Evolutionary Machine Learning for Financial Markets

Sorin Alexe
QML Alpha LLC, Stamford, CT, USA

www.qmlalpha.com
Abstract

Machine Learning (ML) and Artificial Intelligence (AI) become increasingly important to many areas of Science and Technology. The increased interest for the use of ML to develop financial quantitative strategies raises new challenges, as data are time series, dynamic, stochastic and noisy. The proposed approach is to generate a large number of signals and to apply consensus methods to obtain models with increased performance. This requires advanced model generation techniques and high level of automation. We present a generative algorithm that operates over an algebra of portfolios and applies mathematical and ML operations to generate “expressions”. Grammar-like derivations are used to generate hypotheses that are backtested, filtered and analyzed.
Outlines

• History of Investment
• Portfolio Construction
• Finding Alphas
• Evolutionary Framework
• QML Platform Demo Session
• Conclusions
History of Investment - Bonds

• The First Bond Ever. (~2400 BC)
The first recorded bond in history dates back to 2400 B.C. – a stone discovered at Nippur, in Mesopotamia (Iraq).

• Venetians Create Advanced Bond Markets (1100 AD)
Venice began issuing government bonds to fund its wars, known as the presiti. The city continued to evolve its bond market throughout the 14th century, when denizens of Venice could purchase and trade government securities, which paid the owner an endless annuity at a set rate.

• The First Official Government Bond
The first ever government bond was issued by the Bank of England in 1693 to raise money to fund a war against France. These first bonds were a mix of both lottery and annuity.
History of Investment - Stocks

- In the early modern period, the Dutch developed several financial instruments and helped lay the foundations of modern financial system. The Dutch East India Company (VOC) became the first company in history to issue bonds and shares of stock to the general public. In other words, the VOC was officially the first publicly traded company, because it was the first company to be ever actually listed on an official stock exchange.

- The Dutch East India Company (Dutch: Vereenigde Oostindische Compagnie; VOC) was an early megacorporation founded by a government-directed amalgamation of several rival Dutch trading companies (voorcompagnieën) in the early 17th century. It was established on March 20, 1602, as a chartered company to trade with India and Indianised Southeast Asian countries when the Dutch government granted it a 21-year monopoly on the Dutch spice trade.
History of Investment - Futures

• In 1697, the rice exchange received an official license from the Shogunate (government). By 1710, merchants were trading futures contracts based on the perceived future value of rice. 1710 is the official date at which the modern futures exchanges market is thought to have begun.

• In finance, technical analysis is an analysis methodology for forecasting the direction of prices through the study of past market data, primarily price and volume.

• Speculation opportunities appeared:
  Buy a $100 oil future contract, sell a $110 oil future contract.
Portfolio Construction

A Portfolio is a basket of securities. Example: 1 million shares IBM and 2 million shares Apple: \( W = (1 \times $132.22, 2 \times $190.15) \)

\[ \text{https://finance.yahoo.com/quote/IBM?p=IBM} \]
\[ \text{https://finance.yahoo.com/quote/AAPL/?p=AAPL} \]

Performance Metrics

- **Return:** \( R_P = W^T R \)
- **Risk (volatility):** \( \text{Risk}_P = \sqrt{W^T Q W} \)

Q – covariance matrix

**Goal** (equivalent formulations)

1. maximize portfolio return for a certain level of risk
2. maximize risk adjusted return
3. minimize risk for a certain target return
4. maximize Sharpe Ratio \( (R_P / \text{Risk}_P) \)

**Example**

<table>
<thead>
<tr>
<th>Return</th>
<th>Volatility</th>
<th>SR</th>
</tr>
</thead>
<tbody>
<tr>
<td>2%</td>
<td>0%</td>
<td>infinity</td>
</tr>
<tr>
<td>3%</td>
<td>0.7%</td>
<td>4.29</td>
</tr>
<tr>
<td>4%</td>
<td>2%</td>
<td>2.00</td>
</tr>
<tr>
<td>7%</td>
<td>8%</td>
<td>0.88</td>
</tr>
<tr>
<td>11%</td>
<td>19%</td>
<td>0.58</td>
</tr>
<tr>
<td>14%</td>
<td>30%</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Markowitz Models (1952, later Nobel prize for Economics)
Estimate R and Q - Two Big Challenges

**Estimations for R = Alpha Models** (signals, predictors, factors):
- Economic factors: Value and Growth
- Behavioral patterns: Momentum and Reversion
- Economic cycles: Bulls and Bears

**Estimations for Q** (factor risk models: style factors, statistical factors, sectors, industries):

\[
Q = B^T \Sigma B + \Delta^2
\]

Example: 2000 assets, 20 factors:
Q(2000,2000), B(20,2000), \( \Sigma(20,20), \Delta^2 = \text{diag}(2000) \), i.e., use 41210 numbers to represent 2001000 (Q and \( \Sigma \) are symmetric)

Note. If W is orthogonal to all columns of B then \( W^T Q W = W^T B^T \Sigma B W + W^T \Delta^2 W = W^T \Delta^2 W \) (specific risk)

**Advantage:** Q is easy to invert via **Woodbury matrix identity:**

\[
(A + U C V)^{-1} = A^{-1} - A^{-1} U (C^{-1} + V A^{-1} U)^{-1} V A^{-1}
\]

So, an \( O(n^3) \) computational complexity becomes \( O(n^2) \).
Finding Alphas

Examples:

1. Fama and French Three Factor Model (CAPM): adding size risk and value risk factors to the market risk factor
   https://www.investopedia.com/terms/f/famaandfrenchthreelfactormodel.asp

   \[ R_{it} - R_{ft} = \alpha_{it} + \beta_1 (R_{Mt} - R_{ft}) + \beta_2 SMB_t + \beta_3 HML_t + \varepsilon_{it} \]
   - \( R_{it} \) is the total return of a stock or portfolio, \( i \) at time \( t \)
   - \( R_{ft} \) is the risk free rate of return at time \( t \)
   - \( R_{Mt} \) is the total market portfolio return at time \( t \)
   - \( R_{it} - R_{ft} \) is expected excess return
   - \( R_{Mt} - R_{ft} \) is the excess return on the market portfolio (index)
   - \( SMB_t \) is the size premium (small minus big)
   - \( HML_t \) is the value premium (high minus low)
   - \( \beta_{1,2,3} \) refer to the factor coefficients

2. Reversion Model: negative 5-day returns, factor and group neutralized
   SR: 0.3 \( \rightarrow \) 1.54

3. Long Term (252 day) – Short Term (20) Momentum spread:
   SR: 0.4 \( \rightarrow \) 1.54
For a long time linear regression was the standard for the construction of predictive models in finance. It used to be present in 90% of the models involved in quantitative investment and trading.

Lately, the Machine Learning approach gains significant visibility. About 30% of the 10,000 hedge funds in the USA involve Machine Learning and Artificial Intelligence algorithms to design their trading strategies.

Igor’s design (WorldQuant) claims to achieved 26 million alphas, aiming for 100M. This book is interesting as it does not disclose the techniques of finding alphas, but rather described the framework that should be deployed for a successful implementation. As usual, details remain to be defined.

However, linear regression is just one of the methods that could detect signals and information in data.

Machine Learning seems to be a better choice, and the financial industry demonstrated an increased appeal in this domain. However, applying ML blindly can be harmful. No algorithm (ready from the shelf) is ready to face the complexity of high noise, uncertain dynamics and low resolution of the market data. The combination of feature extraction and inductive learning, adapted to a time series format has been proven to be a feasible approach.

We propose the use of a kernel of learning techniques embedded in a automation framework. The evolutionary approach is an additional layer on top of Machine Learning traditional or novel algorithms.
Evolutionary Framework

- Large (infinite) number of states
- A system that can represent on state at a time
- A set of operations (transition matrix) that can change the state of the system sequentially
  - Deterministic
  - Stochastic
- A semantics that validates states
- Examples
  - Digital computers: 0/1 (meaning)
  - Biology: Genetic mutations (survivorship)
  - Finance: Collection of all portfolios over a common universe (investment grade)

Example: epic John Horton Conway's “game of life”
Evolution in Biology & AI
The Machine Arrival

• Artificial Intelligence is developing at increased speed and starts to have major impact for society, economy and finance.
• 1997: IBM Deep Blue computer defeated Kasparov in the second chess match
• 2011: IBM Watson computer won ‘Jeopardy!’
• 2015: Google’s AlphaGo AI defeated a human 9-dan GO champion
Operations and Transformations Kernel

• Mathematical
• Technical Analysis
• Control (risk, holdings and trading)
• ML supervised:
  • Regression (assets and portfolio level predictions)
  • Classification (marginal classifiers could become signals)
  • Aggregation techniques (consensus for signal strength augmentation and noise cancellation)

• ML unsupervised
  • Feature Extraction (risk models)
  • Clustering (groups)
Next Layer: Automation

- Natural Computing Approach
- Infinite Space
- Sample large number of hypothesis
- Backtest and filter
- Assemble the findings
Algebra of Portfolios

• A portfolio is a 2 dimensional set of weights:
  • Time (daily)
  • Assets (tickers)

• Performance is based on the economic effect:
  • Profit & Loss
  • Risk exposure
    • Sharpe, Sortino, Sterling and Calmar ratios

• Operations can be applied to a set of portfolios to create new ones:
  • Averaging (mathematical transformations)
  • Learning (ML algorithms and researcher interactions)
  • Hedging (control)
  • Evolutionary
Evolutionary Alpha Discovery

The space of all searchable portfolios

A large number of portfolios is sampled from the search space

High performing portfolios are retained

Retained portfolios are combined into one model
Sector Learning

• Start with a portfolio, say Stochastic indicator for 10 days
• Project it on sectors and subindustry neutralize all components
• Apply Mean-Variance optimization

All statistics on this page are illustrative and not based on any actual statistical analyses
Pattern Based Inductive Learning

• Select 2 explanatory attributes (return and volume)
• Define positive return (in blue) and negative return (red) classes
• Generate positive and negative patterns (rules)
• Construct positive and negative indices
• Build the alpha model as the difference between the positive index and the negative index
Pattern Based Portfolio

Input: Short Sentiment data (9 attributes, factor neutralized)

Parameters:
• Grid size = 8
• Positive patterns ratio > 51%
• Negative pattern < 49%
• Learning period = 6 years (2 mil data points)

Findings:
• 232 positive patterns
• 24 negative patterns

Portfolio generated: SR = 1.4, PnL = 7%

All statistics on this page are illustrative and not based on any actual statistical analyses
Feature Extraction
Hedging with Risk Models

<table>
<thead>
<tr>
<th>Alpha Name</th>
<th>Sharpe Ratio</th>
<th>PnL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stochastic</td>
<td>0.93</td>
<td>9.10%</td>
</tr>
<tr>
<td>Stochastic_Traditional</td>
<td>1.87</td>
<td>6.19%</td>
</tr>
<tr>
<td>Stochastic_PCA</td>
<td>2.28</td>
<td>6.69%</td>
</tr>
</tbody>
</table>

All statistics on this page are illustrative and not based on any actual statistical analyses.
Evolutionary Alpha Discovery

- Use attributes, factors and groups as input variables
- Combine the variables using operations:
  - Technical indicators
  - Hedging operations
  - Aggregation Methods
  - Machine Learning Algorithms

Example: Alpha_01=Expression(Rank(MACD(ret,10,20))*Stochastic(close,5))

Abstractize to a Grammar

[port]=Stochastic([Price],[days])|MACD([Return],[days],[hist])
[Price]=close|high|low|open
[Return]=ret
[Alpha]=[port]*[port]
[days]=5|10
[hist]=20|60|60
[StartSymbol]=[Alpha]

All statistics on this page are illustrative and not based on any actual statistical analyses
Evolutionary Alpha Discovery

- Generate many “portfolio expressions” of the same kind
- Backtest, select and aggregate/learn

MACD(ret,10,60)*Stochastic(low,5)
Stochastic(close,10)*MACD(ret,10,20)
Stochastic(high,5)*Stochastic(open,10)
Stochastic(close,5)*Stochastic(low,10)
Stochastic(close,10)*MACD(ret,5,20)
Stochastic(close,5)*Stochastic(low,5)
MACD(ret,10,60)*Stochastic(high,5)
Stochastic(high,10)*MACD(ret,10,20)
MACD(ret,5,60)*MACD(ret,10,20)
MACD(ret,5,60)*MACD(ret,5,60)

Pattern based alpha model

All statistics on this page are illustrative and not based on any actual statistical analyses
QML Platform Demo Session

• Interaction
  https://www.youtube.com/edit?o=U&ar=3&video_id=8eBm2R8BbdU

• Machine Learning
  https://www.youtube.com/edit?o=U&ar=3&video_id=FN-cUhHfLq8

• Evolution
  https://www.youtube.com/edit?o=U&ar=3&video_id=c3CVmEomBc
Conclusions

• Machine Learning is going to transform the society in the near future.
• Financial industry will be affected too: faster and better investment decisions, order flow management and execution
• This raises the question: Is information going to be commoditized? To the point where the value is precisely known? Or is it going to be still an advantage to those on the upfront wave of research?