Machine Learning In Quantitative Finance: A kernel of truth

Vulgarizing Summary

Lennert Van der Schraelen
KULeuven

2020
June

In this document, we give a non-technical summary of the thesis 'Machine Learning In Quantitative Finance: A kernel of truth' which aims to be understandable for a broad audience.

An option is a financial product which gives the investor the right, not an obligation, to buy or sell a stock satisfying some predetermined terms and conditions. After calibrating a financial model with the market, one often relies on time-consuming techniques in order to price such kind of products. In the financial world, this can be problematic since these calculations needs to be repeated on thousands of option prices. In this thesis, we propose to train a statistical model (a regression model) which tries to predict the price of an option in a fast way using the parameters of the product and the financial model. However, the price we have to pay for this speed-up is some loss in accuracy.

Our regression model needs to be able to predict in an accurate way for a broad parameter space. We use a machine learning technique called 'Gaussian process regression' (GPR) which is a technique achieving very good results. We derive this regression model ourself and give the necessary theoretical background. However, the amount of calculations for standard GPR rises cubic with the data. Hence, it is infeasible to use this model when one wants to train really large datasets. Although pricing options using standard GPR achieves a huge speed up, the more data points we use the slower the prediction time.

Since we are able to generate a dataset using the standard time-consuming techniques for pricing options, one can try to exploit the structure of the data. Despite these methods seems promising, we obtain problems by using models and options which needs to a lot of parameters to price. Hence, this is not an appropriate solution for our problem.

Next, we propose the use of the so called sparse methods. We try to approximate the exact Gaussian model using a set of 'inducing inputs'. Using these methods, we can train in a very fast way. Furthermore, the prediction time is now only dependent on the amount of inducing points and not on the amount of data points. However, we again loose some accuracy. Nonetheless, we still obtain problems for training for a large dataset and introduce a model which enables training using only a small part of the data at the time.

Subsequently, we discuss methods which only need a product with a matrix and a vector in order to train the regression model. Using this, we try to exploit the structure of the inducing points and some special form of objects we used in our model. Furthermore, we discuss a whole new framework enabling much faster training of exact GPR and also achieving a considerable speed-up for the other models. This framework also allow us to execute our calculations in parallel which can be exploited. We also include a section using advanced learning techniques, i.e. Bayesian and deep methods. However, these methods often slow down calculations and predictions which makes them less suitable for our purpose.

Finally, we give a extensive data study pricing options and comparing most of the models we discussed. We conclude that GPR and its extensions are very suitable for pricing options since we are able to price options in a very accurate and fast way by using models which are very feasible to train.