 |
| Top 5 | 31.3267 | 31.3250 | 0.0017 | Worst 5 | 6.1971 | 10.1500 | 3.9529 |

Scenario III: Learning American option pricing with market data (Implied volatility excluded)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Model Predicted | Real Price | Absolute Difference |  | Model Predicted | Real Price | Absolute Difference |
| Top 1 | 43.7248 | 43.7250 | 0.0002 | Worst 1 | 0 | 90.3750 | 90.3750 |
| Top 2 | 2.4702 | 2.4700 | 0.0002 | Worst 2 | 0 | 70.6500 | 70.6500 |
| Top 3 | 24.9496 | 24.9500 | 0.0004 | Worst 3 | 0 | 45.3750 | 45.3750 |
| Top 4 | 50.4762 | 50.4750 | 0.0012 | Worst 4 | 0 | 40.3750 | 40.3750 |
| Top 5 | 15.1275 | 15.1250 | 0.0025 | Worst 5 | 0 | 30.5000 | 30.5000 |

Scenario IV: Learning American option pricing with market data (Historical prices included)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Model Predicted | Real Price | Absolute Difference |  | Model Predicted | Real Price | Absolute Difference |
| Top 1 | 20.6499 | 20.6500 | 0.0001 | Worst 1 | 69.8399 | 100.3750 | 30.5351 |
| Top 2 | 2.8403 | 2.8400 | 0.0003 | Worst 2 | 16.1906 | 31.7000 | 15.5094 |
| Top 3 | 33.5755 | 33.5750 | 0.0005 | Worst 3 | 11.2859 | 26.7000 | 15.4141 |
| Top 4 | 43.7983 | 43.8000 | 0.0017 | Worst 4 | 7.2137 | 21.7000 | 14.4863 |
| Top 5 | 0.2083 | 0.2100 | 0.0017 | Worst 5 | 31.9789 | 41.9500 | 9.9711 |