Machine Learning for Quantitative Finance

FOREX Forecasting / General Trading

Prepared by: Kevin Jeswani, WeCloudData Data Science Bootcamp, 2022

Professionals in the financial services industry have always been heavily tuned into advanced applied mathematics for forecasting, risk evaluation, and decision making. Machine learning (ML) is used heavily in the industry, with deep learning (DL) approaches catching on. In parallel, online platform providers offering sleek, simple, and affordable software solutions are expanding into every market. These platforms include QuantConnect, Quantrocket, Alpaca, just to name a few – each with the purpose of bringing quantitative finance into the hands of people without traditional finance backgrounds. The beauty of it all lies in the interface of terabytes of historical and live data, the ability to store, manipulate, and build powerful machine learning and deep learning models all at users' finger tips, then deploy them to make live trades. As such, this analytics plan is geared to clients who have always been interested in day trading, but need a helping hand with building and deploying machine learning models.

State of the Industry

The potential of machine learning and big data for quantitative finance and casual/personal algorithmic trading is colossal. Machine learning use cases in finance include general process automation, security enhancement, underwriting and credit scoring, chatbots, and algorithmic trading – the latter being of most interest to this plan.

Data collection

Application Programming Interfaces (APIs) provide access to historical data (intra- and end-of-day) and live data (within milli-seconds) for high frequency trading. This information is directly embedded into the programming environment and the data can be parsed into meaningful structural, simplified, then stored in large databases.

Forecasting

ML can be harnessed to take hundreds of thousands of lines of financial information and process it into meaningful insights. In trading, this would mean stock prediction which can be done with various levels of model complexity for short or long term needs. Technical analysis that is typically taught to traders can be arduous to deal with as there are many different strategies to encompass when trying to trade. ML has the potential to process and package these classical best practices into production-line models.

Automated Trading

Automated trading goes beyond simple process automation and enters the realm of utilizing continual machine learning model training with the potential of deep reinforcement learning. The ability to train an agent to make sounds trading decision is achieved by providing rewards for successful trades and punished for poor decisions. Essentially, trading gamification

Phase 1

Data collection

An API service such as EODHD will be selected and subscription package will be purchased. Data will then be gathered for historical data of forex exchange rates, stock options, stock price, stock fundamental data, and other available indicators. While only a small fraction of data will be used to develop baseline models, the rest of the data will be stored and assembled in Phase 2.

Exploration

To flatten the learning curve, a handful of FOREX pairs will be analyzed (10-15). Interactions between them will be studied (correlation & cointegration). The seasonality and trends in the currency pairs will be examined closely to gain insight into what models may be the most appropriate to use in forecasting this data. Simple statistical tests will be performed to support this along with graphical representations of these seasonality and trend features.

Data Modelling

Classical time series forecasting approaches will first be explored to build on the fundamental breakdown of trend and seasonality. Once these basics are understood, boosted tree models will be tested with SkForecast, which allows for time-series cross validation and training multivariate time series in a recursive multi-step environment while account for exogenous variables. Preliminary testing with an LSTM neural net will be done.

Phase 2

Big Data Assembly

The gathered stock data will be assembled into a proper database management system so that big data tools such as pyspark can be used to parse and analyze the data. Data enhancement may also be conducted to collect news articles either from the EODHD API or by webscraping. The database will ideally encompass 20+ forex pairs and go back further into the past if possible. A potential pivot may be to collect data on a stock index such as SPX and all the 500 stocks to perform a meta analysis of stock options data.

Data Modelling

This phase will dive deeper into hybrid models, specifically with neural nets. A greater hyperparameter tuning space will be explored for these exercises. Big data techniques will have to be implemented in handling this processing, and cloud computing will most likely have to be employed.

Deep Reinforcement Learning

Deep reinforcement learning will be explored with the aim of testing the forecasting models in a simulated trading environment. Packages like FinRL already exist for this purpose, which is connected to OpenAI Gym.