ML in finance paper

ten applications of financial machine learning

the complexity of financial systems have exceeded the modelling capability of basic **econometric** quantitative analysis. Also the data utilised for financial decision making has evolved form simple numbers and datasets.

Machine learning algorithms can identify complex patterns in financial datasets, but it can also easily overfit, leading to false discoveries. But ignoring the hype around ML in finance ML has demonstrated is prowess in the field and many companies old and new are rushing to explore, develop and implement their own computer assisted financial decision making.

Financial applications of ML

1. Asset Pricing

Machine learning can model nonlinear predictor interactions and volatility better than standard economic method in predicting asset valuations. The asset pricing or risk premium extracted from the machine learning models can also help in understanding complex economic mechanisms under the hood. Machine learning can also be used to identify similar assets to conduct relative value analysis.
Also ML can be used in generation of synthetic datasets for backtesting investment strategies using Variational Autoencoder allowing researchers to obtain enough data to prevent overfitting in model and backtesting.

2. Risk Management and Portfolio Construction

Standard hedging methods involves several theoretical assumptions and ignores market frictions such as transaction costs, liquidity constraints etc. This hampers in generating more accurate hedging positions and higher frequency.
Deep Reinforcement learning models are purely empirical, greek and model free approach that relies on few theoretical assumptions and hence generate higher accuracy hedges. The deep reinforcement model incorporate hierarchical

relationships which are not recognised by traditional methods, clustering together highly correlated securities creating more efficient solutions.

These hierarchical clustering algorithms for construction of investment portfolio have shown to have better performance compared to standard Markowitz Critical line Algorithm.

3. Outlier Detection

Outliers are observations that differs from the rest of the observations due to structural break or regime switch. many ml models like regression models are sensitive to outliers. So algorithms like RANSAC can deal with outliers preventing them from biasing model estimates.

4. Bet Sizing

For any primary model taking buy-sell decision, there needs to be a secondary model that decides the size of the bet. A meta labeling classification algorithm can learn betting sizing by predicting the gain or loss from the primary model predictions.

5. Sentiment Analysis

Sentiment analysis methods are used to identify sentiment of text. it is used to analyze sentiments expressed in public filing texts like 10-Ks etc press release, conference calls, news articles and analyst calls,and public sentiment. and learn the price reaction to the sentiments

6. Feature Importance

Traditional regression methods require researchers to variables provide all variables (variables selection) and establish correct structural equation that bind those variables. It is important to get both of the part correct, ie getting the variables right is conditional on getting the specification right and vice versa.

ML methods like mean decrease impurity and mean decrease accuracy can help in selecting the important features, extracting the important features from the redundant and noise features.
algorithm indicates the presence of strong noise and multicollinearity which can help in understanding the relations of the features with each other

7. Credit Rating and Analyst Recommendations

credit rating agencies involve complex logic that can't be reflected by simple equations and formulas. But ML algorithms have been successful in replicating recommendations produced by credit rating agencies, and hence have been actively incorporated to supporting the analysts.

8. Execution

ML algorithms can price an low frequently traded bond based on modeling liquidity and similar trades based on their common features and and derive theoretical price and learn to execute the trade at most efficient price.

9. Big Data Analysis

The quantity and granularity of economic data has improved dramatically over the past few
years. the large collection of data can provide unmatched insight into the economic functionings, but handling of such unstructured data it difficult. ml can help in recognizing the complex relationships in the data and give a clearer and unbiased economic statistics.

One successful example is the use of web-scraping algorithms by MIT's Billion Prices Project
(BPP). This project collects daily price fluctuations associated with tens of millions of products
sold by thousands of online retailers in almost 100 countries, which allows BPP to produce daily
statistics of inflation and purchasing power parities. Specifically, BPP applies unsupervised ML
to classify a large number of items into unique products . In countries where inflation statistics do not exist, are inaccurate, or are manipulated, BPP allows researchers to obtain real-time statistics with important implications for debt and foreign exchange markets.

10. Controlling for Effects and Interactions

ml can be used to analyze the effect of a variable X on variable Y which is also dependent on variable Z. here we can replace the variable z with the prediction for

Y based on the variable Z hence helping us understand the effect of X while controlling the effect of variable Z , hence helping us understand interactions by modelling complex interaction effects, involving hierarchical, nonlinear and non-continuous relationships.

Conclusion

Machine learning techniques have not only performed better in several standard financial problems traditionally approached with standard tools but also solved many problems beyond the reach earlier.But it is important to analyze a result of a ml model for risk of overfitting, and other problems. Overall the future of finance and econometrics deeply involves AI and computations.