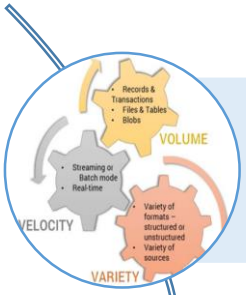
Big Data and AI Strategies

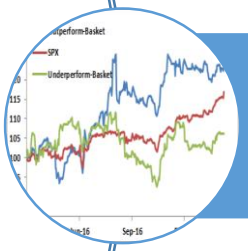
Machine Learning and Alternative Data Approach to Investing

Rajesh T Krishnamachari, Ph.D.



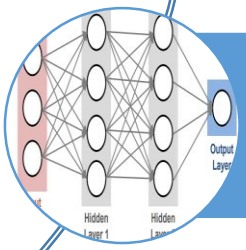
Part One:

Overview of Big Data and Machine Learning



Part Two:

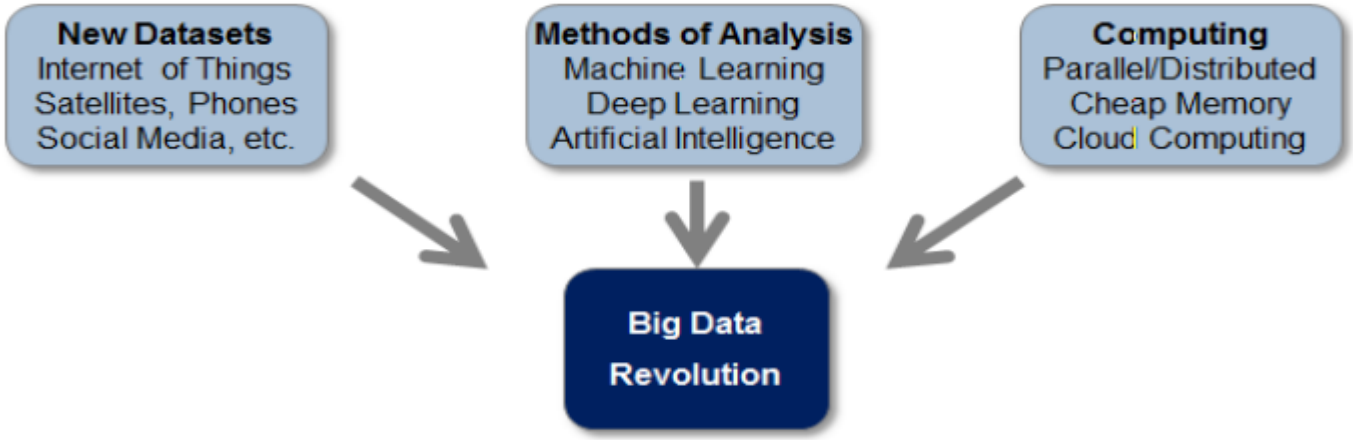
Alternative Data

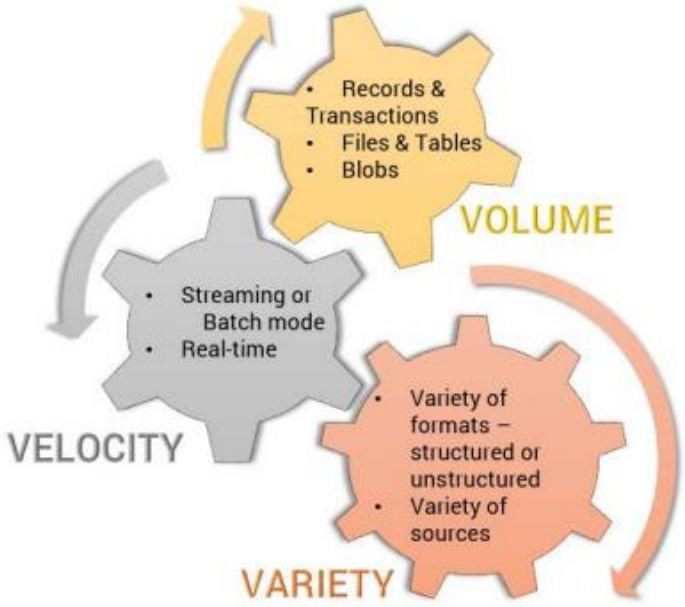


Part Three:

Machine Learning and A.I.

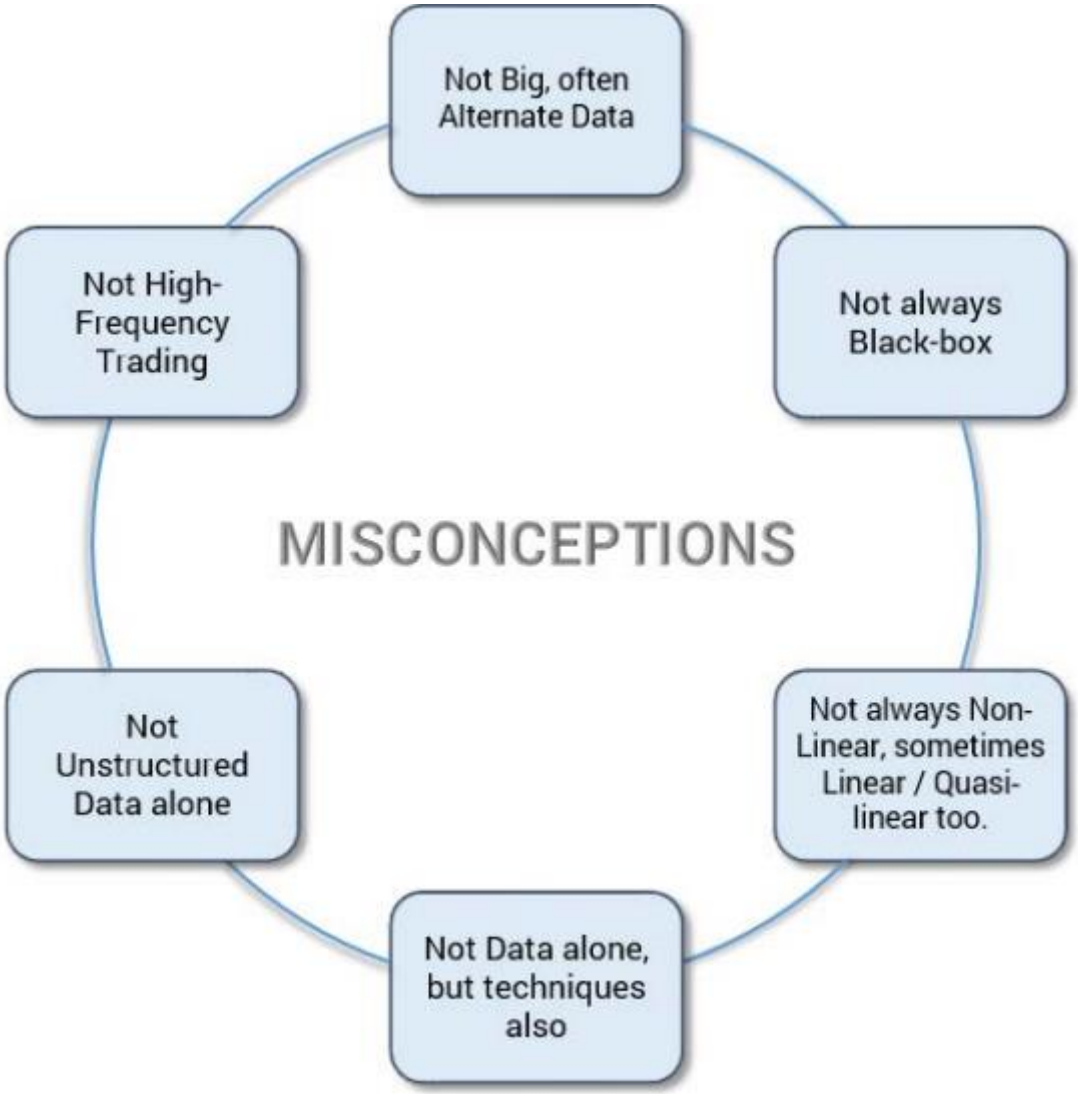
Factors Leading to Big Data Revolution



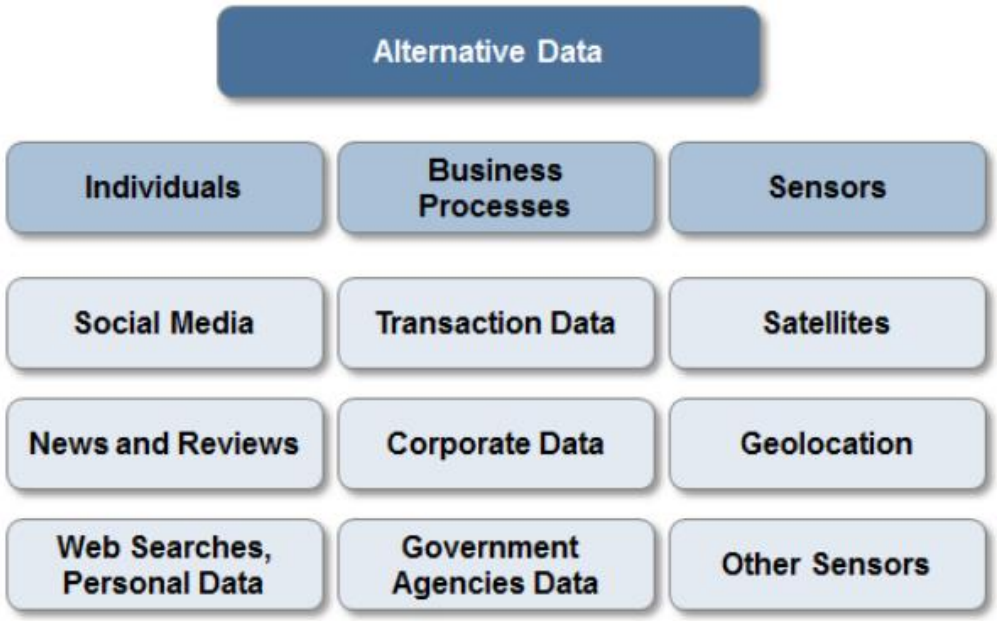


Link: Google "Big Data and AI Strategies PDF"

Common Misconceptions around Big Data in Investing/Trading



Classification of Alternative Data Sets



A. Data from individual activity
   1. Social media
      i. Investment professional social media
      ii. Social network sentiment
      iii. Blogs, picture and video analytics
   2. News and reviews
      i. Mobile data content and reviews
      ii. News sentiment
   3. Web searches and personal data
      i. Email and purchase receipt
      ii. Web search trends

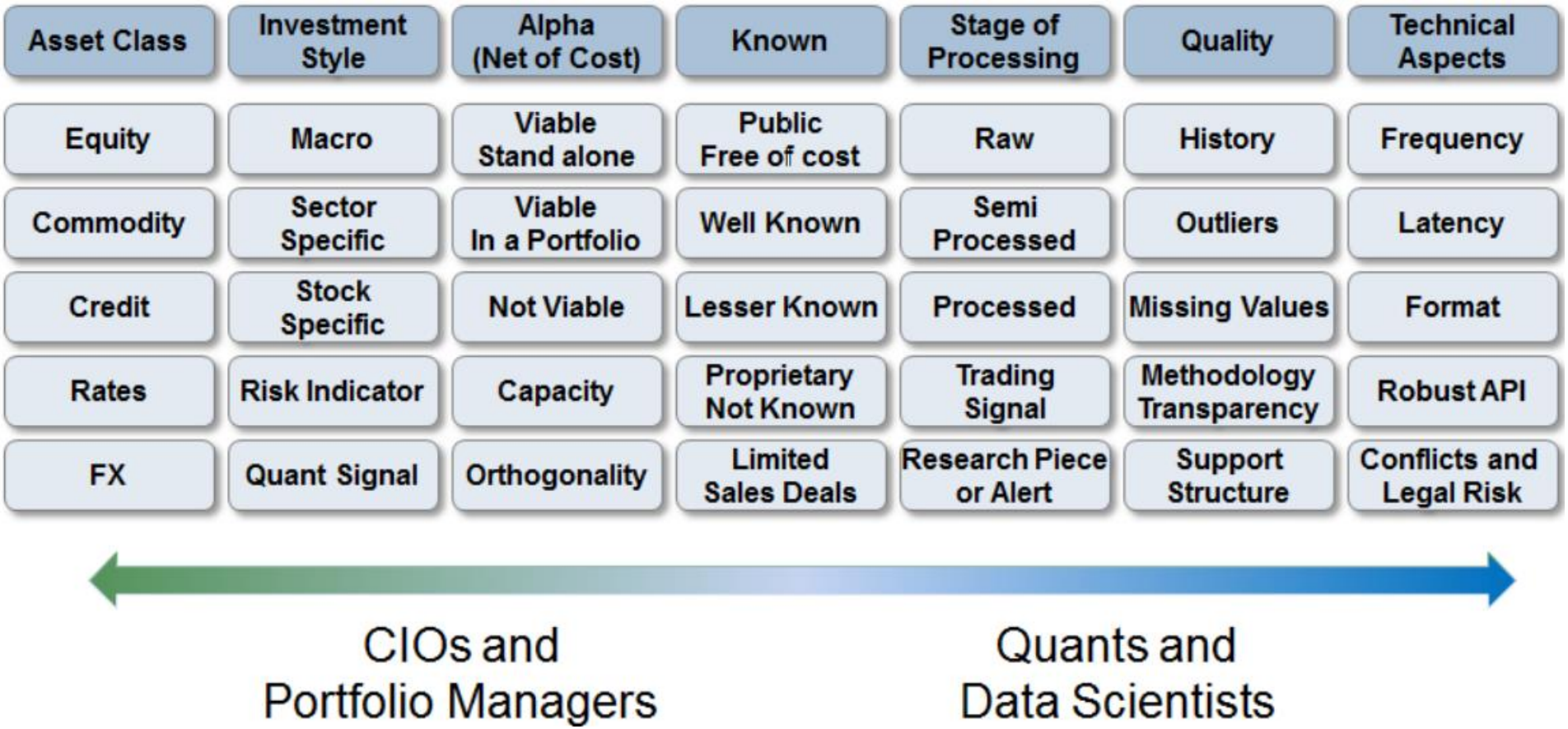
B. Data from business processes
   1. Transaction data
      i. Other commercial transactions
      ii. E-commerce and online transactions
      iii. Credit card data
      iv. Orderbook and flow data
      v. Alternative credit
   2. Corporate data
      i. Sector data (C.Discretionary, Staples, Energy/Utilities, Financials, Health Care, Industrials, Technology, Materials, Real Estate)
      ii. Text parsing
      iii. Macroeconomic data
      iv. Accounting data
      v. China/Japan data
   3. Government agencies data
      i. Federal or regional data
      ii. GSE data

C. Data from sensors
   1. Satellites
      i. Satellite imagery for agriculture
      ii. Satellite imagery for maritime
      iii. Satellite imagery for metals and mining
      iv. Satellite imagery for company parking
      v. Satellite imagery for energy
   2. Geolocation
   3. Other sensors

D. Data aggregators

E. Technology solutions
   1. Data storage (databases)
   2. Data transformation
   3. Hadoop and Spark framework
   4. Data analysis infrastructure
   5. Data management and security
   6. Machine learning tools
   7. Technology consulting firms

Attributes of Alternative Data



Information Content of an Alternative Data Set

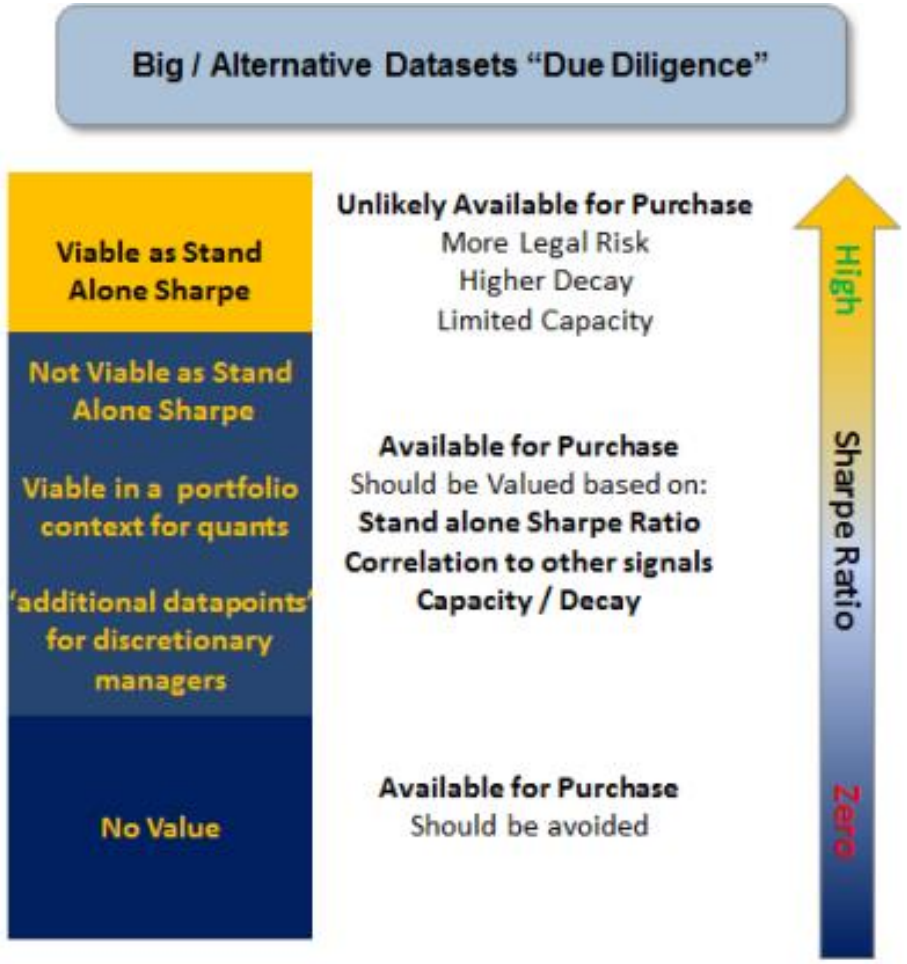
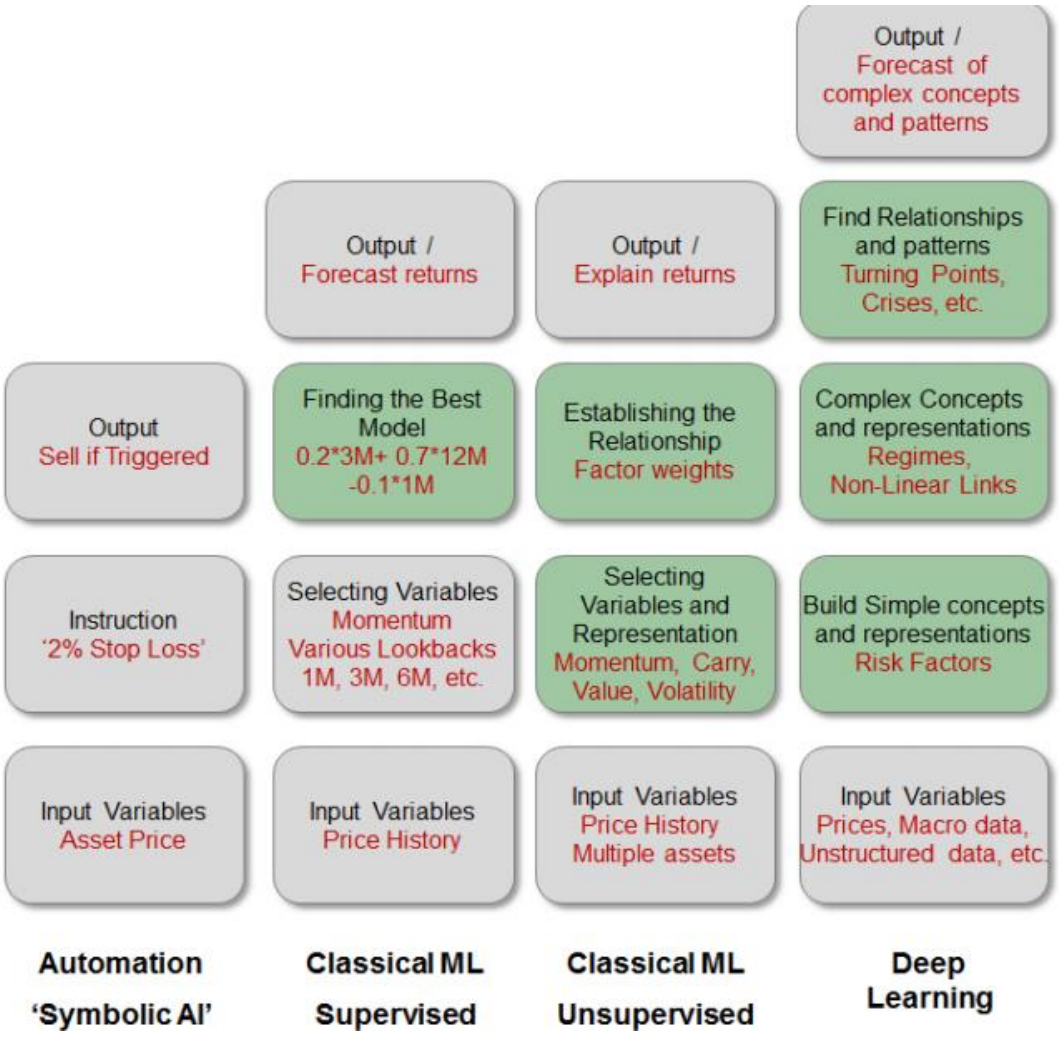
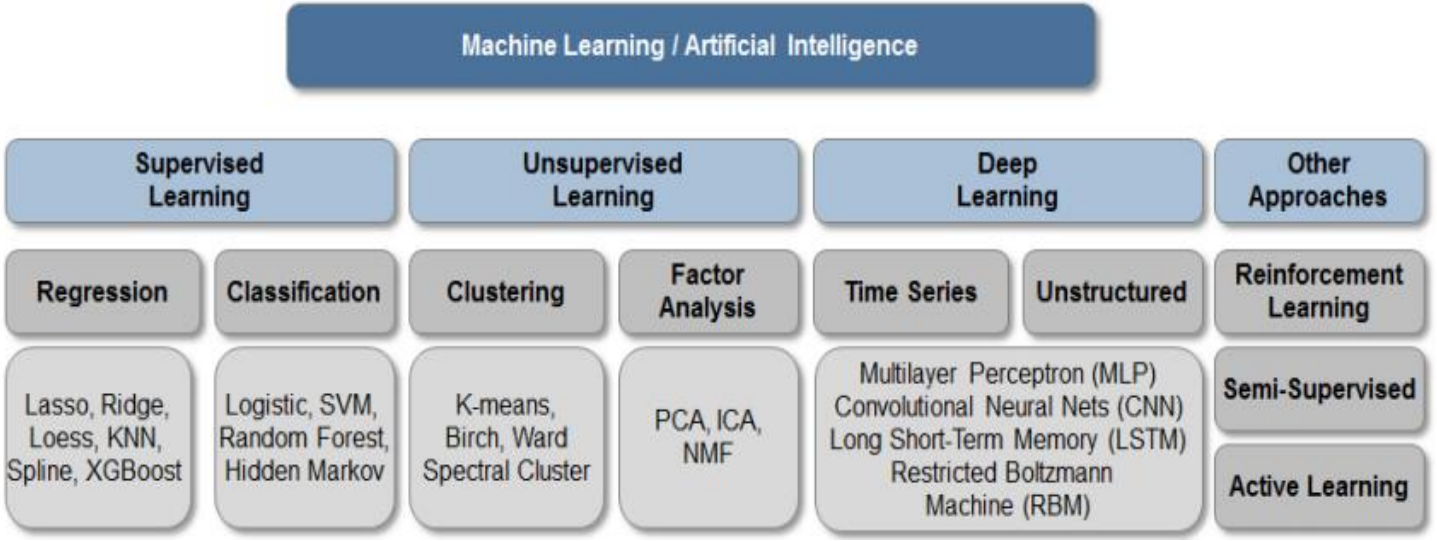


Illustration of Machine Learning Categories

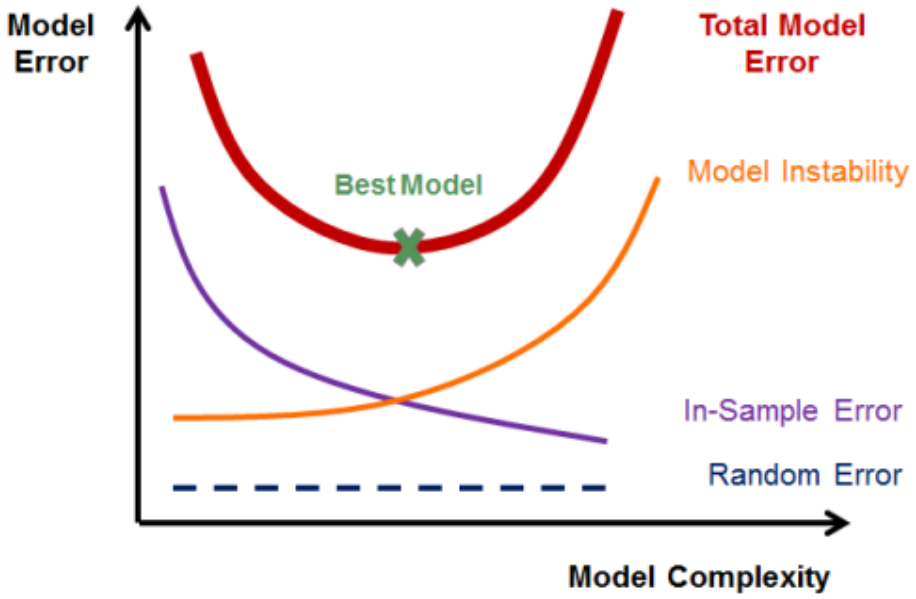


Machine Learning Techniques

Taxonomy of Machine Learning Algorithms



Trade-off between 'model bias' and 'model variance'

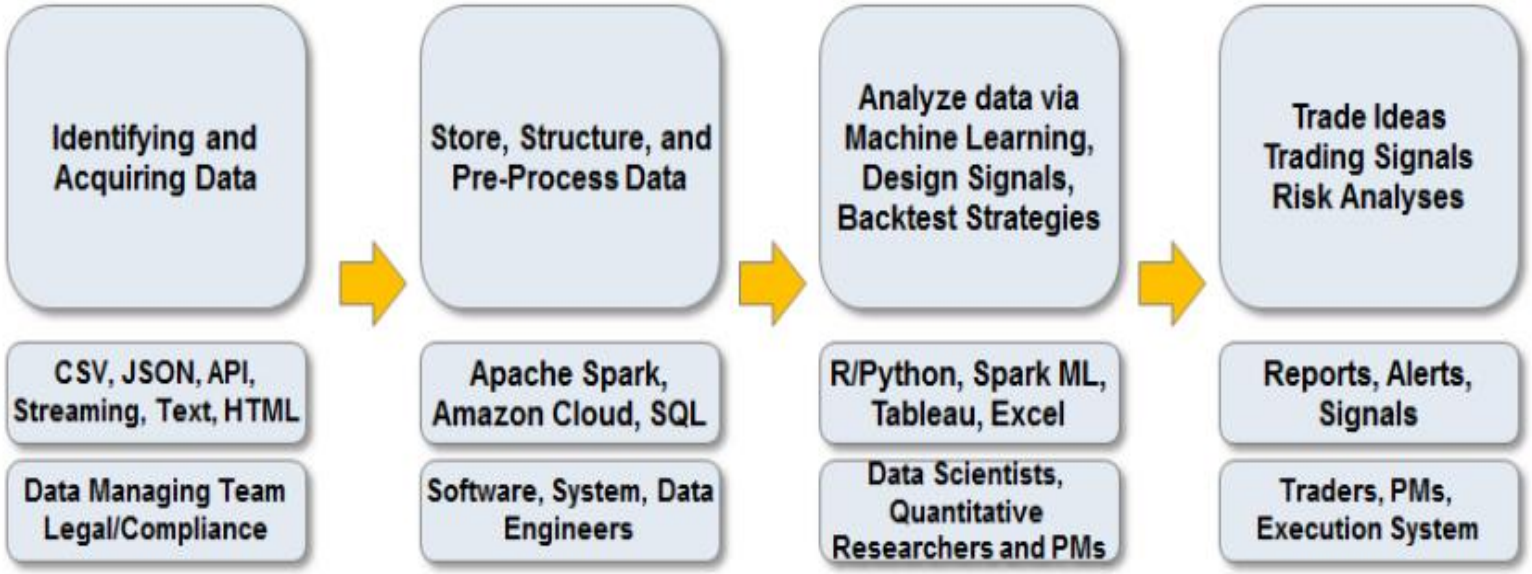


Techniques to find the right model:

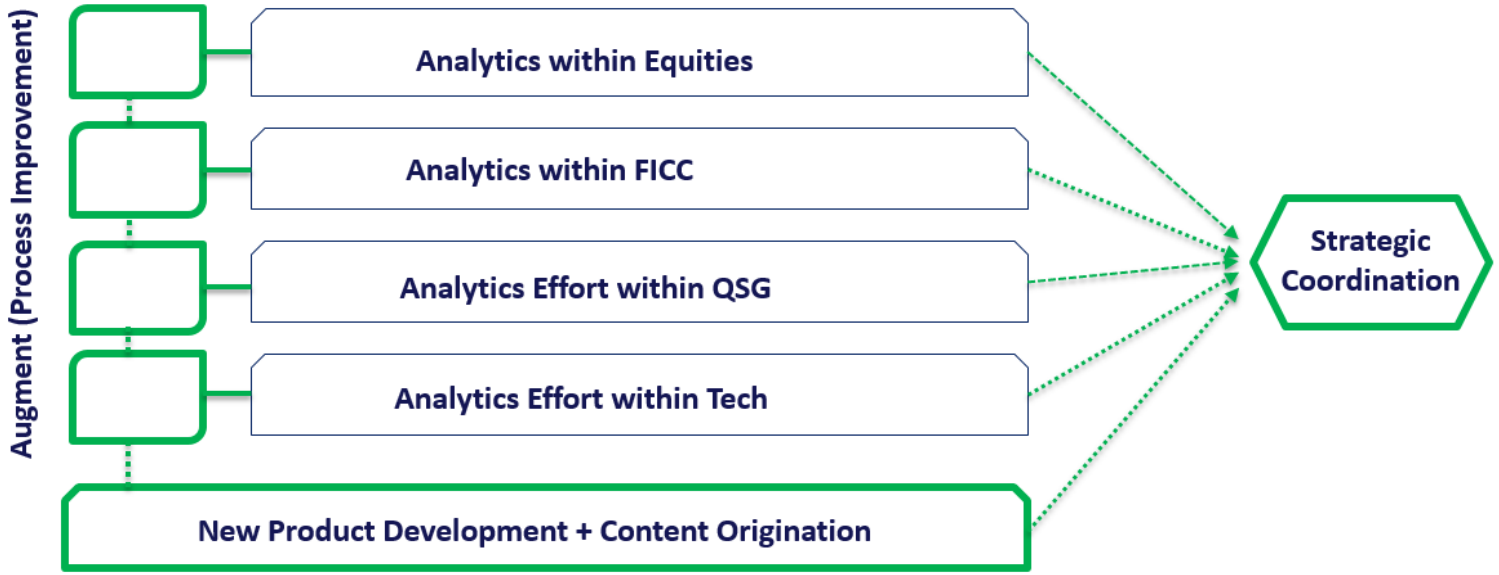
- Analytical Approach: AIC, BIC
- Efficient Sample Re-use: cross-validation, bootstrap
- Regularization: penalty function

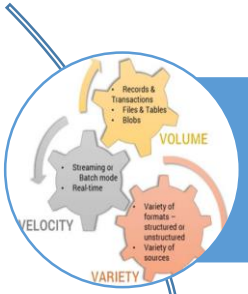
Big Data Workflow

Data to Trade Ideas

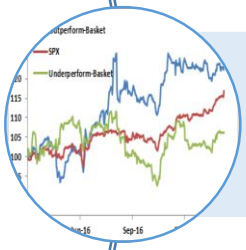


BAML Workflow

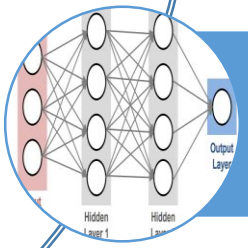




Part One:
Overview of Big Data and Machine Learning

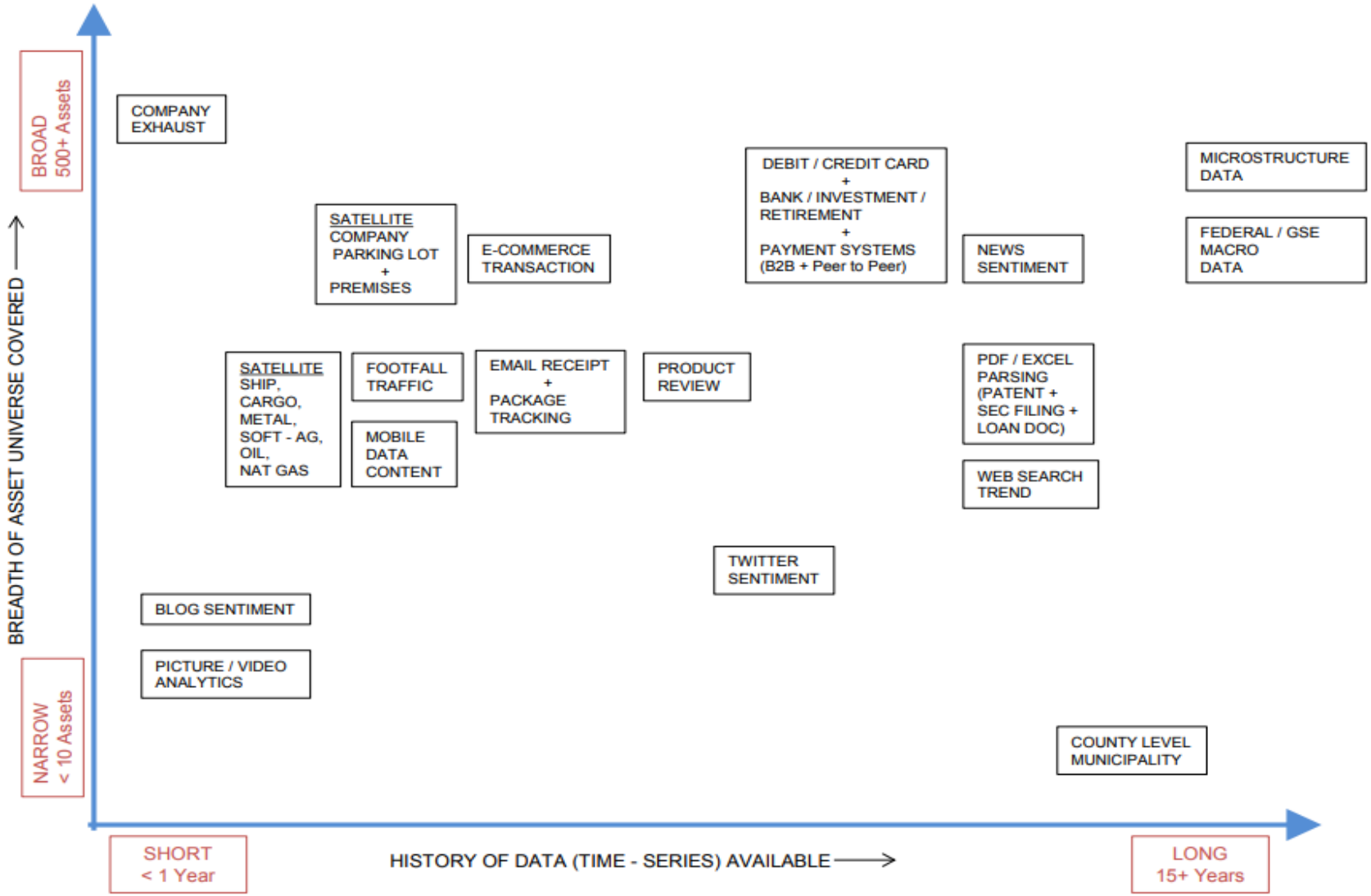


Part Two:
Alternative Data



Part Three:
Machine Learning and A.I.

Typical Length of History for Alternative Data Sets



Alternative Data From Individual Activity

Sentiment Analysis

- Named entity extraction
- Theme and category extraction
- Intention and sentiment
- Relevance and influence

Limitations

- Winograd's schema and linguistic idiosyncrasies

Cross-asset

- Ravenpack
- Gnip (Twitter)
- Descartes Labs (Commodity)
- Social Alpha (ETF)

Client focus

- Data Minr (Fundamental)
- LexAnalytics (raw NLP engine for quant)
- DataSift (uses LexAnalytics)

Others

- GDELT (news)
- Inferess (Asia)
- Repustate (multi-language)
- Heckyl (for VC/PE firms)

Specialized websites

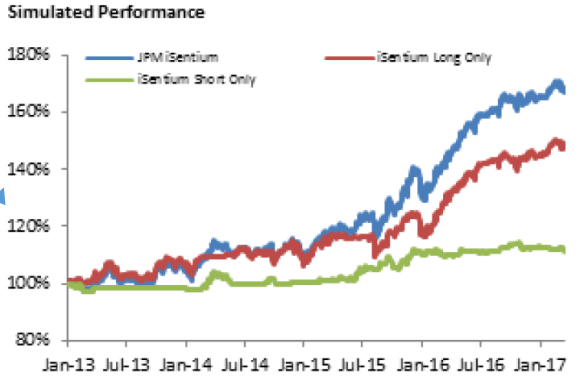
- App Annie/AppTopia (for mobile apps)
- Yipit (website)
- Google Trends (search)
- Return Path (email data)

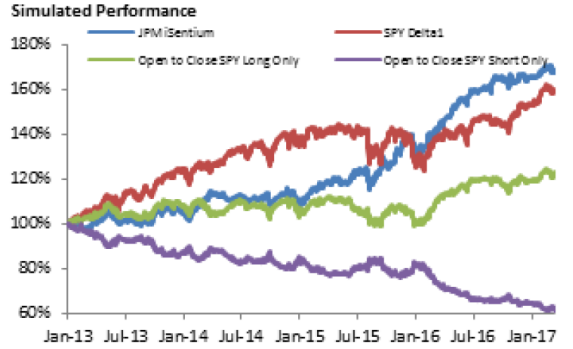
Alternative Data From Individual Activity

Case Study: Using Twitter to trade S&P 500

Construction of iSentium Daily Directional Indicator

1. The universe is limited to the 100 stocks which are most representative of the S&P 500, filtered using tweet volume and realized volatility measures.
2. Tweets are assigned a sentiment score using a patented NLP algorithm.
3. By aggregating tweet scores, a sentiment level is produced per minute between 8:30 AM and 4:30 PM every day. Sentiment for the day is aggregated using an exponentially weighted moving average over the past ten days.
4. S&P 500 returns are forecasted using a linear regression over the sentiment scores for the past two days, with betas evolved via a Kalman filter.



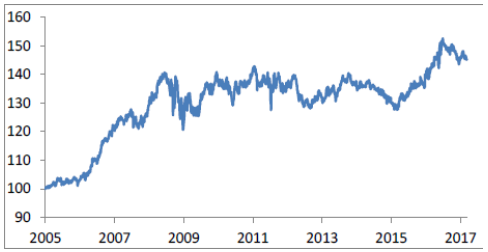


Strategy/Index	Return Ann. (%)	Volatility (%)	IR	Max DD (%)
iSentium L/S (JPM iSentium Index)	13.74	9.79	1.40	-8.10
iSentium – Act only on Long signal	10.33	8.74	1.18	-7.29
iSentium – Act only on Short signal	2.83	4.46	0.63	-4.66
S&P 500 Index	12.08	12.76	0.95	-12.08
S&P 500 – Long from open to close only	5.25	9.83	0.95	-5.25
S&P 500 – Short from open to close only	-10.97	9.83	-1.12	-10.97

Alternative Data From Individual Activity

Case Study: Using News Sentiment to Trade Eq/Bonds/FX/Comdty (Ravenpack)

Daily performance of long (top 3)/short (bottom 3): Bond, Equity, FX



Performance of signals with different lookback windows: Bond, Equity, FX

Time	Returns	Sharpe
1D	-0.26	-0.10
1Wk	0.03	0.01
1M	1.18	0.45
2M	0.29	0.11
3M	0.12	0.05

Time	Returns	Sharpe
1D	2.76	0.21
1Wk	5.62	0.44
1M	1.95	0.14
2M	2.27	0.16
3M	0.78	0.06

Time	Returns	Sharpe
1D	0.48	0.07
1Wk	3.24	0.43
1M	2.53	0.32
2M	1.32	0.16
3M	-0.22	-0.03

Correlation of sentiment strategy with traditional Risk Premia: Bond, Equity FX

Risk Premia		1	2	3	4	5
Volatility - Bond	1					
Value - Bond	2	-0.04				
MoM - Bond	3	0.00	0.63			
Carry - Bond	4	-0.04	0.49	0.44		
Sentiment - Bond	5	-0.03	0.04	0.03	0.10	

Risk Premia		1	2	3	4	5
Volatility - Equity	1					
Value - Equity	2	0.16				
MoM - Equity	3	0.02	0.14			
Carry - Equity	4	0.16	0.03	-0.16		
Sentiment - Equity	5	0.04	0.09	0.08	-0.04	

Risk Premia		1	2	3	4	5
Volatility - FX	1					
Value - FX	2	-0.02				
MoM - FX	3	-0.06	-0.06			
Carry - FX	4	0.22	0.08	0.01		
Sentiment - FX	5	-0.03	-0.03	-0.02	-0.02	

Translating RavenPack News Feed into Daily Sentiment Score

RavenPack provides 50 data fields for each event. We analyzed data since 2005 for each asset of interest.

1. Step One: isolate all unique events on a given day specific to a certain "ENTITY_NAME". Entity name was set equal to the currency, commodity or country name. We set a cutoff time of 4 PM EST to reflect NY market close.
2. Step Two: RavenPack provides a field called "RELEVANCE", which is an integer score between 0 and 100. A higher value indicates that the mention of the entity is more integral to the underlying news story. We used RELEVANCE as a cut-off filter, ignoring stories with a value < 75.
3. Step Three: RavenPack provides a field called "EVENT_SENTIMENT_SCORE" or ESS. It is a granular score between -1.00 and +1.00 that represents the news sentiment for a given entity. The average of ESS for all filtered events was designated as the sentiment value for the day. We forward-filled the sentiment for days on which no news was received by the analytics engine.

Alternative Data From Business Activity

Public Agencies

- International: IMF, WorldBank, WTO
- Federal: Fed Reserve, China
- City: San Francisco, New York

Commercial Transactions

- Point of sale: Nielsen
- Intercompany payments: D&B
- Consumer transaction: Yodlee, Second Measure, Earnest
- Newer Transaction: Bill of lading (EagleAlpha), Online retail (Slice)
- Building Permit: BuildFax

Other Private Agencies

- Traditional Private Agencies: Edmunds (Car), Redbook (SSS trends), SNL Financial (Cable/Broadcast)
- Market microstructure: Tick Data

Challenges in alternative data from business activity

- Sampling bias: demography, geography, income
- Short history; Error prone (seasonality)
- Alternate flows (DDA accounts) exist
- New ideas: Predict inflection point in BLS non-farm payroll

Alternative Data From Business Activity

Case study: Emailed Receipts to trade US Equities

80% of all online purchases, >5000 retailers

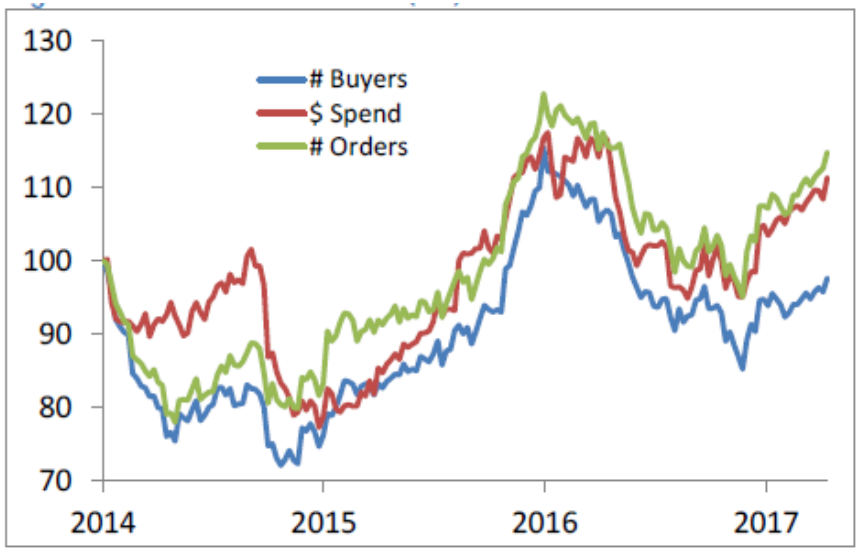
Sharpe ratios using long/short strategies for dollar spend, buyer count and order count data

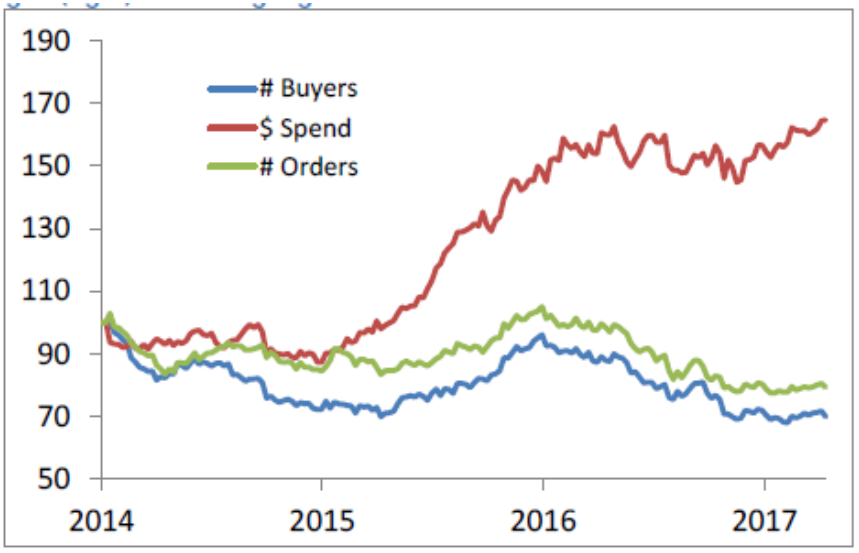
Dollar Spend Data	Top 6/ Bottom 6
Level	0.29
Z-score 4 week	1.13
Z-score 5 week	0.72
Z-score 6 week	0.67

Buyer Count Data	Top 6/ Bottom 6
Level	0.02
Z-score 4 week	-0.71
Z-score 5 week	-0.49
Z-score 6 week	0.04

Order Count Data	Top 6/ Bottom 6
Level	0.36
Z-score 4 week	-0.49
Z-score 5 week	-0.14
Z-score 6 week	0.11

Performance of level and time-series z-score of changes





Alternative Data From Data From Sensors

Satellite Data

- Car count (Orbital Insight)
- Wheat/Corn (RezaTec)
- Maritime (Windward)
- Cushing oil (Genscape)
- Copper/Zinc (RS Metrics)
- Challenges: clouds, seasonality, history

Geolocation Data for Footfall

- From ads – triangulation (Placed)
- From 3G and WiFi (AirSage)
- From apps (Advan Research)

Other Sensors

- Store Front: Cameras (Nomi) / Thermal (Irisys)
- Foot (ShopperTrak)
- Ceiling (RetailNext)
- Combined (Percolata)

Satellite Launch
- Planet Labs

AI to process images
- Orbital Insights

Trading Signals
- RS Metrics

Alternative Data From Data From Sensors

Case Study: Using Cellular Location to Estimate Retail Sales

Success rate in predicting next quarter sales using footfall data

Sector	Number of companies	Success rate for sales beats	Success rate for sales misses
C. Discretionary	43	58%	57%
Consumer Staples	6	25%	80%
Industrials	5	33%	30%
Health Care	3	63%	50%
Financials	11	91%	9%
Telecomms	2	25%	50%
Energy	3	40%	86%
Materials	1	100%	100%
Technology	2	25%	50%
Real Estate	4	50%	0%
Total	80	60%	52%

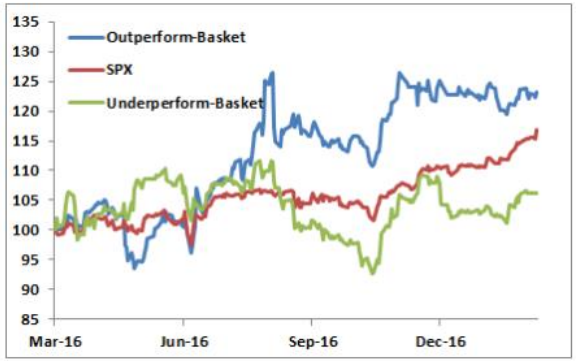
Signal Generation from Foot Fall Traffic Data

$$traffic = \frac{\sum_{across\ select\ apps} \%\ devices\ in\ store * devices}{\sum_{across\ select\ apps} devices}$$

$$Sales_{i+1} = Sales_i \frac{\sum_{each\ day\ in\ Q\ i+1} traffic}{\sum_{each\ data\ in\ Q\ i} traffic}$$

If $Sales_{i+1}$ > analyst estimates, long the stock, else short.

Performance of long and short legs (left) and long/short strategy vs S&P500 (right)



Strategy	Mean	Volatility	Sharpe ratio
Outperform-Basket	24.31%	20.36%	1.19
Underperform-Basket	7.79%	17.08%	0.46

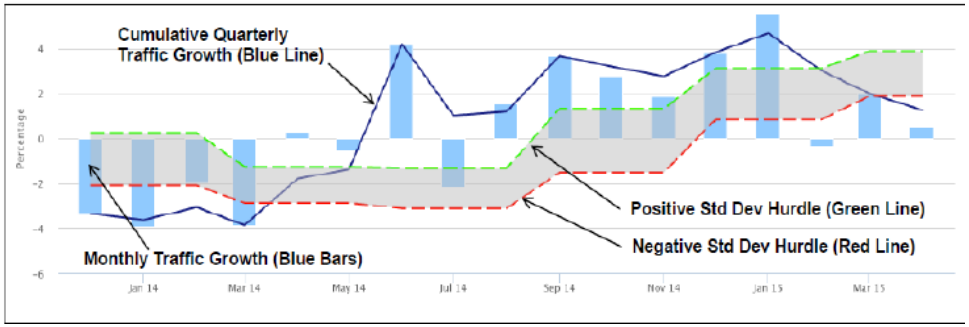
Strategy	Mean	Volatility	Sharpe ratio
Long-Short Strategy	16.52%	18.81%	0.88
SPX Index	17.12%	10.19%	1.68

Alternative Data From Data From Sensors

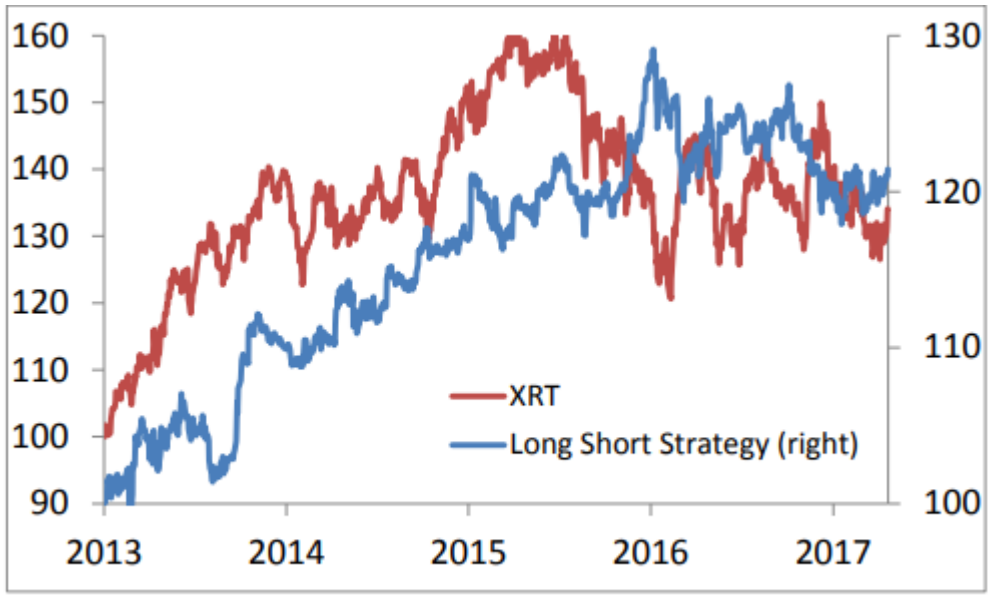
Case Study: Using Car Counts to Trade Retail Stocks

Sample satellite image (RS Metrics)

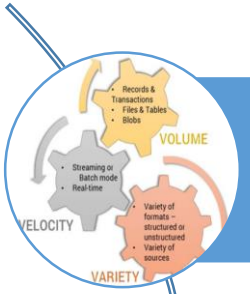
Bollinger bands for car count growth used for signal



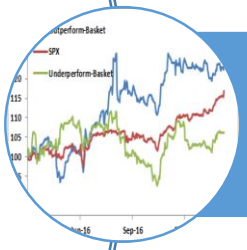
Performance of retail stock strategy using car count signal



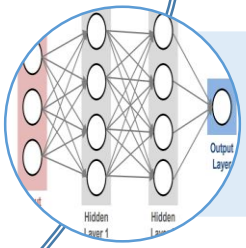
Strategy	Mean	Vol	Sharpe Ratio
L/S Strategy	4.8%	7.0%	0.68
Retail Sector	8.2%	16.9%	0.49



Part One:
Overview of Big Data and Machine Learning



Part Two:
Alternative Data



Part Three:
Machine Learning and A.I.

Statistics vs ML, and what can ML do for me?

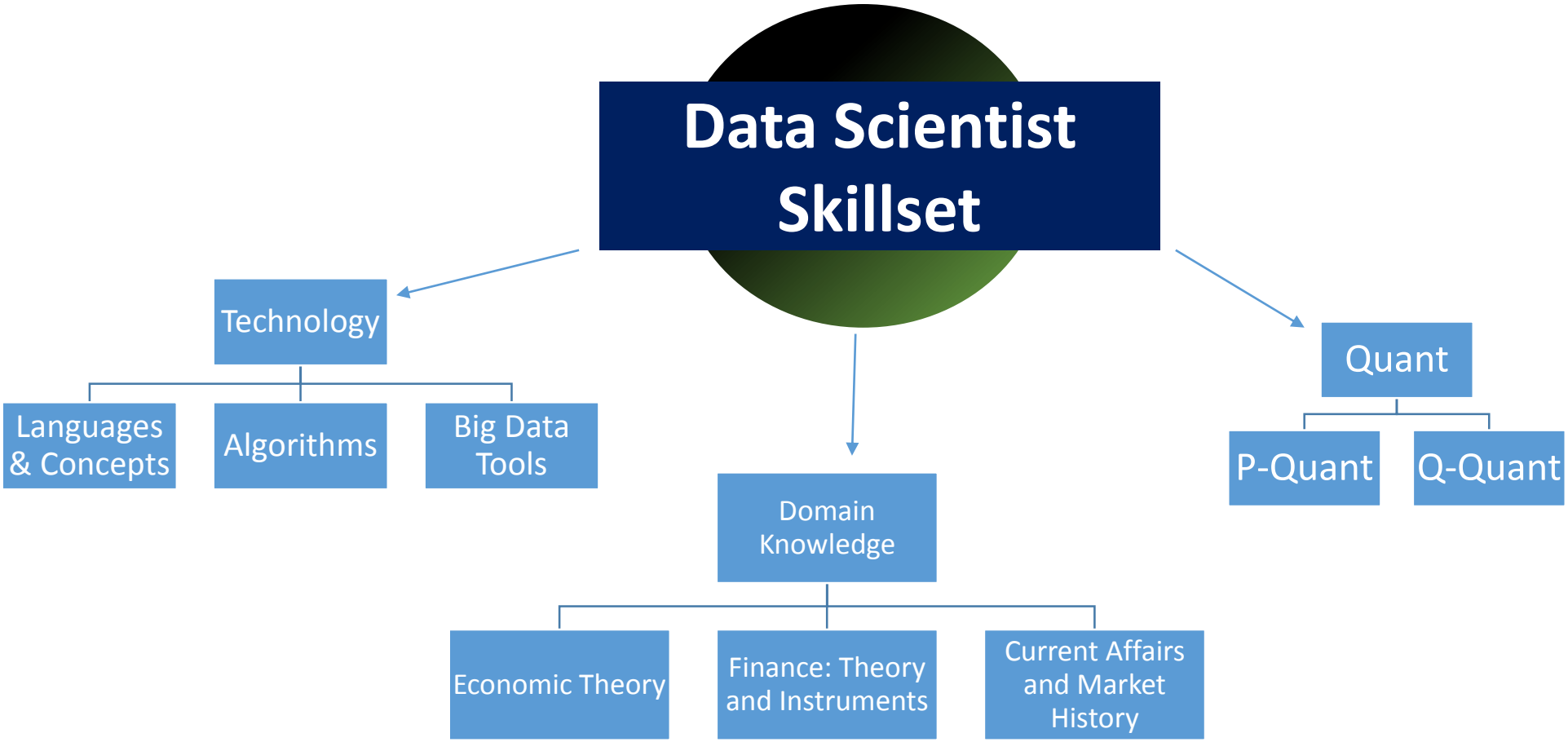
Statistics vs Machine Learning: terminology

Term in Statistics	Equivalent Term in Machine Learning
Statistical Learning	Classical Machine Learning
Independent Variable, X	Input Feature, attribute
Dependent Variable, Y	Output Feature, response
In-Sample	Training Set
Out-of-Sample	Test Set
Estimate, Fit a Model	Learn a Model
Model Parameters	Model Weights
Regression	Supervised Learning
Clustering and Dimensionality Reduction	Unsupervised Learning
Classifier	Hypothesis
Response (e.g. 0 or 1)	Label
Data Point	Example, Instance, Training Sample

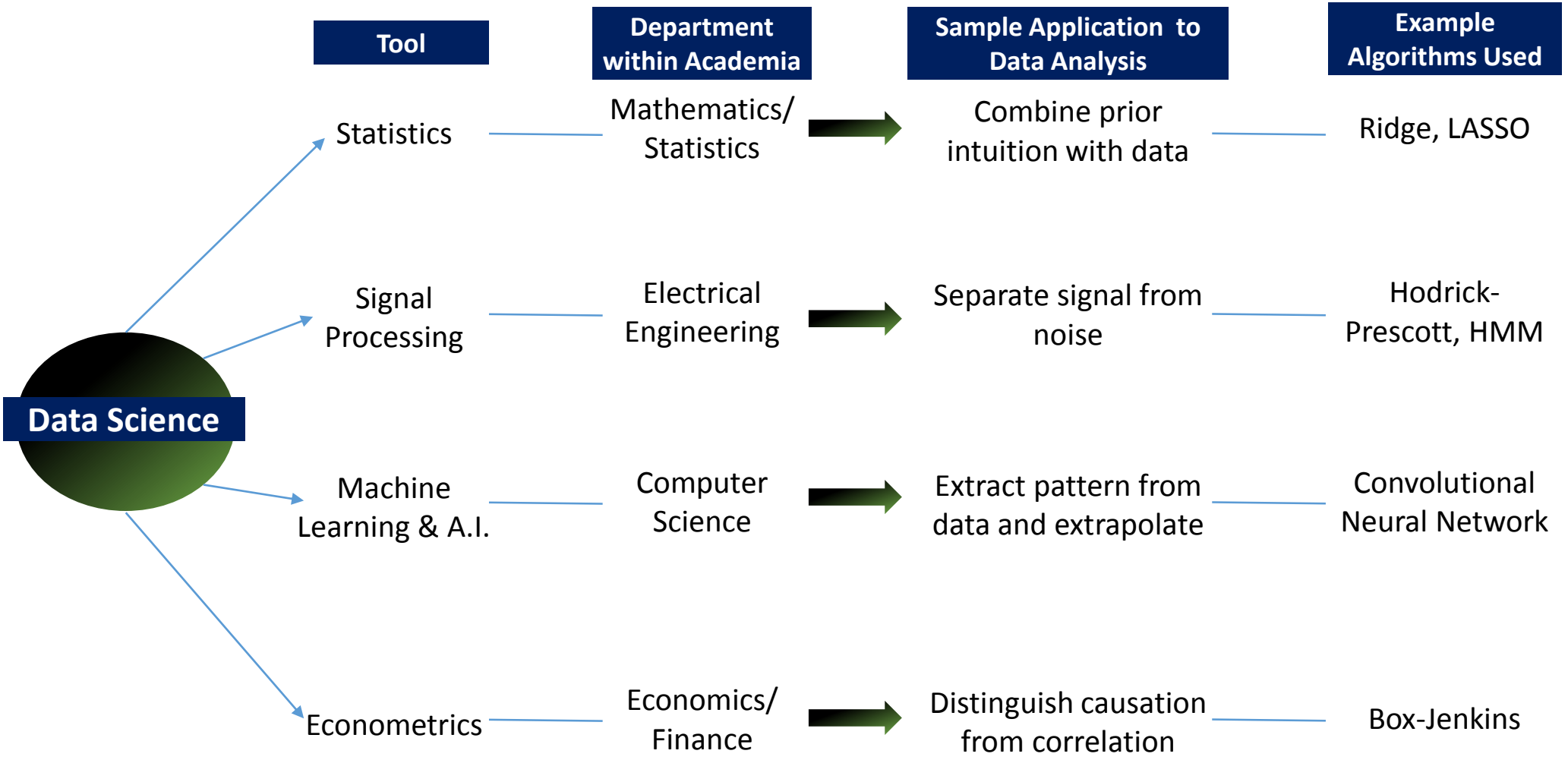
Frequently used Machine Learning methods

Question	Data Analysis Technique
Given set of inputs, predict asset price direction	Support Vector Classifier, Logistic Regression, Lasso Regression, etc.
How will a sharp move in one asset affect other assets?	Impulse Response Function, Granger Causality
Is an asset diverging from other related assets?	One-vs-rest classification
Which assets move together?	Affinity Propagation, Manifold Embedding
What factors are driving asset price?	Principal Component Analysis, Independent Component Analysis
Is the asset move excessive, and will it revert?	
What is the current market regime?	Soft-max classification, Hidden Markov Model
What is the probability of an event?	Decision Tree, Random Forest
What are the most common signs of market stress?	K-means clustering
Find signals in noisy data	Low-pass filters, SVM
Predict volatility based on a large number of input variables	Restricted Boltzmann Machine, SVM
What is the sentiment of an article / text?	Bag of words
What is the topic of an article/text?	Term/InverseDocument Frequency
Counting objects in an image (satellite, drone, etc)	Convolutional Neural Nets
What should be optimal execution speed?	Reinforcement Learning using Partially Observed Markov Decision Process

Skillset of a Data Scientist within Markets



What is “Data Science” ?

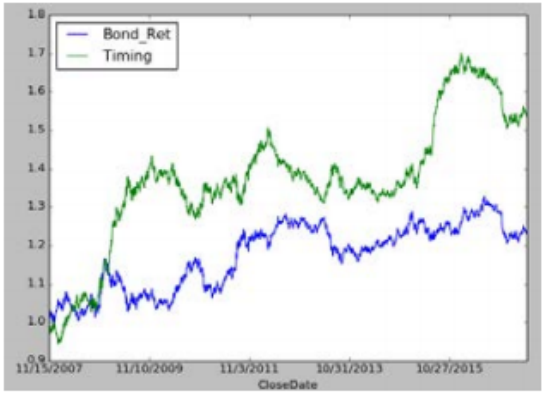


Identifying key drivers via regularization

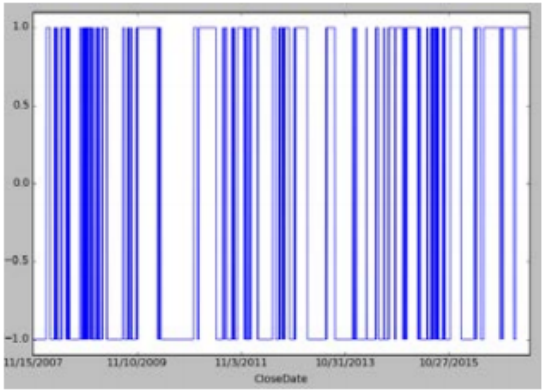
Application: Improving CTA Strategy

Result for 7-10 Treasury Bond Index: Lasso ($\alpha = 0.001$) yields IR = 0.67

PnL vs Always-Long



Buy/Sell Indicator



Performance Analytics

	Annualized Return (%)	Sharpe Ratio
IEF	2.30	0.32
IEF- Lasso	4.86	0.67
Correlation	-19.4%	

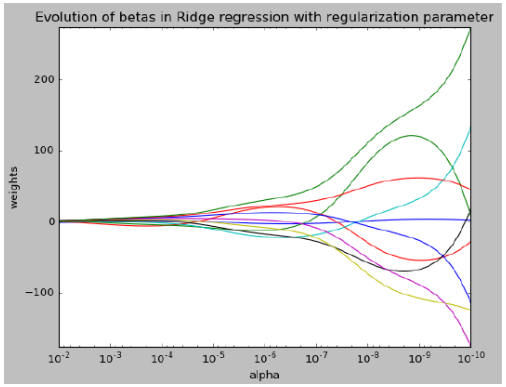
Sample Signals Chosen(12 to 16 factors chosen)

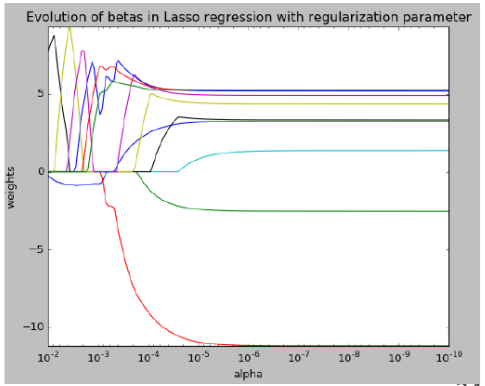
	1M	3M	6M	12M
Bond	✓	✓	✓	✓
Gold	✓	✓	✓	✓
Dollar	✓	✓	✓	✓
Equity	✓	×	✓	×

Predicting returns (Y) of 4 assets: S&P500, 7-10Y Treasury Bond Index, US Dollar, Gold
Momentum-based Features (X): 1,3,6,12M Lagged returns of same 4 assets (16 total)

Objective Functions

- **Lasso:** $[Y - (\beta_0 + \sum \beta_i x_i)]^2 + \alpha \sum |\beta_i|$
- **Ridge**: $[Y - (\beta_0 + \sum \beta_i x_i)]^2 + \alpha \sum \beta_i^2$
- **Elastic Net**: $[Y - (\beta_0 + \sum \beta_i x_i)]^2 + \alpha_1 \sum |\beta_i| + \alpha_2 \sum \beta_i^2$



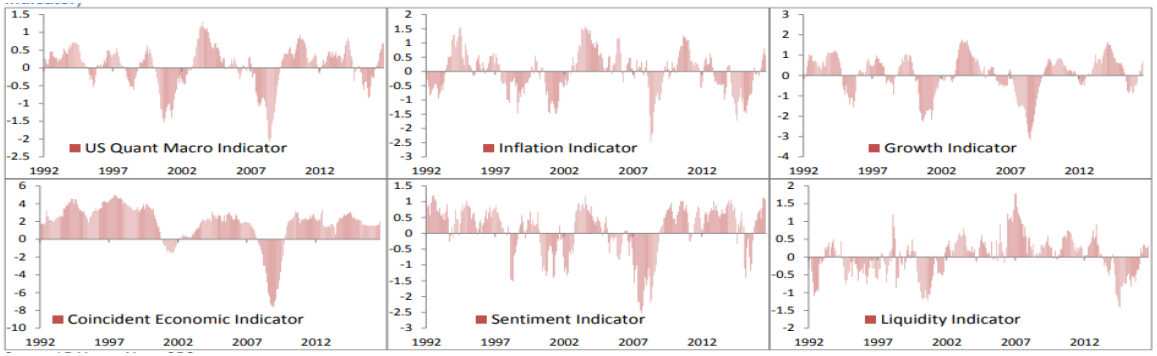


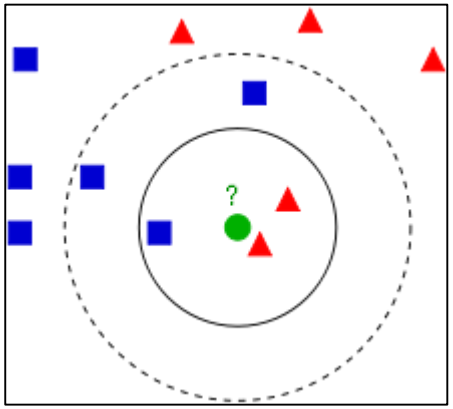
Macro Regime Identification for Style Rotation

K-Nearest Neighbors and LOESS

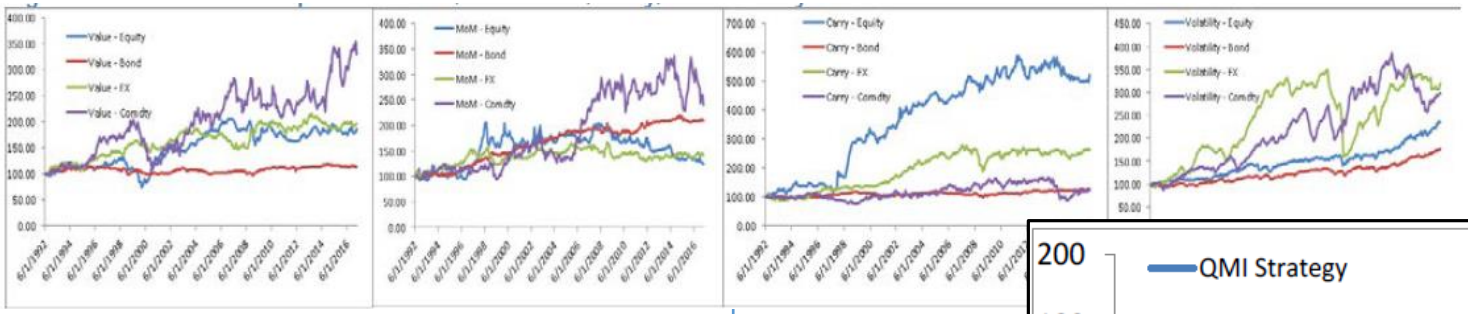
K-Nearest Neighbors to identify the Macro Regime for a long only strategy

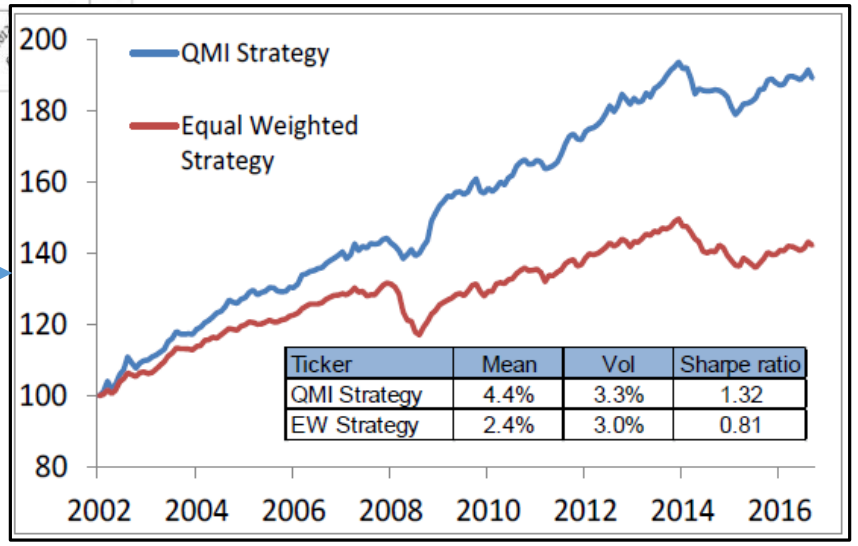
Using the values of macro indicator as vector in Euclidean space, find the K dates in history most like today. Then, long the S risk premia strategies that performed best.





Strategy results with K=2, S=14





Ticker	Mean	Vol	Sharpe ratio
QMI Strategy	4.4%	3.3%	1.32
EW Strategy	2.4%	3.0%	0.81

Predicting returns for ETF Sector rotation

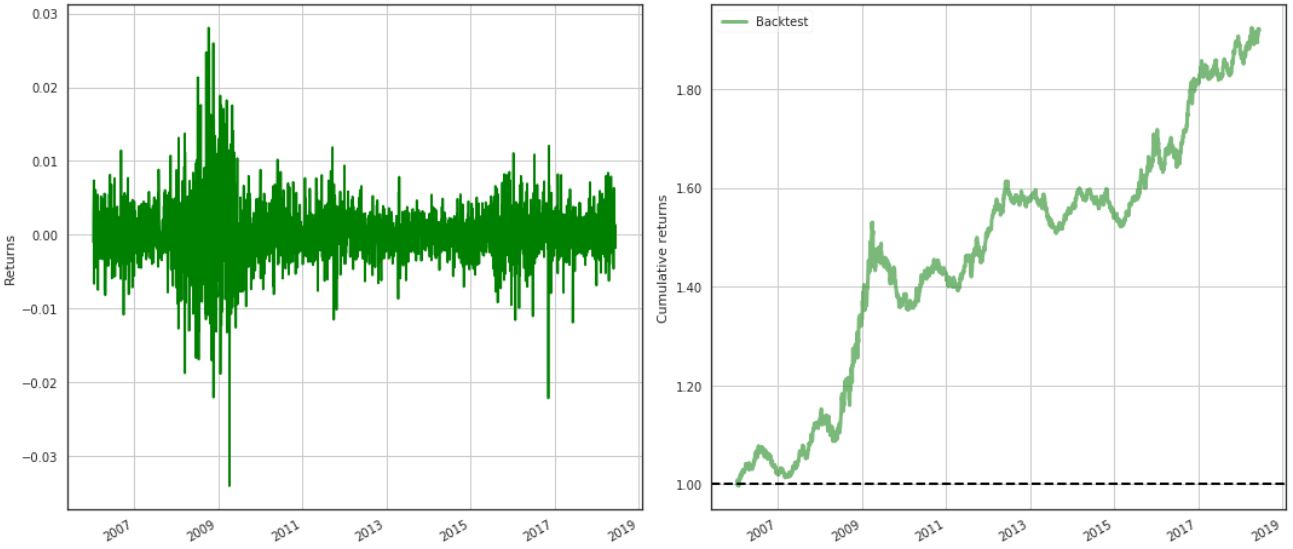
Extreme Gradient Boosted Trees using Macro Factors

Predicting sector returns (Y_k): 9 US Sector ETFs *(financials, energy, utilities, healthcare, industrials, technology, cons staples, cons discretionary, materials)*

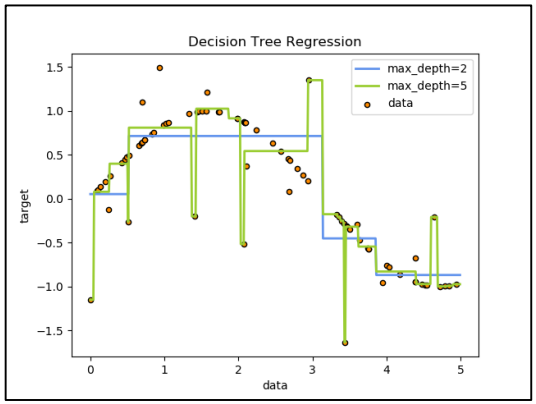
Macro-style Features (X_i): *Oil, Gold, Dollar, Bonds, Economic Surprise Index, 10-2Y spread, IG and HY credit spreads*

Long/short strategy performance 25 estimators and a maximum depth of 3: Sharpe 0.91

Daily (left) and cumulative returns (right).



The predictive power of the base estimators (decision trees-left) are enhanced through boosting and regularization.



Objective Function (with MSE training loss)

$$obj(\theta) = \sum_i (y_i - \hat{y}_i)^2 + \Omega(\theta)$$

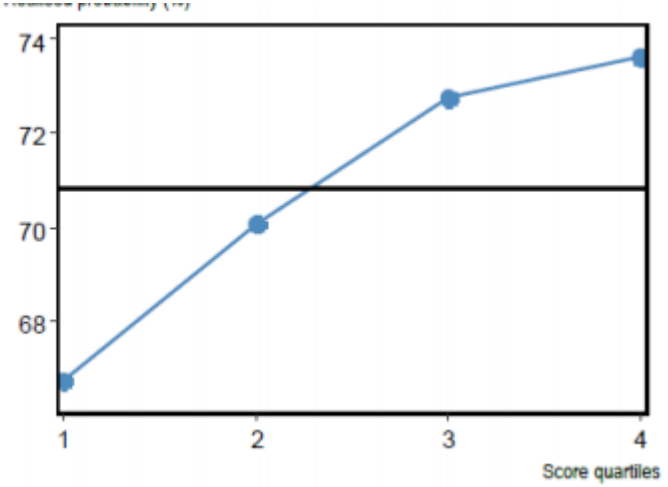
$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum w_j^2$$

Equity Call overwriting based on Accounting Parameters

Logistic Regression

VALUE *[30%]* P/E Vs Market (12mth fwd EPS) *[34%]* P/E Vs Country Sector (12mth fwd EPS) *[33%]* EPS Growth (FY1 mean to FY2 mean) *[33%]*	**MOMENTUM/Technical** *[20%]* 12Mth Price Momentum *[75%]* 1Mth Price Reversion *[25%]*
GROWTH *[30%]* Earnings Momentum 3mth (Risk Adj.) *[34%]* Net Revisions to mean FY2 EPS *[33%]* 1Mth change in consensus recomms *[33%]*	**QUALITY** *[20%]* Historical ROE *[50%]* Earnings Certainty (Var. in forecast EPS) *[50%]*

Factor	Coef Estimate	Z-value
3M Realised Volatility	-0.36	-6.1
Historical ROE	-0.06	-1.5
1M Price Momentum	-0.05	-1.2
1Y Earnings Yield vs. Country Sector	-0.08	-1.2
EPS Growth	-0.05	-1.0
1Y Earnings Yield	-0.05	-0.8
Earnings Momentum 3M	-0.03	-0.5
Net revisions to Mean FY2 EPS	-0.02	-0.4
1M Change in Consensus Recs	0.00	-0.1
Earnings Certainty	0.01	0.2
12M Price Momentum	0.11	2.3



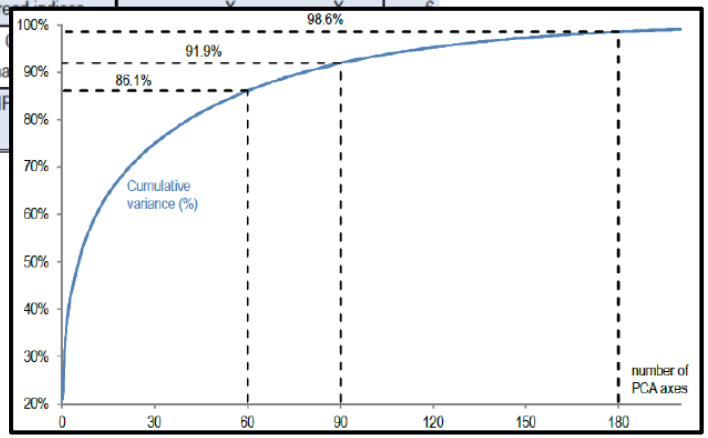
Predicting FX Vol Option PnL using Macro Factors

Principal Component Analysis + Support Vector Machines

PCA to reduce dimensionality and algorithm for classification to buy/sell/remain neutral on ATM EURUSD

Market data from 377 Features…

Market data type	Market data	Level	1 week change	1M change	Count
FX realised vols	2M realised vols in USD vs G10, MXN, BRL, ZAR, TRY, INR and KRW	X	X	X	45
FX ATM vols	1M, 3M and 1Y ATM in same pairs	X	X	X	135
FX skews	3M 25D RRs	X	X	X	45
FX spots	FX Spots in 15 G10 and EM USD pairs		X	X	30
Depo rates / Basis	3M FX Forward drops		X	X	30
Interest Rates	10Y Gov yields: US, Japan, UK, Germany, France, Italy, Spain, Australia		X	X	16
Equity Indices	S&P, Nikkei, FTSE, E-Stoxx, ASX, Mexbol, Bovespa, KOSPI, Hang Seng		X	X	18
Commodities	Gold and Brent spot		X	X	4
Credit spreads	CDX IG and HY, iTraxx spre				
EASI indices	Global, US, CAD, EU, UK, SEK, Japan, AU, NZ, China				
IMM positions	USD, EUR, JPY, GBP, CHF NZD, MXN, RUB, Gold				



… to predict ATM EURUSD outcome

Algorithm	Raw Data	Normalised	PCA 180	PCA 90	PCA 60
kNN	69.0%	83.9%	83.7%	83.3%	84.1%
SVC (polynomial kernel - degree 3)	59.5%	74.1%	82.4%	84.1%	84.1%
SVC (linear kernal)	62.1%	80.8%	83.3%	83.3%	82.4%
Ridge Regression	81.0%	80.8%	73.9%	68.4%	69.3%
Gaussian NB	67.2%	68.4%	69.2%	69.5%	69.3%
Linear Discriminant Analysis	81.2%	81.2%	73.4%	68.6%	68.2%
Logistic Regression	79.3%	79.9%	74.1%	68.6%	68.0%
CART Decision Tree	75.7%	76.1%	61.9%	68.6%	67.6%
PAC Regression	48.5%	76.6%	63.2%	65.7%	60.3%
SGD Regression	41.2%	76.4%	69.3%	64.9%	57.5%

		Predicted			
		Long	Neutral	Short	Recall
Actual	Long	**30**	26	0	53.6%
	Neutral	4	**304**	16	93.8%
	Short	0	46	**96**	67.6%
	Precision	88.2%	80.9%	85.7%	

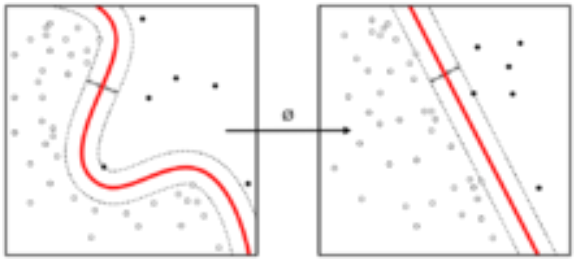
Support Vector Machine and Classification

Nice analytical properties made SVM popular in academic circles…

$$\min_{w,b,\zeta} \frac{1}{2} w^T w + C \sum_i \zeta_i$$

$$s.t.\ y_i(w^T \phi(x_i) + b) \geq 1 - \zeta_i$$

$$\zeta_i \geq 0, i = 1, \ldots, n$$



Measuring up

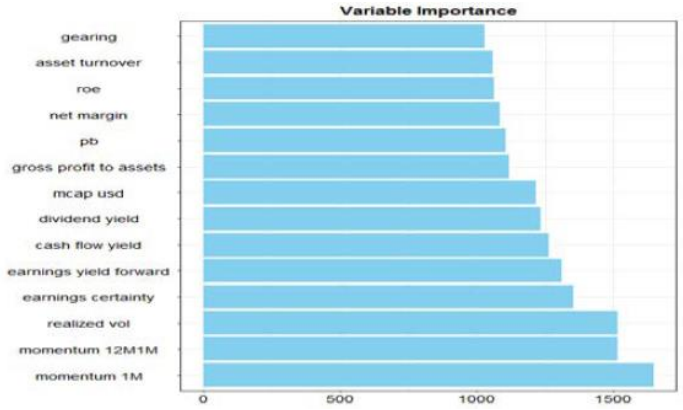
- **Precision:** $\frac{T_p}{T_p + F_p}$
- **Recall:** $\frac{T_p}{T_p + F_n}$
- **Harmonic Mean (F1):** $\frac{1}{\frac{1}{P} + \frac{1}{R}}$

Equity L/S using Accounting Parameters

Random Forests

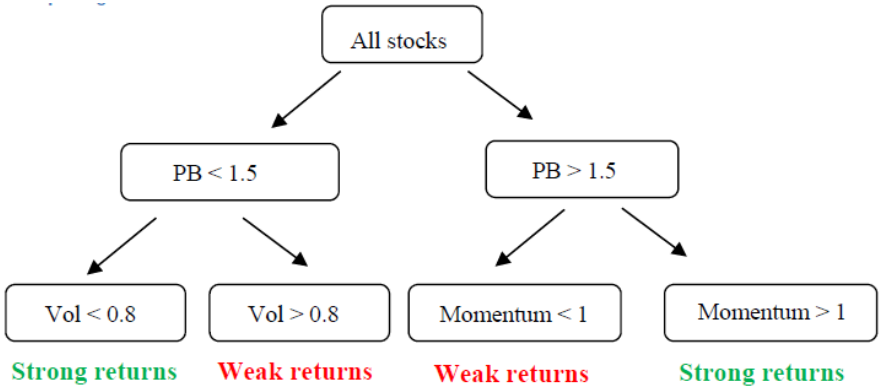
14 Accounting parameters to predict the returns of 1400 single name stocks

Factors	
Price-to-book ratio	Earnings Certainty
Gross profit / Total assets	Cash flow yield
ROE	Dividend yield
Net Margin	Realized vol
Asset Turnover	1M momentum
Gearing	12M-1M momentum
Forward earnings yield	Market cap



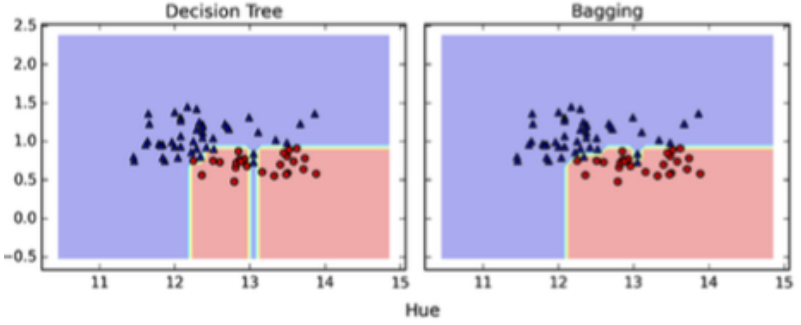
Basket	Cum. Return	CAGR	Volatility	IR	Max Drawdown	Hit Ratio
1 (low)	3.0%	1.0%	11.3%	0.09	27.7%	37.3%
2	7.9%	2.5%	10.7%	0.24	22.7%	37.8%
3	6.4%	2.1%	10.7%	0.19	23.6%	38.3%
4	12.8%	4.1%	10.6%	0.38	21.9%	37.0%
5 (high)	19.2%	6.0%	11.1%	0.54	20.9%	39.5%
L/S	15.4%	4.8%	4.2%	1.16	6.9%	37.7%

Bagging Decision Trees: Random Forests



Reducing Variance via Bagging:

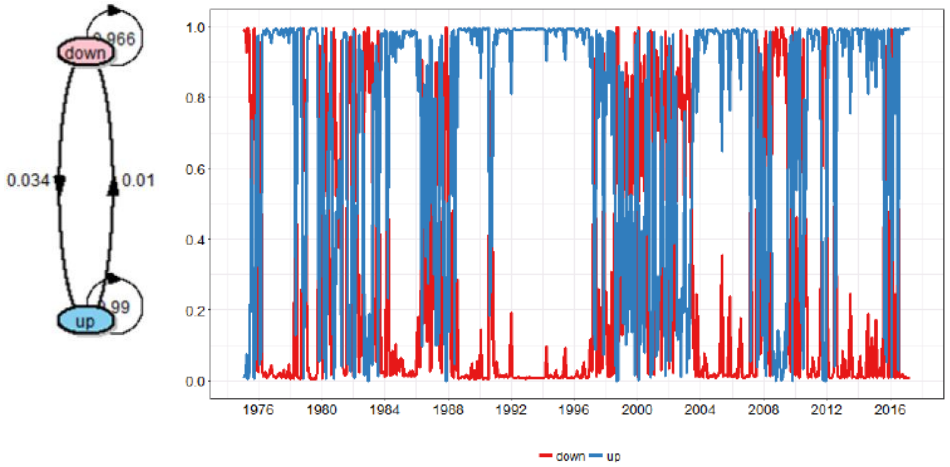
$$Variance \sim \rho\sigma^2 + \frac{1-\rho}{B}\sigma^2$$



Regime Detection For Market Timing

Hidden Markov Model

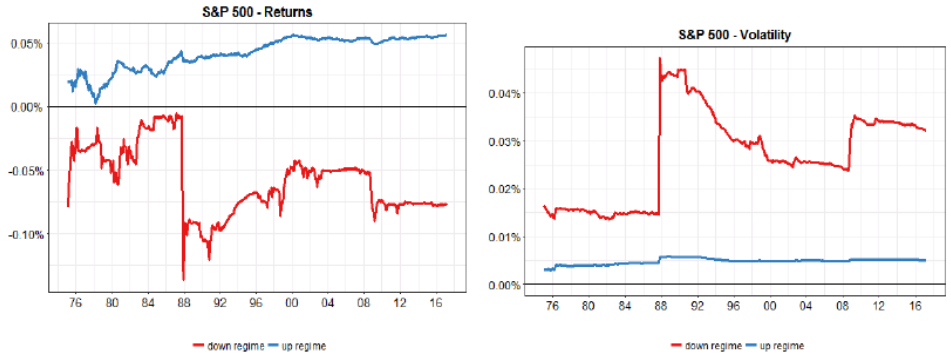
Using HMMs to estimate the probability of being in an up or down 'state' on the last trading day of each month.



Long/short S&P 500 when model predicts up/down gives better results than long S&P only.

	Long-only	Timing
CAGR (%)	8.2	6.1
Volatility (%)	16.8	11
Information ratio	0.49	0.56
Max DD (%)	56.8	38.4

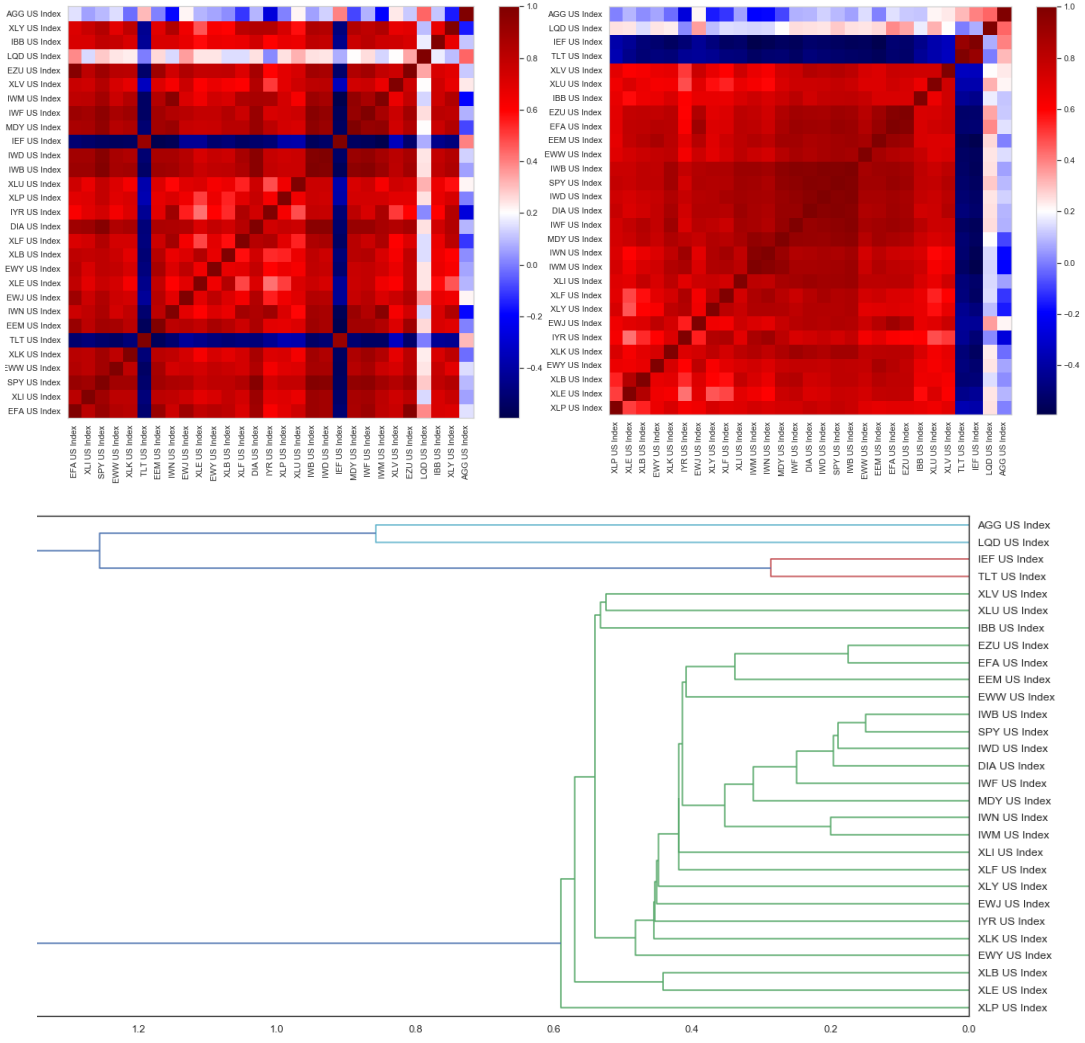
Returns and volatility clearly show characteristics of their respective regimes



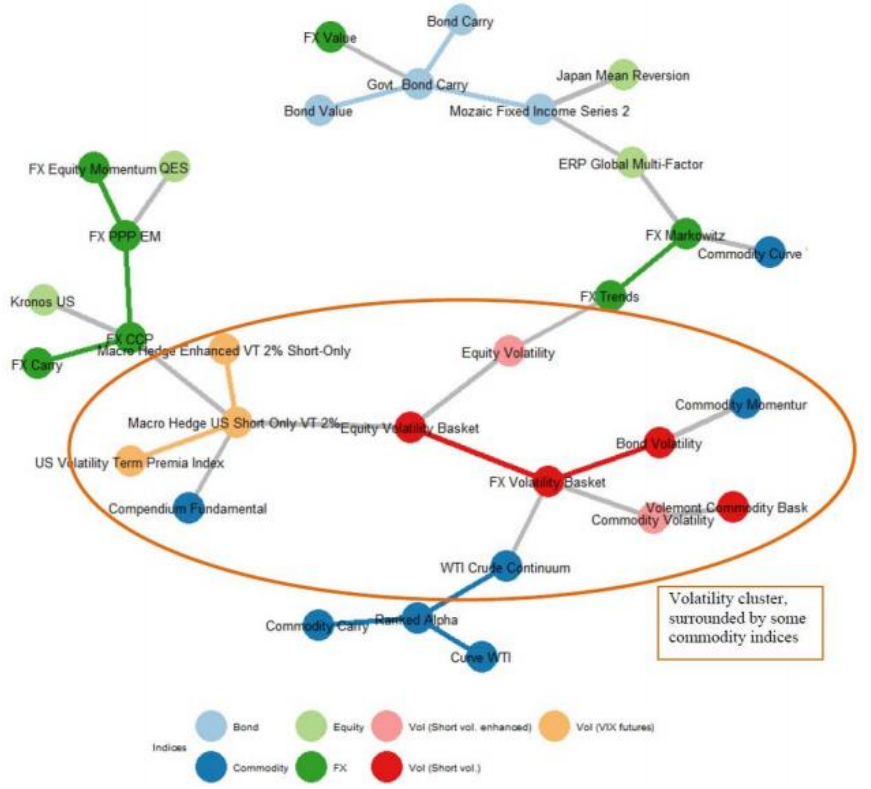
Uncovering structure in US Equities and Risk Premia

Unsupervised Learning Techniques

Hierarchical Clustering US ETFs



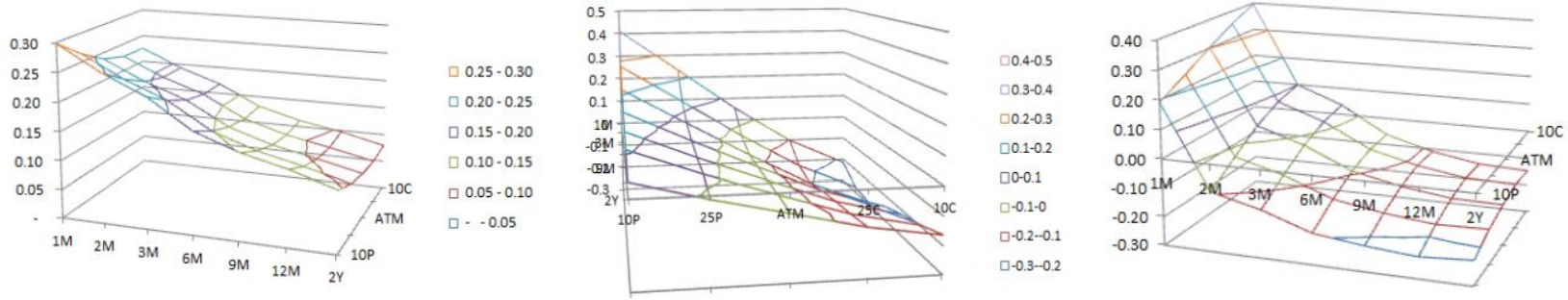
Minimum Spanning Tree: Risk Premia



Understanding the USD-JPY Volatility Surface

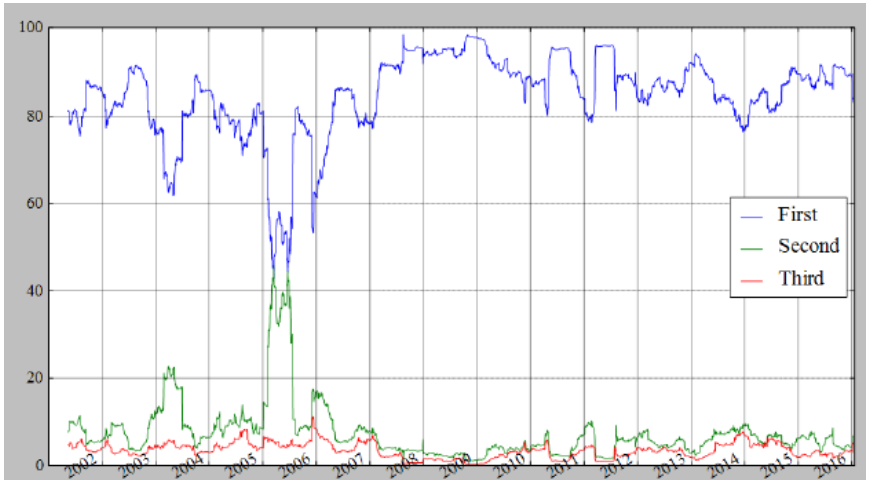
Principal Components Analysis (PCA)

First 3 principal components of USDJPY…



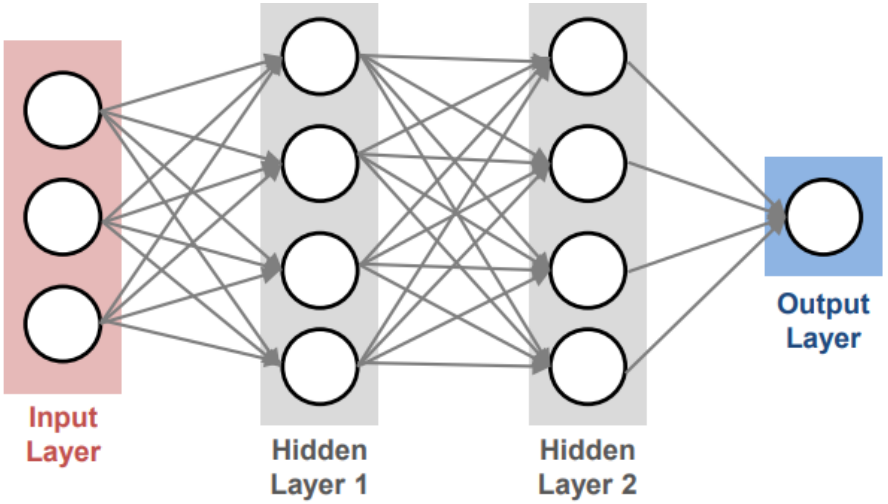
Explains a large proportion of the variance

Number of principal components	Percentage of variance explained (%)
1	88.04
2	4.82
3	3.66
4	1.85
5	0.60



Deep Learning and AI

Example Neural Network with 2 hidden layers (4 neurons each), 3 input variables, and 1 output variable



Additional attributes that characterize a neural network

Feature of Neural Network	Role in Network Design and Performance	Most Common Example	Other Examples Used in Practice
Cost Function	Used to calculate penalty/error in prediction versus true output	Mean squared error (for regression), Binary cross-entropy (for classification)	Mean absolute error, Categorical cross-entropy, Kullback-Leibler divergence, Cosine proximity, Hinge/Squared-Hinge, log-cosh
Optimizer	Used to calibrate network weights based on error	Stochastic Gradient Descent or SGD	RMSprop[54], Adagrad, Adadelta, Adam/Adamax/Nestorov-Adam
Initialization Scheme	Used to initialize network weights	Xavier (including Glorot-Normal and Glorot-Uniform)	Ones/Zeros/Constant, Random Normal/Uniform, Variance Scaling, Orthogonal, Le Cun – Uniform, He – Normal/Uniform
Activation Function	Used at the end of each neuron after the weighted linear combination to get non-linear effect	ReLU (for all intermediate layers), Linear (for final layer in regression), Sigmoid (for final layer in classification)	Softmax/Softplus/Softsign, Leaky/Parametrized ReLU, tanh, Hard Sigmoid
Regularization Scheme	Used to penalize large weights to avoid overfitting	Dropout	L1/L2 regularization for kernel, bias and activity

Predicting returns for ETF Sector rotation

Neural Networks with Macro Factors

Predicting sector returns (Y_k): 9 US Sector ETFs *(financials, energy, utilities, healthcare, industrials, technology, cons staples, cons discretionary, materials)*

Macro-style Features (X_i): *Oil, Gold, Dollar, Bonds, Economic Surprise Index, 10-2Y spread, IG and HY credit spreads*

MLP	Bond	Commodity	Equity	FX
Carry	-17.8%	0.5%	5.5%	9.7%
Momentum	-10.4%	0.2%	0.7%	-1.4%
Value	-18.7%	-1.7%	-0.3%	-11.2%
Volatility	0.4%	0.9%	8.4%	8.8%
Beta	-17.0%	7.4%	25.6%	2.7%

Example code for Neural Network architecture

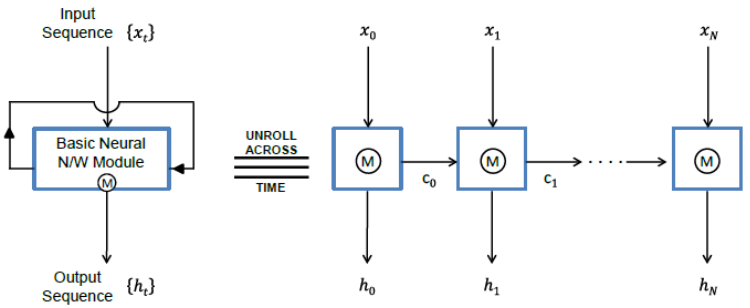
8 neurons at input layer, 8 neurons in hidden layer, single output

```
model = Sequential()
model.add(Dense(8, activation='relu',input_dim=8))
model.add(Dropout(dropout))
model.add(Dense(8, activation='relu'))
model.add(Dropout(dropout))
model.add(Dense(1))
```

Future: Remembering & Picturing trends

LSTM and CNN

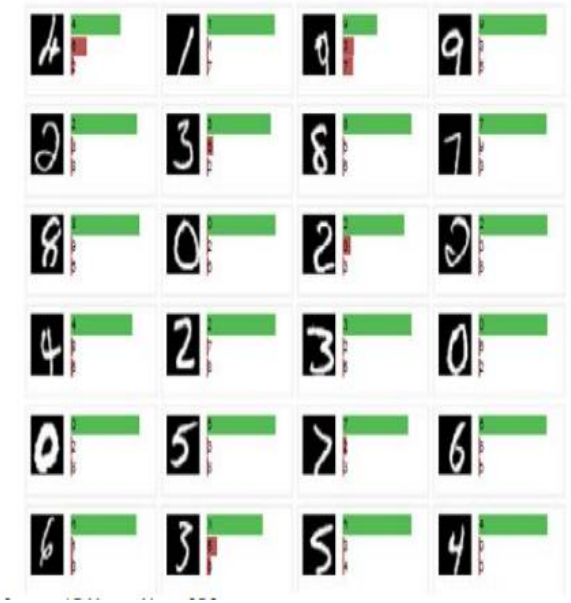
Long Short-Term Memory Network architecture is a class of Recurrent Neural Networks (RNNs), allowing for retention of recent events.

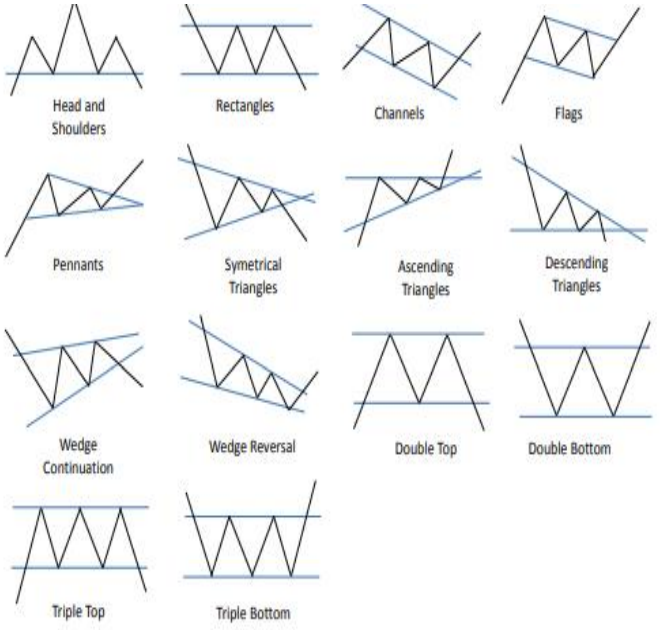


Initial results are not conclusive, but future work in progress

Prediction for 2-month look-ahead period by LSTM for SPX price levels

Convolutional neural networks are state of the art image classifiers (left: classifying handwritten digits). Potentially, the same technology can be used to identify technical patterns.

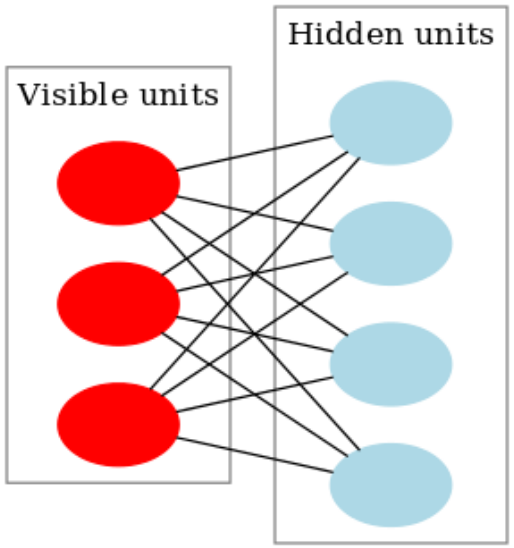




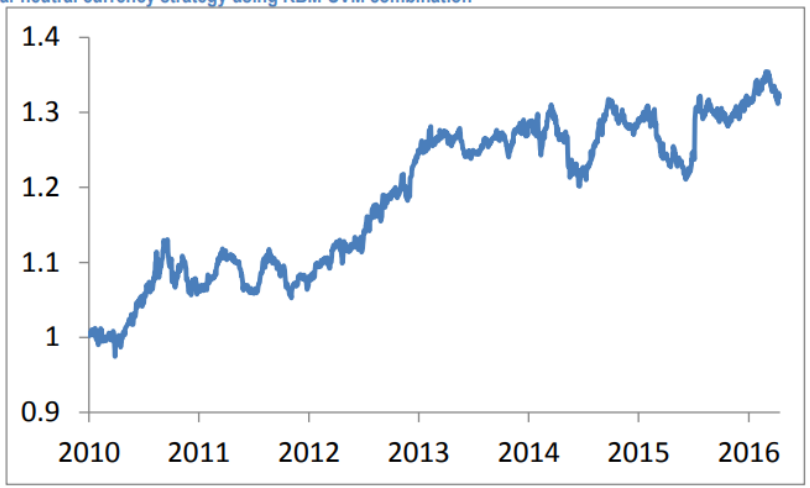
Neural Networks for Dimensionality Reduction

FX trading example using Restricted Boltzmann Machines

RBM: A graphical model for factor analysis



PnL of dollar neutral currency strategy using RBM-SVM combination



Performance	
Annual Return	4.5%
Annual Volatility	6.7%
Sharpe Ratio	0.7
ρ to SPX	13.8%
ρ to DXY	-6%

Theory

$$E(v,h) = a^T v - b^T h - v^T W h$$

$$P(v,h) = \frac{1}{Z} e^{-E(v,h)}$$

$$\underset{W}{argmax} \prod_{v \in V} P(V)$$

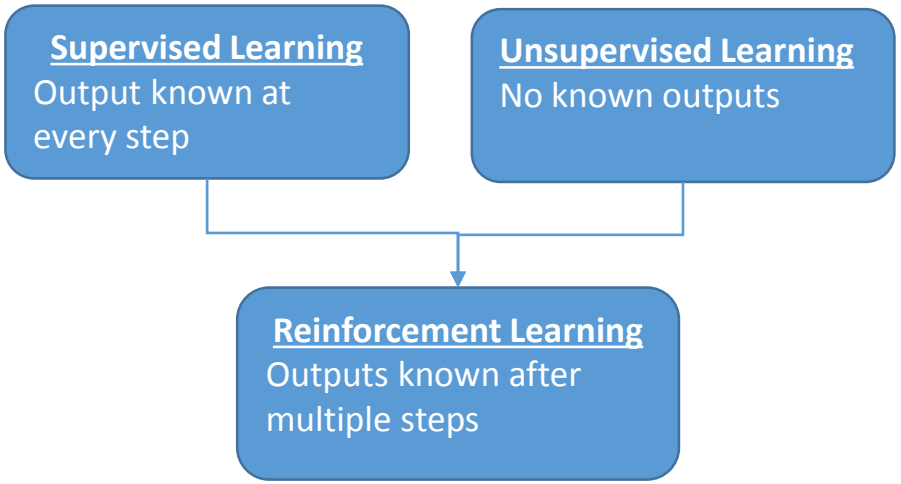
Data	Features	Neural Network : Dimensionality Reduction	Supervised Learning: Predict next day returns	Long/short
•FX Assets: AUD, CAD, CHF, DKK, EUR, GBP, JPY, NOK, NZD, SEK	•Lagged daily return over past 10days (=100 features)	•RBM on scaled inputs (μ=-, r=1) to get to 20 factors	•Support vector regression (c:regularization term, γ: RBF kernel	•Daily rebalancing: Long top 3, short bottom 3

Other applications: Collaborative filtering, topic modeling, classification

Hype: TBD in initial phase of deep learning

Reinforcement Learning: Premise for Ever-Improving AI

From self-driving cars to algorithmic execution

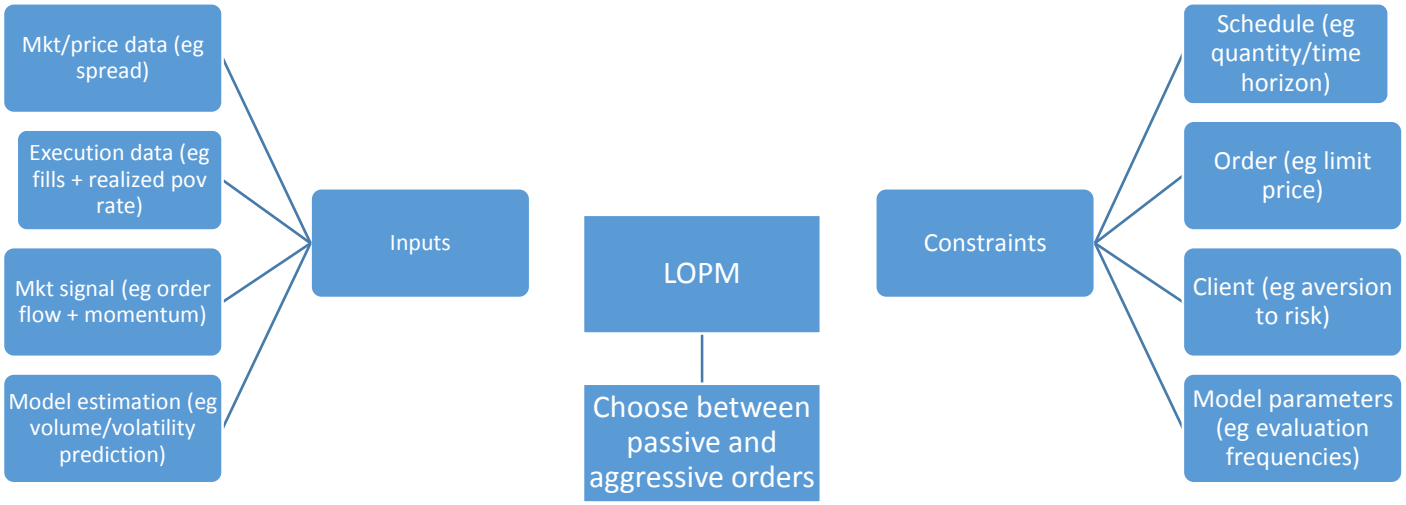


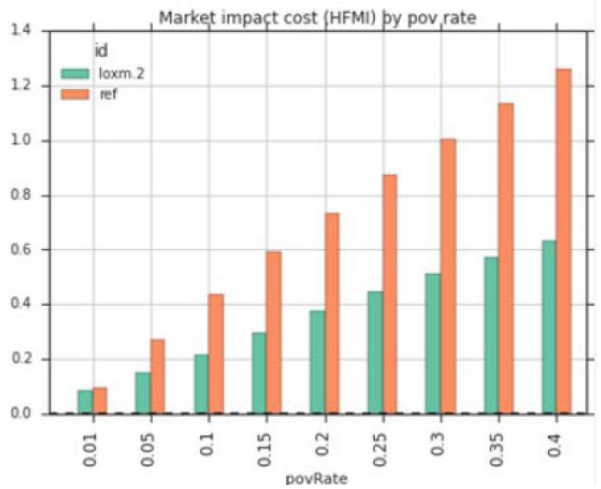
- PnL/outcome not known after one step, but after multiple steps.
- Make predictions *repeatedly* and improve *relentlessly*.
- Evolved as approximation to dynamic programming when state space is unknown.
- 2 challenges define **Deep Q-Learning**:
  - Explore vs Exploit
  - Credit Assignment Problem

Bellman Equations: Handling partially observable Markov decision models

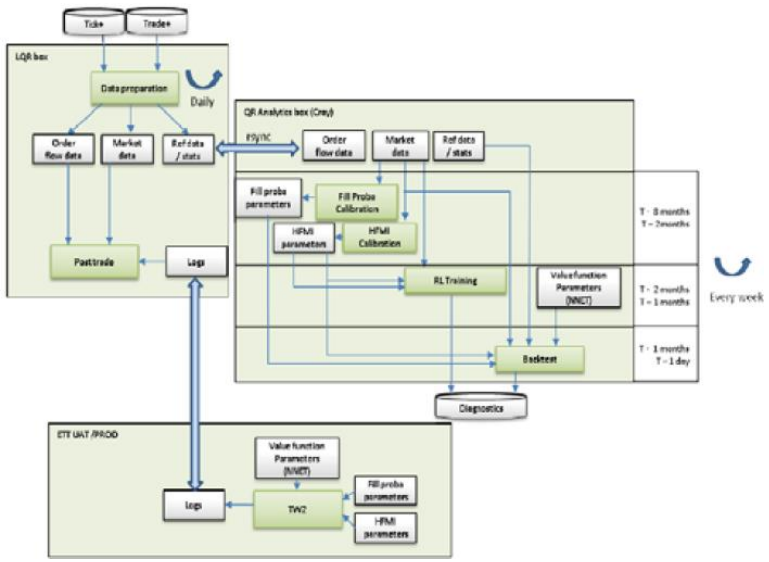
- Sequence of status/actions/reward: $s_0, a_0, r_0, \dots, s_n, a_n, r_n$
- Discounted future reward: $R_t = r_t + \gamma r_{t+1} + \cdots + \gamma^{n-t} r_n$
- $Q(s_t, a_t) = \max_{a_t} R_{t+1} \rightleftarrows \pi(s) = \max_a Q(s, a)$
- $Q(s_t, a_t) = r_t + \max_{a_{t+1}} Q(s_{t+1}, a_{t+1})$

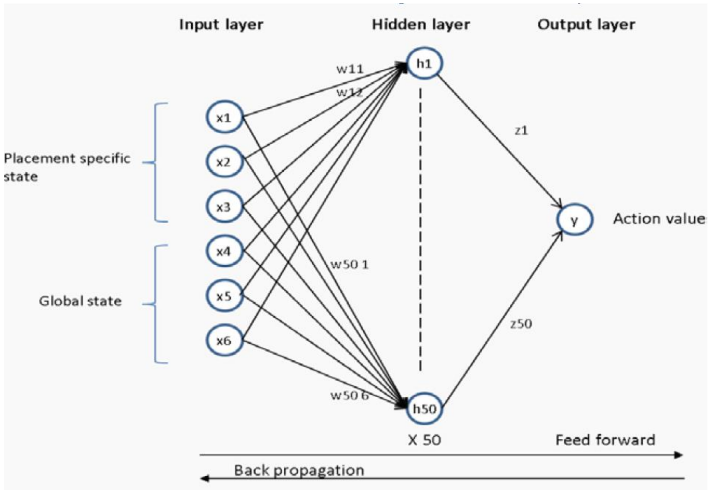
Reinforcement Learning in Practice





- Attempts to incorporate within HFT and sell-side algorithmic execution: *Limited success so far*
- Competitor's attempt to use Deep Q-learning in Limit Order Placement module
- <u>Aim</u>: Minimize slippage within constraints





Conclusions

Evolution or Revolution	• Big/Alternative Data and Machine Learning are here to stay
Taxonomy of Alternative Data	• By individuals, business processes and sensors – with sentiment, transaction, geo-location and satellite as exemplars
Taxonomy of Data Analysis	• Tools drawn from econometrics, signal processing, statistical learning and AI
Impact of Machine Learning	• Impact across the investment landscape, including stock selection, sector/style rotation, yield generation and portfolio construction.