Introduction

Options pricing is a hot topic in the derivatives market. The Black Scholes (BSM) Model is perhaps the most famous formula in all of finance to achieve this task with great accuracy. It is a partial differential equation that requires 5 input variables: the strike price of an option, the spot price, the time to expiration, the risk-free rate, and the volatility. However, it makes certain assumptions that are not often realistic, and predictions deviate from real-world results. For example, the model can only be used to price European options as it does not take into account the fact that American options can be exercised before the contract expiration date. Therefore, some more accurate models have been developed, such as the Binomial Option Pricing Model (BOPM). Applying deep learning to options pricing is a promising field, with not a lot of prominent research being done on the topic at this point in time.

Assumptions

I will often refer to the BSM model in this project as it serves as the perfect baseline for my work. The following are the assumptions of the BSM model.

- No dividends are paid out during the life of the option.
- Markets are random because market movements can't be predicted.
- There are no transaction costs in buying the option.
- The risk-free rate and volatility of the underlying asset are known and constant.
- The returns of the underlying asset are normally distributed.
- The option is European and can only be exercised at expiration.

Rationale

From the assumptions, the BSM model makes many parametric assumptions about pricing options, which limits usefulness. For example, it assumes that the returns of the underlying asset are normally distributed. Deep learning excels at approximating non-linearities given enough data and neurons (I will ignore the complexities of bias-variance trade-off for the time being). Moreover, the classical BSM model only takes in 5 parameters as previously mentioned. Therefore, I believe that if a deep learning model can relax some of these assumptions and incorporate more of these features as part of the model, the project can be quite successful.

Model Choice

While there are numerous deep learning models to choose from, one thing that is worth experimenting with is the association of data with relation to time. In this project, I am curious to explore how a CNN-LSTM model retains long-term dependencies and extracts features from time-series data.

Research Target

Build and fine-tune a CNN-LSTM model on historical equity options price data, which will provide a more reliable model to price options by relaxing some of the strong assumptions of the Black Scholes model. The CNN-LSTM should be able to perform better than the Black Scholes model in order for the project to be a success.

Research Plan and Deliverables

1. Find a dataset for training. Some candidate datasets are below:
    - https://datarade.ai/data-products/optionmetrics-ivydb-us-historical-option-prices-and-implied-optionmetrics
    - https://www.barchart.com/options/price-history
2. Build the first step of a data pipeline that will clean and extract useful features from the dataset (stock price, strike price, implied volatility, etc.). The following operations may be performed:
    - Noise removal
    - Normalisation
    - Dimensionality reduction
3. Build a CNN-LSTM model that will predict the premium price of a single US call option
4. Train the model on the training dataset split
    - Test different model architectures
    - Tune hyperparameters such as learning rate
    - Optimize model
5. Measure the accuracy of the model with the testing data
    - Cross-validation
6. Assess and compare model accuracy with respect to the baseline Black Scholes model

Optional tasks, if time permits:

7. Generalize the model to work on both European and American put and call options
8. Deploy the model