Machine Learning Applications in Asset Management

Alberto G Rossi

17 89
GEORGETOWN UNIVERSITY

Plan of the Talk

- What is Machine Learning?
- Why is ML useful in Finance and Asset Management?
- Machine Learning in Asset Management. Why Now?
- A Cautionary Tale of the

  "*Promises of Machine Learning in Asset Management*"

Machine Learning

What is Machine Learning?

- Subset of artificial intelligence
- Process of teaching a computer how to learn from data

What Does It Do?

- Identifies patterns and relationships in the training data
- Allows for the prediction of future values and events

Prediction is CRUCIAL in decision-making under uncertainty

Finance IS decision-making under uncertainty

- Credit decisions (mortgages, loans, ...): Predict default
- Risk Management: Predict bad scenarios
- Financial advice: Predict clients' needs
- Fraud detection: Predict when a transaction is fraudulent
- ...
- **Asset Management: Predict stock returns**

AI/ML in Asset Management-I (Time-series Predictability)

Consider the simplest possible investment decision:

- You want to create a portfolio for one year using
  - A riskless security (e.g., T-Bills) with return R_f
  - A risky asset (e.g., S&P 500 index) with R
- You have a *quadratic utility function*,
  - $U(E[R_p], \sigma) = E[R_p] - \gamma\sigma^2$
  - Increasing in $E[R_p]$: you like higher expected returns
  - Decreasing in σ: you don't like risk
  - γ is the coefficient of risk aversion: how risk affects you

AI/ML in Asset Management-II (Time-series Predictability)

Main Goal of Asset Allocation

- Choose fraction "*w*" in risky asset to maximize utility
- Formally, we want to maximize utility with respect to w:

$$\max_{w} \{U(E[R_p], \sigma)\} = \max_{w} \left\{ \underbrace{w\mathrm{E}[R] + (1-w)R_f}_{\text{Expected Return}} - \underbrace{\gamma\left(w^2\sigma^2\right)}_{\text{Penalty for Risk}} \right\}$$

- Take the derivative wrt w, and set it equal to zero:

$$\mathrm{E}[R] - R_f - 2\gamma w\sigma^2 = 0$$

- Solve it for w and obtain:

$$w^* = \frac{\mathrm{E}[R] - R_f}{2\gamma\sigma^2}$$

AI/ML in Asset Management-III
(Time-series Predictability)

- Consider the optimal-weight formula:

$$w^* = \frac{\mathrm{E}[R] - R_f}{2\gamma\sigma^2}$$

- Keeping everything else fixed, invest:
  1. More in stocks if risk premium increases, $\mathrm{E}[R] - R_f \uparrow$
  2. Less in stocks if they become riskier, $\sigma \uparrow$
  3. Less in stocks if you become more risk averse, $\gamma \uparrow$

- Modeling $\mathrm{E}[\mathrm{R}]$ and σ is crucial if it changes over time
  (The more precise they are, the better-off you are)

AI/ML in Asset Management-IV
(Cross-Sectional Predictability)

Usually performed via Factor investing:

- Form portfolios of stocks sorted based on a characteristic
- For example, Size Factor: Return of a portfolio
  - Long small capitalization stocks
  - Short large capitalization stocks
- These portfolios should generate high returns on average
- Can use at most 2 or 3 characteristics to create portfolios

AI/ML in Asset Management-V (Cross-Sectional Predictability)

BUT. The literature discovered 300+ factors!

Paper available here

Table 1
Factor classification

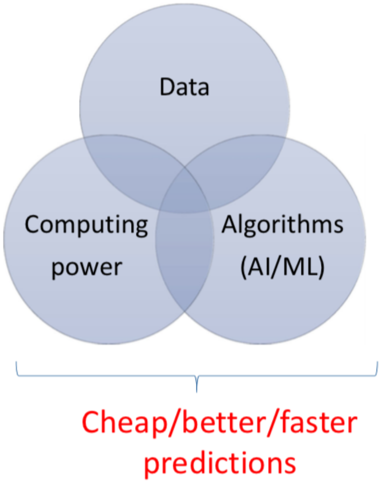
Risk type		Description	Examples
Common (113)	**Financial** (46)	Proxy for aggregate financial market movement, including market portfolio returns, volatility, squared market returns, among others	Sharpe (1964): market returns; Kraus and Litzenberger (1976): squared market returns
	Macro (40)	Proxy for movement in macroeconomic fundamentals, including consumption, investment, inflation, among others	Breeden (1979): consumption growth; Cochrane (1991): investment returns
	Microstructure (11)	Proxy for aggregate movements in market microstructure or financial market frictions, including liquidity, transaction costs, among others	Pastor and Stambaugh (2003): market liquidity; Lo and Wang (2006): market trading volume
	Behavioral (3)	Proxy for aggregate movements in investor behavior, sentiment or behavior-driven systematic mispricing	Baker and Wurgler (2006): investor sentiment; Hirshleifer and Jiang (2010): market mispricing
	Accounting (8)	Proxy for aggregate movement in firm-level accounting variables, including payout yield, cash flow, among others	Fama and French (1992): size and book-to-market; Da and Warachka (2009): cash flow
	Other (5)	Proxy for aggregate movements that do not fall into the above categories, including momentum, investors' beliefs, among others	Carhart (1997): return momentum; Ozoguz (2009): investors' beliefs
Characteristics (202)	**Financial** (61)	Proxy for firm-level idiosyncratic financial risks, including volatility, extreme returns, among others	Ang et al. (2006): idiosyncratic volatility; Bali, Cakici, and Whitelaw (2011): extreme stock returns
	Microstructure (28)	Proxy for firm-level financial market frictions, including short sale restrictions, transaction costs, among others	Jarrow (1980): short sale restrictions; Mayshar (1981): transaction costs
	Behavioral (3)	Proxy for firm-level behavioral biases, including analyst dispersion, media coverage, among others	Diether, Malloy, and Scherbina (2002): analyst dispersion; Fang and Peress (2009): media coverage
	Accounting (87)	Proxy for firm-level accounting variables, including PE ratio, debt-to-equity ratio, among others	Basu (1977): PE ratio; Bhandari (1988): debt-to-equity ratio
	Other (24)	Proxy for firm-level variables that do not fall into the above categories, including political campaign contributions, ranking-related firm intangibles, among others	Cooper, Gulen, and Ovtchinnikov (2010): political campaign contributions; Edmans (2011): intangibles

The numbers in parentheses represent the number of factors identified. See Table 6 and http://faculty.fuqua.duke.edu/~charvey/Factor-List.xlsx.

ML allows to incorporate information from all factors jointly

Machine Learning in Asset Management. Why Now?

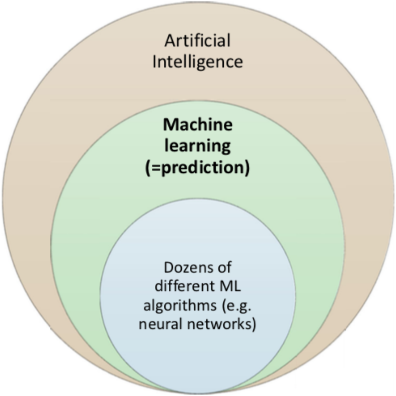
Requirements for good Predictions?



Algorithms (AI/ML)

Big (really big) picture

- Machine Learning: “prediction based on data”



Basic (simple) idea of a predictive model

- Predictive model:

$$Y = f(X)$$

- Y is the outcome we want to predict (e.g., stock returns)
- X (features/variables) are variables predicting Y
- f is the function (i.e., algorithm) linking X to Y

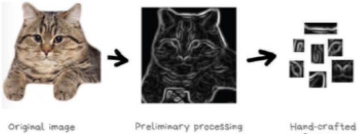
A huge number of real problems are prediction-based

Simple illustrative example

- Image recognition—10, 000 images of cats or dogs
- $Y = 0$ (CAT) or $Y = 1$ (DOG)

- $\mathbf{X} \rightarrow$ attributes of the picture (divide in different pixels)

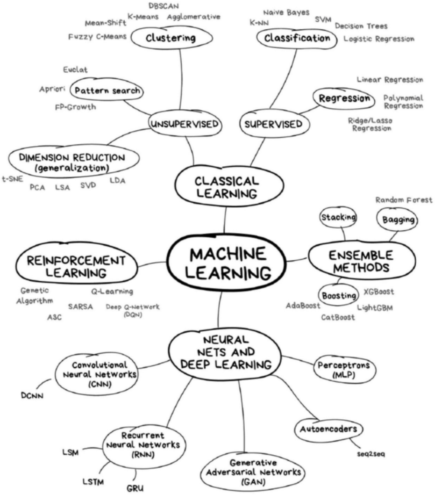


  - E.g., 1000 pixels (parts of a picture) values
  - $X = X_1, X_2, ..., X_{1000}$
  - $Y = f(X_1, X_2, ..., X_{1000})$

How do we determine the parameters of f?

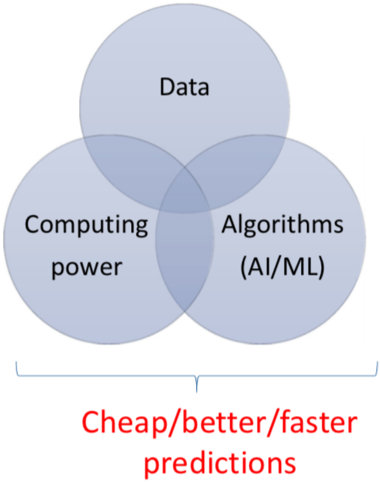
- Minimize prediction error (e.g., predicting dog when cat)

How many methods exist to determine f?

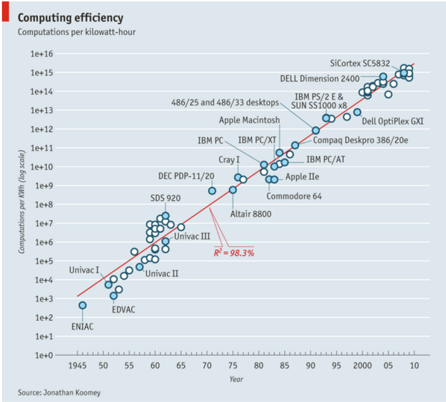


- Different tools depending on X and Y

Requirements for Good Predictions?

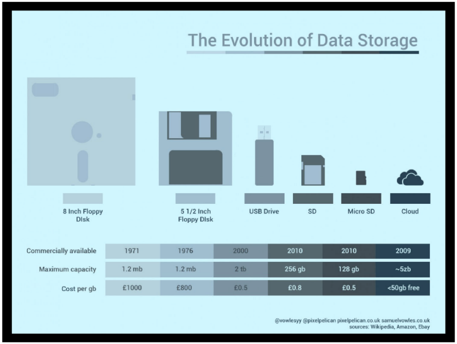


Computing power



- GPU more recently

Data Storage and Costs



Digitization of the economy

- Digitization creates data as a by-product
- Digitization lowers the cost of data gathering
  - More of existing type of data
  - New types of data (Example: Digital footprint)
- Huge amount of data useful for predictions

"Alternative" data

- Geolocation (foot traffic)
- Email receipts
- Point-of-sale transactions
- Website usage
- Satellite images
- Social media posts
- Online browsing activity
- ...
- Product reviews
- Flight and shipping trackers

$\rightarrow$ **Virtually all of them are used by asset managers**

Real-time Machine Learning in the Cross-Section of Stock Returns

Bin Li
Wuhan University

Alberto G Rossi
Georgetown University

Xuemin Yan
Lehigh University

Lingling Zheng
Renmin University

(A Cautionary tale of the Promises of Machine Learning in Asset Management)

Motivation

Key task in asset management

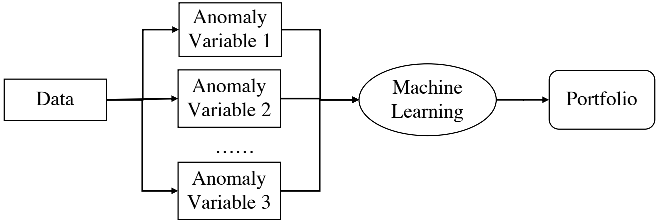
- Identify undervalued and overvalued assets
- Hundreds of signals have been discovered

Event	Market	Valuation	Fundamental
Change in Asset Turnover	52-Week High	Advertising/MV	Accruals
Change in Profit Margin	Age-Momentum	Analyst Value	Age
Change in Recommendation	Amihud's Measure	Book-to-Market	Asset Growth
Chg. Forecast + Accrual	Beta	Cash Flow/MV	Asset Turnover
Debt Issuance	Bid/Ask Spread	Dividends	Cash Flow Variance
Dividend Initiation	Coskewness	Earnings-to-Price	Earnings Consistency
Dividend Omission	Idiosyncratic Risk	Enterprise Component of B/P	Forecast Dispersion
Dividends	Industry Momentum	Enterprise Multiple	G Index

Big debate: if and when these signals lose predictive power

Motivation

- Recent years: ML methods have been introduced



- General conclusions:
  - *"ML provides large economic gains to investors."*
  - *"ML yields highly profitable investment strategies."*

Motivation

- One concern: these studies use the 200-300 signals discovered and published over the years.
- Strategies based on subsequently discovered anomaly variables cannot be implemented in real-time
- Decline in anomalies' post-publication performance (McLean and Pontiff, 2016)

Results reported may overestimate the promises of ML

This Paper

- We examine ML strategies based on a "universe" of over 18,000 fundamental signals
- Signals are constructed from financial statement variables using permutational arguments

1. Stem from investors' fundamental analysis
2. Implementable in real-time
3. Side-step the issue of data mining and look-ahead bias

Compare our results to the ones reported by other studies

Baseline Setting

- Sample: 15,035 stocks from July 1963 to June 2019
- 170,262 firm-year observations: $t - 1$'s fundamental signals + year t's annual excess return
- 18,113 signals.
- Computationally intensive. Some results required 30 days of computations on 1,056 cores.

Boosted Regression Trees (BRT)

$$R_{i,t+1} = f\left(\mathbf{x}_{i,t}|\theta\right) + \epsilon_{i,t+1}$$

In estimating $f(\cdot)$, we face several challenges:

1. Relation btw signals and returns may be nonlinear
2. Traditional methods may face "curse of dimensionality"

We use BRTs as a baseline method because they:

1. exhibit strong predictive performance across fields
2. can handle high-dimensional data sets b/c they perform
    - variable selection
    - shrinkage
3. Good interpretability

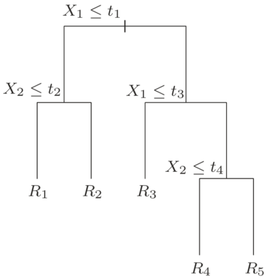
Regression trees

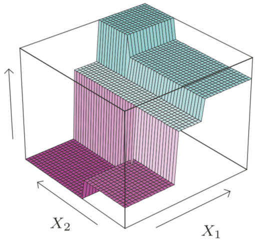
A regression tree, $\mathcal{T}_J$, with J regions (states) and parameters $\Theta_J = \{S_j, c_j\}_{j=1}^{J}$ can be written as:

$$\mathcal{T}(x, \Theta_J) = \sum_{j=1}^{J} c_j \; I \; (x \in S_j).$$

- $S_1, S_2, ..., S_J$: J disjoint states
- $x = (x_1, x_2, ..., x_P)$: P predictor ("state") variables
- Dependent variable is constant, c_j, within each state, S_j

Regression Trees: Intuition





Key features:

- Partitioning using lines parallel to the coordinate axes
- Recursive binary partitioning
- Very hierarchical
- Use less and less data $\rightarrow$ overfit

Boosting

A Boosted Tree Model is a sum of Regression Trees:

$$f_B(x) = \sum_{b=1}^{B} \mathcal{T}(x; \Theta_{J,b}).$$

The B-th boosting iteration fits a tree on:

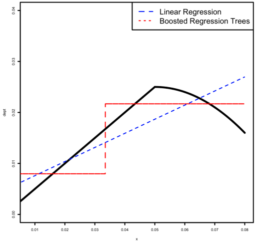
$$\hat{\Theta}_{J,B} = \arg\min_{\Theta_{J,B}} \sum_{t=0}^{T-1} [e_{t+1,B-1} - \mathcal{T}(x_t; \Theta_{J,B})]^2$$

where $e_{t+1,B-1} = y_{t+1} - f_{B-1}(x_t)$ are the residuals of the model with "B-1" iterations.

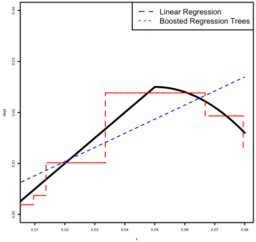
To minimize the current residuals, the B-th tree finds:

- The optimal splitting regions, $S_{j,B}$
- The optimal constants, $c_{j,B}$

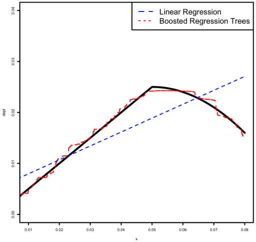
BRT vs linear models



(a) 1 iteration



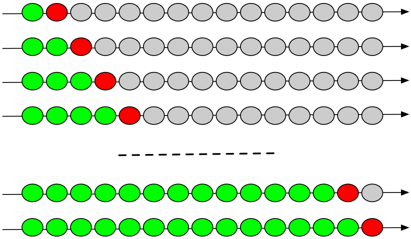
(b) 5 iterations



(c) 1,000 iterations

Model Selection

- Baseline results: Expanding window
- Training sample: 1963 - 1986; validation: 12 years
- Test sample: 1987 - 2018
- Model selection: grid search across hyperparameters



Performance Evaluation

1. BRTs Predict stock returns
2. Also use Neural Networks (NNs)
3. Form decile portfolios based on expected returns (equally-weighted)
4. Construct High-Low (H-L) long-short portfolios

Results with Previously Discovered Signals

- 94 predictors Green et al. (2017) **(GHZ)**
- 207 predictors from Chen and Zimmermann (2022) **(CZ)**

	(GHZ)			**(CZ)**		
Method	*H-L*	*t-stat*	*Sharpe*	*H-L*	*t-stat*	*Sharpe*
BRT	42.84%	8.91	2.35	62.16%	9.91	3.68
NN1	45.36%	9.74	2.69	44.64%	7.63	2.82
NN2	46.44%	9.93	2.95	39.12%	6.32	2.23
NN3	47.28%	10.09	2.84	38.16%	6.22	2.32
NN4	35.28%	9.59	2.27	36.00%	6.07	2.32
NN5	20.40%	5.92	1.33	19.56%	5.14	1.23

Results with Universe (18,113) of Signals

Method	*H-L*	*t-stat*	*Sharpe*
BRT	11.40%	3.26	1.02
NN1	12.96%	6.09	1.16
NN2	12.36%	4.10	0.75
NN3	14.04%	5.32	1.10
NN4	11.88%	5.53	0.98
NN5	9.60%	3.79	0.74

Main Takeaways

1. Many of the outstanding results associated with ML do not account for the fact they use signals discovered over time

2. Feature engineering, i.e., the choice of signals fed to ML methods, is crucial to their performance.

3. ML methods are not a panacea in asset management because financial markets evolve over time

Conclusions

- ML and AI tools are becoming very widespread
- Asset Management is no exception
- Should assess their effectiveness carefully
    - across different contexts
    - over time

before deploying them