Frequently Asked Questions

Contents

---

1. What are the applications of machine learning in trading?

In trading, quants utilise machine learning algorithms for forecasting prices, for sentiment analysis, for finding patterns, and for adjusting high-frequency trading machines with changing markets. Forecasting simply means predicting the prices of any number of asset classes in the future. Also, in trading, one can use machine learning for inference, which is the process of finding the variables impacting the stock value.

Sentiment analysis implies analysing the sentiment of daily news, tweets, and other social media posts above certain topics. After analyzing, the machine summarises the strength of positive and negative information. These summaries can also be categorized on the basis of emotions or sentiments for easy access.

Patterns imply the consistency of datasets which is crucial for any trader. In case there are irregularities in the dataset, there will be faulty outcomes or predictions.

High-frequency trading machines imply the set up for facilitating buying and selling of stocks for traders. This setup works at a very fast speed and tuning changes in the market is a difficult task. But, machine learning algorithms can do it quickly. This helps save a lot of time and efforts of the traders.

2. How to implement machine learning in trading?

There are a few steps for implementing machine learning in trading, which are as follows:

Step 1 - Define the problem statement

You must define what you are trying to predict in the first step. It can be the future price, future return, . ll signal, optimization of portfolio allocation etc.

Step 2 - Collect and pre-process data

There should be reliable data with no duplicacies, redundancies, or any such issues which make data unfit for feeding to the algorithm since unclean data leads to erroneous results.

Step 3 - Split data

Then you need to split the data into training and testing datasets. Training dataset helps to train the machine learning algorithm , whereas, testing dataset helps to test the trained information.

Step 4 - Feature engineering

Feature engineering is important for finding the most relevant features of the dataset which describe the problem aptly. This way, the predictive power increases and the result turns out to be accurate.

Step 5 - Model selection

Then the appropriate model needs to be selected on the basis of the choses problem or input.

Step 6 - Train, validate and optimize

In this step, you train the model on the training data, then measure the performance on validation data and again go back to optimize, re-train and evaluate again until accuracy is achieved.

Step 7 - Backtest your test data

In the last step, you will backtest the test data which was taken out in the 3rd step. Thus, the final and optimized model from the previous step is run on the test data. This helps you see the realistic expectation of how the model will perform on the unseen market data when you start trading live.

3. Which machine learning books are helpful?

For beginners

Paid books:

- 6 Practical Books for Beginning Machine Learning
- Best Machine Learning Resources for Getting Started
- Machine Learning
- Machine Learning using Python

Free books:

- The Hundred-Page Machine Learning Book
- Neural Networks & Deep Learning, Nielsen (FREE)

- Understanding Machine Learning: From Theory to Algorithms
- Reinforcement Learning: An Introduction, Second Edition
- Support Vector Machines Succinctly

For advanced learners

Paid books:

- Statistics Books for Machine Learning
- Top Resources for Learning Linear Algebra for Machine Learning
- Top Books on Natural Language Processing
- Practical Machine Learning Books for the Holidays
- Python Machine Learning Books
- Hands-On Machine Learning for Algorithmic Trading

Free books:

- Python Data Science Handbook
- Natural Language Processing with Python
- Automate the Boring Stuff with Python
- Think Bayes

4. What is the relevance of Python in machine learning?

Basically, machine learning helps the traders with complex mathematical/statistical computations. For example, machine learning provides algorithms which automate the processes like predicting the market trend. Be it machine learning or any other set up involving complex mathematical/statistical computations, Python is a computer language which is easier to use as compared with other computer languages. It is developed to be a faster application for programming. Moreover, it does not have a complex syntax, unlike C++. There are few advantages listed below which you can take a look at:

- Python is a free, open-source, and cross-platform language, which has a rich library for almost every task imaginable and a specialized research environment.
- Python is very easy to learn. It is almost like reading English language and provides simple and efficient handling of data structures.
- It is an excellent choice for automated trading when the trading frequency is low to medium, and for trades which do not last less than a few seconds.
- It has multiple APIs and libraries that can be linked to make it optimal, cheaper, and allow greater exploratory development of multiple trade ideas.

5. Which firms use ML in their trading strategies?

Although there are several firms using ML in their trading strategies, we have listed a few major firms here. The list goes like this:

- J.P Morgan Chase (New York)

- Morgan Stanley (New York)
- Goldman Sachs (New York)
- Bridge Water Associates (Westport, Conn.)
- Equbot (San Francisco)
- Renaissance Technologies (New York)

6. Is there a benchmark to tell me if my strategy is really good or not?

There are a few ways with which you can evaluate the effectiveness of your strategy, and they are:

1. Maximum drawdown - It measures the peak-to-trough decline in the value of the portfolio.
2. Sharpe ratio - The Sharpe ratio is the excess return calculated as total returns less the risk-free rate of return per unit of volatility. Generally, risk-free return is the return on the risk-free assets such as government bonds. The excess returns are due to the 'extra risk' taken by the investor on investing in risky assets.
3. Annualised volatility - This shows the volatility on an annual basis.
4. Annualised returns - This is the average amount of money earned each year over a period of time.

8. What are the common mistakes while using machine learning for trading?

Common mistakes while using machine learning include:

- Using default-loss function as it is

This implies that the model using default-loss function may not be able to align with your objective. It is so because the default loss function of binary classifiers weighs the false positives as well as the false negatives equally. For example, in case you want to predict as to when a downtrend of stock may happen. In this scenario, the loss function should penalise false negatives more and also in proportion to the value of the stock.

- Using linear models for non-linear work

Although linear regression is simple, there are many nonlinear interactions between the predictor variables which are needed to be encoded manually. For example, if you want to predict the trend of the market. And, in case of favourable events, you want the trade order price to not be more than $60 a share. Also, you want the order to be cancelled if not executed. In such a scenario, you will need models with support vector machine or tree-based classifiers which can help with non-linear interactions.

- Forgetting about outliers

You should never forget about the outliers since they may cause you to feed faulty data to the algorithm. For example, during the prediction of the value of your stocks, in case unusual spikes in the value are shown, you must figure out the reason behind the spike. If the reason is something like a mechanical error or a measurement error, it is required to filter these outliers. Or else, the data you feed to the algorithm will provide erroneous results.

9. Which papers published for machine learning in trading are relevant?

There are some of the papers we recommend, and these are:

- Machine Learning for Trading

In multi-period trading with realistic market impact, determining the dynamic trading strategy that optimizes expected utility of final wealth is a hard problem. In this paper we show that, with an appropriate choice of the reward function, reinforcement learning techniques (specifically, Q-learning) can successfully handle the risk-averse case. We provide a proof of concept in the form of a simulated market which permits a statistical arbitrage even with trading costs. The Q-learning agent finds and exploits this arbitrage.

- Machine Learning in Finance: A Topic Modeling Approach

This paper provides a first comprehensive structuring of the literature applying machine learning to finance. They have used a probabilistic topic modelling approach to make sense of this diverse body of research spanning across the disciplines of finance, economics, computer sciences, and decision sciences.

- Applying Machine Learning to Trading Strategies

This paper proposes a machine learning approach to building investment strategies that addresses several drawbacks of a classic approach. To demonstrate our approach, we use a logistic regression algorithm to build a time-series dual momentum trading strategy on the S&P 500 Index. Our algorithm outperforms both buy-and-hold and several base-case dual momentum strategies, significantly increasing returns and reducing risk. Applying the algorithm to other U.S. and international large capitalization equity indices generally yields improvements in risk-adjusted performance.