Financial Data Professional Institute
A Conversation with Tony Guida
"Long Term Machine Learning Predictions for US equity"

Mehrzad Mahdavi, Executive Director, FDP Institute
Kathy Wilkens, Senior Advisor, FDPI Curriculum
Mirjam Dekker, Project Manager, FDP Institute

www.fdpinstitute.org
March 5, 2020
Agenda

- Welcome
- Introductions
  - Tony Guida
    Executive Director, Sr. Quant Research
    Ram Active Investment
  - Katherine Wilkens
    Senior Curriculum Advisor FDPI
  - Mehrzad Mahdavi
    Executive Director FDPI
  - Mirjam Dekker
    Project Manager FDPI
- User case “Long Term ML Predictions for US Equity”
- FDP Curriculum
- Q & A
[LT ML predictions for EQ]
An empirical exercise
A bit of Epistemology
A New Way for Research
Asset pricing vs Empirical Asset Pricing

• Econometrics vs machine learning
• Share a common goal: build a predictive model
• Radical difference remains in the “learning” part
• Econometrics is a beta question while ML is an alpha answer
• From a practitioner standpoint ML more suited to high dimensional non-linear signals’ space
• Poses the problem of maximizing “factor zoo”
[LT ML predictions for EQ]
definitions and concepts
eXtreme Gradient Boosting: quick introduction

General objective of tree ensemble for K trees

\[ \hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F} \]

\[ \text{Obj} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \]

Training on loss

\[ \hat{y}_i^{(0)} = 0 \]
\[ \hat{y}_i^{(1)} = f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i) \]
\[ \hat{y}_i^{(2)} = f_1(x_i) + f_2(x_i) = \hat{y}_i^{(1)} + f_2(x_i) \]
\[ \vdots \]
\[ \hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \]

Complexity of the trees

Additive training

http://xgboost.readthedocs.io/en/latest/model.html#
Wisdom of the crowd in ML

**Simple example:**
Assuming independent classifiers
Classifier has an error rate $\varepsilon < 0.5$
Ensemble prediction better than random guess
If $\varepsilon > 0.5$ for each classifier, ensemble wrong prediction will increase

Boosted Tree example

Source: “Machine Learning for Factor Investing” Coqueret, Guida (2020 Chapman & Hall)
Measuring the Quality of a ML model

- **Left axis (vertical)** of the matrix shows **Actual**
- **Top axis (horizontal)** shows **predicted**

<table>
<thead>
<tr>
<th>Outperformed</th>
<th>Underperformed</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Positive: Stock WAS classified as outperforming and DID outperformed</td>
<td>False negative: Stock was NOT classified as outperforming and DID outperformed</td>
</tr>
<tr>
<td>False Positive: Stock WAS classified as outperforming and did NOT outperformed</td>
<td>True Negative: Stock was NOT classified as outperforming and DID not outperformed</td>
</tr>
</tbody>
</table>
Beyond confusion matrix

- **Fp**: false positive. Stock predicted to outperform and that did not outperform out of sample.
- **Fn**: false negative. Stock predicted to underperform that outperform out of sample.
- **Tp**: true positive. Stock predicted to outperform which outperform out of sample.
- **Tn**: true negative. Stock predicted to underperform which underperform out of sample.

**Precision**: $\frac{Tp}{(Tp + Fp)}$
Precision could be defined as a rate of successful prediction for sector neutral outperforming stocks.

**Recall**: $\frac{Tp}{(Tp + Fn)}$
Recall could be defined as a true rate, since we include the instances that have been wrongly classified in negative.

**Accuracy**: $\frac{(Tp + Tn)}{(Tp+Tn+Fp+Fn)}$
This is the accuracy level used in the cross validation part.

**F1 score**: $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$
[LT ML predictions for EQ] dataset & E.D.A
Objective, data and protocol

• We will compare different labels corresponding to different prediction horizon for cross sectional returns
  • (1M, 3M, 6M, 9M, 12M, 18M, 36M)
• Investment universe is US stocks (~1500)
• Full dataset from Dec-1999 until Dec-2019
• (~100) features, monthly normalised in percentrank.
• Dataset pre-processed, outliers removed, focusing on training on the tails of the distribution (top and bottom 25%) excluding the top 1% avoiding to train on high vol.
• Split the dataset between Training (80%) and Testing (20%)
• Rolling window of 60 months
Features engineering: Training on tails

Training trees on tails with applications to portfolio choice

Guillaume Cogueret & Tony Guida

Abstract

In this article, we investigate the impact of truncating training data when fitting regression trees. We argue that training times can be curtailed by reducing the training sample without any loss in out-of-sample accuracy as long as the prediction model has been trained on the tails of the dependent variable, that is, when ‘average’ observations have been discarded from the training sample. Filtering instances has an impact on the features that are selected to yield the splits and can help reduce overfitting by favoring predictors with monotonous impacts on the dependent variable. We test this technique in an out-of-sample exercise of portfolio selection which shows its benefits. The implications of our results are decisive for time-consuming tasks such as hyperparameter tuning and validation.
Features correlation example

Source: Guida, Coqueret. Chapter 7, Ensemble Learning Applied to Quant Equity – Big Data and Machine Learning in Quantitative Investment
Creating the dataset

Some features examples
- Fundamental trailing
- Price based
- Volume based
- Risk based
- Composites
Rolling Windows for training (case for 12M forward)

In this example we use a rolling window of **60 months** to predict the **12M forward performance** of a stock.

Dec 99

TRAINING (60 Months)

Training set (80%)  Test set (20%)

Dec 04

1st Prediction using trained model, Dec 05

Jan 00

[.................]

Jan 05

2nd Prediction using trained model, Jan 06

Feb 00

[......]
[LT ML predictions for EQ] Building & Training models
Hyperparameters:

- **The learning rate, \( \eta \):** it is the step size shrinkage used in update to prevent overfitting. After each boosting step, we can directly get the weights of new features and \( \eta \) actually shrinks the feature weights to make the boosting process more conservative.

- **The maximum depth:** it is the longest path (in terms of node) from the root to a leaf of the tree. Increasing this value will make the model more complex and more likely to be overfitting.

- **Regression \( \lambda \):** it is the \( L^2 \) regularization term on weights (mentioned in the technical section) and increasing this value will make the model more conservative.

- **gamma:** minimum loss reduction required to make a further partition on a leaf node of the tree. The larger, the more conservative the algorithm will be.

<table>
<thead>
<tr>
<th>model</th>
<th>max_depth</th>
<th>eta</th>
<th>round</th>
<th>eval_metric</th>
<th>subsample</th>
<th>col_by_sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>XGB</td>
<td>5</td>
<td>1%</td>
<td>150</td>
<td>error</td>
<td>0.8</td>
<td>0.8</td>
</tr>
</tbody>
</table>
LT vs the rest: impact on training

We compare the accuracy in training and test for each rebalancing. Training parameters are kept the same across models/horizon.
LT vs the rest: impact on training

We compare the accuracy in training and test for each rebalancing. Training parameters are kept the same across models/horizon.
# Training model: quality measures

<table>
<thead>
<tr>
<th>Model</th>
<th>[train]</th>
<th>[test]</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1M</td>
<td>43.7%</td>
<td>47.6%</td>
</tr>
<tr>
<td>R3M</td>
<td>40.7%</td>
<td>48.0%</td>
</tr>
<tr>
<td>R6M</td>
<td>38.9%</td>
<td>47.7%</td>
</tr>
<tr>
<td>R9M</td>
<td>37.6%</td>
<td>46.5%</td>
</tr>
<tr>
<td>R12M</td>
<td>36.1%</td>
<td>46.4%</td>
</tr>
<tr>
<td>R18M</td>
<td>32.8%</td>
<td>46.2%</td>
</tr>
<tr>
<td>R24M</td>
<td>30.6%</td>
<td>42.8%</td>
</tr>
<tr>
<td>R36M</td>
<td>27.4%</td>
<td>40.3%</td>
</tr>
</tbody>
</table>
Interpretability breakdown – 1M preds.
Interpretability breakdown – 12M preds.
Interpretability breakdown – 36M preds.
Interpretability: simple avg feature importance
[LT ML predictions for EQ]
Analysing portfolios results
Decile performance’s analysis: monotonicity

<table>
<thead>
<tr>
<th></th>
<th>Avg annual net performance: net of TC gross of mc</th>
<th>R1M</th>
<th>R3M</th>
<th>R6M</th>
<th>R9M</th>
<th>R12M</th>
<th>R18M</th>
<th>R24M</th>
<th>R36M</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td></td>
<td>9.06%</td>
<td>11.98%</td>
<td>11.38%</td>
<td>11.77%</td>
<td>10.09%</td>
<td>10.60%</td>
<td>10.07%</td>
<td>9.61%</td>
</tr>
<tr>
<td>D2</td>
<td></td>
<td>8.64%</td>
<td>11.59%</td>
<td>11.67%</td>
<td>12.03%</td>
<td>12.53%</td>
<td>11.94%</td>
<td>12.33%</td>
<td>11.25%</td>
</tr>
<tr>
<td>D3</td>
<td></td>
<td>9.08%</td>
<td>10.28%</td>
<td>9.76%</td>
<td>12.39%</td>
<td>12.52%</td>
<td>12.38%</td>
<td>14.01%</td>
<td>12.57%</td>
</tr>
<tr>
<td>D4</td>
<td></td>
<td>9.89%</td>
<td>11.39%</td>
<td>10.37%</td>
<td>11.69%</td>
<td>13.40%</td>
<td>10.42%</td>
<td>14.73%</td>
<td>13.44%</td>
</tr>
<tr>
<td>D5</td>
<td></td>
<td>12.44%</td>
<td>12.61%</td>
<td>12.54%</td>
<td>12.39%</td>
<td>12.12%</td>
<td>13.71%</td>
<td>14.27%</td>
<td>13.49%</td>
</tr>
<tr>
<td>D6</td>
<td></td>
<td>11.73%</td>
<td>13.61%</td>
<td>13.68%</td>
<td>13.10%</td>
<td>11.90%</td>
<td>14.97%</td>
<td>16.25%</td>
<td>15.03%</td>
</tr>
<tr>
<td>D7</td>
<td></td>
<td>11.74%</td>
<td>13.58%</td>
<td>12.17%</td>
<td>13.93%</td>
<td>13.28%</td>
<td>15.00%</td>
<td>17.02%</td>
<td>15.19%</td>
</tr>
<tr>
<td>D8</td>
<td></td>
<td>11.61%</td>
<td>13.39%</td>
<td>13.10%</td>
<td>11.96%</td>
<td>15.41%</td>
<td>16.86%</td>
<td>19.33%</td>
<td>18.17%</td>
</tr>
<tr>
<td>D9</td>
<td></td>
<td>11.93%</td>
<td>15.30%</td>
<td>16.17%</td>
<td>16.39%</td>
<td>17.27%</td>
<td>17.89%</td>
<td>22.42%</td>
<td>21.39%</td>
</tr>
<tr>
<td>D10</td>
<td></td>
<td>13.20%</td>
<td>20.00%</td>
<td>20.28%</td>
<td>21.69%</td>
<td>23.47%</td>
<td>25.60%</td>
<td>27.20%</td>
<td>26.49%</td>
</tr>
</tbody>
</table>
Decile turnover’s analysis: look for the tails...

<table>
<thead>
<tr>
<th>avg monthly turnover (buy + sell)</th>
<th>R1M</th>
<th>R3M</th>
<th>R6M</th>
<th>R9M</th>
<th>R12M</th>
<th>R18M</th>
<th>R24M</th>
<th>R36M</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>63.7%</td>
<td>48.2%</td>
<td>43.7%</td>
<td>41.2%</td>
<td>39.6%</td>
<td>37.2%</td>
<td>35.5%</td>
<td>32.4%</td>
</tr>
<tr>
<td>D2</td>
<td>80.9%</td>
<td>71.4%</td>
<td>67.2%</td>
<td>64.6%</td>
<td>63.6%</td>
<td>59.7%</td>
<td>58.9%</td>
<td>55.1%</td>
</tr>
<tr>
<td>D3</td>
<td>84.8%</td>
<td>77.7%</td>
<td>74.0%</td>
<td>70.8%</td>
<td>70.2%</td>
<td>67.8%</td>
<td>67.8%</td>
<td>64.7%</td>
</tr>
<tr>
<td>D4</td>
<td>86.9%</td>
<td>80.7%</td>
<td>77.1%</td>
<td>73.9%</td>
<td>73.6%</td>
<td>72.1%</td>
<td>70.8%</td>
<td>68.4%</td>
</tr>
<tr>
<td>D5</td>
<td>86.9%</td>
<td>81.0%</td>
<td>78.5%</td>
<td>75.5%</td>
<td>75.1%</td>
<td>73.9%</td>
<td>72.0%</td>
<td>69.7%</td>
</tr>
<tr>
<td>D6</td>
<td>86.9%</td>
<td>81.6%</td>
<td>78.5%</td>
<td>75.4%</td>
<td>74.7%</td>
<td>73.3%</td>
<td>72.9%</td>
<td>69.3%</td>
</tr>
<tr>
<td>D7</td>
<td>86.0%</td>
<td>80.7%</td>
<td>77.0%</td>
<td>73.5%</td>
<td>72.7%</td>
<td>72.7%</td>
<td>71.8%</td>
<td>67.7%</td>
</tr>
<tr>
<td>D8</td>
<td>83.5%</td>
<td>78.5%</td>
<td>73.4%</td>
<td>70.0%</td>
<td>68.9%</td>
<td>69.0%</td>
<td>67.9%</td>
<td>64.8%</td>
</tr>
<tr>
<td>D9</td>
<td>80.3%</td>
<td>72.9%</td>
<td>67.6%</td>
<td>64.3%</td>
<td>63.2%</td>
<td>61.6%</td>
<td>60.3%</td>
<td>57.9%</td>
</tr>
<tr>
<td>D10</td>
<td>62.3%</td>
<td>53.1%</td>
<td>47.8%</td>
<td>45.5%</td>
<td>44.8%</td>
<td>42.1%</td>
<td>41.0%</td>
<td>39.2%</td>
</tr>
</tbody>
</table>
## Comparison accross portfolios

<table>
<thead>
<tr>
<th>from Feb 08 until Dec 19</th>
<th>Avg perf p.a. Net of tc (USD)</th>
<th>Vol p.a.</th>
<th>risk/perf ratio</th>
<th>Turnover avg monthly (B+S)</th>
<th>avg annual trading cost</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>D10 port R1M</strong></td>
<td>13.2%</td>
<td>12.34%</td>
<td>1.07</td>
<td>62%</td>
<td>1.87%</td>
</tr>
<tr>
<td><strong>D10 port R3M</strong></td>
<td>20.0%</td>
<td>16.51%</td>
<td>1.21</td>
<td>53%</td>
<td>1.59%</td>
</tr>
<tr>
<td><strong>D10 port R6M</strong></td>
<td>20.3%</td>
<td>17.72%</td>
<td>1.14</td>
<td>48%</td>
<td>1.43%</td>
</tr>
<tr>
<td><strong>D10 port R9M</strong></td>
<td>21.7%</td>
<td>18.68%</td>
<td>1.16</td>
<td>46%</td>
<td>1.37%</td>
</tr>
<tr>
<td><strong>D10 port R12M</strong></td>
<td>24.5%</td>
<td>18.58%</td>
<td>1.32</td>
<td>45%</td>
<td>1.34%</td>
</tr>
<tr>
<td><strong>D10 port R18M</strong></td>
<td>25.6%</td>
<td>19.17%</td>
<td><strong>1.34</strong></td>
<td>42%</td>
<td>1.26%</td>
</tr>
<tr>
<td><strong>D10 port R24M</strong></td>
<td><strong>27.1%</strong></td>
<td>20.44%</td>
<td>1.33</td>
<td>41%</td>
<td>1.23%</td>
</tr>
<tr>
<td><strong>D10 port R36M</strong></td>
<td>26.5%</td>
<td>20.70%</td>
<td>1.28</td>
<td><strong>39%</strong></td>
<td><strong>1.18%</strong></td>
</tr>
<tr>
<td><strong>Universe EW</strong></td>
<td>13.4%</td>
<td>12.03%</td>
<td>1.11</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td><strong>SP500</strong></td>
<td>9.8%</td>
<td>14.90%</td>
<td>0.66</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>
Conclusion

[1] Machine learning is not new but a “new” way for doing research today.

[2] ML used with traditional data proved to add a non-linear adaptative component to alpha prediction

[3] Long term predictions seems to give higher risk-adjusted performance with less turnover than the usual 1M forward horizon.
Important Information

This material is addressed to professional clients for informative purposes only. It is neither an offer nor an invitation to buy or sell investment products and may not be interpreted as investment advice. It is not intended to be distributed, published or used in a jurisdiction where such distribution, publication or use is forbidden, and is not intended for any person or entity to whom or to which it would be illegal to address such a material. In particular, investment products are not offered for sale in the United States or its territories and possessions, nor to any US person (citizens or residents of the United States of America). The opinions herein do not consider individual clients’ circumstances, objectives, or needs. Before entering into any transaction, clients are advised to form their own opinion and consult professional advisors to obtain an independent review of the specific risks incurred (tax, financial etc.). Upon request, RAM AI Group is available to provide more information to clients on risks associated with investments. The information and analysis contained herein are based on sources deemed reliable. However, RAM AI Group does not guarantee their accuracy, correctness or completeness, and it does not accept any liability for any loss or damage resulting from their use. All information and assessments are subject to change without notice. Changes in exchange rates may cause the NAV per share in the investor's base currency to fluctuate. There is no guarantee to get back the full amount invested. Past performances, whether actual or back-tested, are not necessarily indicative of future performance. Without prejudice of the due addressee’s own analysis, RAM understands that this communication should be regarded as a minor non-monetary benefit according to MIFID regulations. Clients are invited to base their investment decisions on the most recent prospectus, key investor information document (KIID) and financial reports which contain additional information relating to the investment product. These documents are available free of charge from the SICAV’s and Management Company’s registered offices, its representative and distributor in Switzerland, RAM Active Investments S.A. and at Macard Stein & Co AG, Paying and Information Agent in Germany; and at RAM Active Investments (Europe) SA – Succursale Milano in Italy. This marketing material has not been approved by any financial Authority, it is confidential and addressed solely to its intended recipient; its total or partial reproduction and distribution are prohibited. Issued in Switzerland by RAM Active Investments S.A. which is authorised and regulated in Switzerland by the Swiss Financial Market Supervisory Authority (FINMA). Issued in the European Union and the EEA by the Management Company RAM Active Investments (Europe) S.A., 51 av. John F. Kennedy L-1855 Luxembourg, Grand Duchy of Luxembourg. The reference to RAM AI Group includes both entities, RAM Active Investments S.A. and RAM Active Investments (Europe) S.A.
FDP Curriculum

1. Introduction to Data Science & Big Data

2. DM & ML: Introduction

3. DM & ML: Regression, LASSO, Predictive Models, Time Series & Tree Models

4. DM & ML: Classification & Clustering

5. DM & ML: Performance Evaluation, Backtesting & False Discoveries

6. DM & ML: Representing & Mining Text

7. Big Data, DM & ML: Ethical & Privacy Issues

8. Big Data and Machine Learning in the Financial Industry

Source: FDP Institute Study guide March 2020 Exam

Sample of the Reading(s):

- Topic 1 – Reading 1.4: Chapters 2, 4 & 5.
- Topic 8 - Reading 8.9 : Chapter 10.

Sample Keywords (of the Guida reading):
Mainstream (p. 336)  
Part of Speech Tagging (p. 349)  
Primary source (p. 336)  
Stemming (p. 350)  
Social media (p. 337)  
Lemmatization (p. 350)  
Sentiment analysis (p. 339)  
Naïve Bayes (p. 355)  
Natural language processing (p.347)  
FNN (p. 363)  
Tokenization (p. 348)  
RNN (p. 363)  
Word filter (p. 348)  
CNN (p. 363)
1. Introduction to Data Science & Big Data

2. DM & ML: Introduction

3. DM & ML: Regression, LASSO, Predictive Models, Time Series & Tree Models

4. DM & ML: Classification & Clustering

5. DM & ML: Performance Evaluation, Backtesting & False Discoveries

6. DM & ML: Representing & Mining Text

7. Big Data, DM & ML: Ethical & Privacy Issues

8. Big Data and Machine Learning in the Financial Industry

Sample Learning Objectives (provided for reading 8.9.1)

Demonstrate proficiency in the following areas:

8.9.1 Natural language processing of financial news. For example:
A. Describe the three categories of sources of news data.
B. Explain the advantages and disadvantages of using the new category of social media.
C. Describe sentiment analysis.
D. Describe the word list approach to sentiment analysis.
E. Describe the three challenges associated with sentiment analysis.
F. Describe the four steps — pre-processing, feature representation, inference and evaluation — in applying NLP to texts.
G. Understand aspects of pre-processing: tokenization, vocabulary, part of speech, stemming and lemmatization.
H. Understand aspects of representation of words as features: bag of words, N-gram, distributed representation

Sample Question:

According to “Natural Language Processing of Financial News,” by Sesen et al., what is the description of a “word list” approach to sentiment analysis?

a) Words appearing in an article are manually labeled as positive or negative
b) A data set that associates words with different sentiments is created
c) The predictive power of a news item is used to assign sentiment labels to words

Answer: b
Source: LO 8.9.1, Reading 8.9, pp 340-341
Kind reminders of upcoming webinars as we go through the Q & A. Add your questions in the chat room please.

Q & A

WEBINAR SERIES
A Conversation With...
Ganesh Mani
Adj. Faculty Carnegie Mellon
“Data Supply Chain Mgmt.”
March 10, 2020
1pm EST

WEBINAR SERIES
A Conversation With...
Rick Roche, CAIA
Man. Dir. Little Harbor Advisors
“Evolution of Machine Learning in Investment Mgmt.”
March 17 - 11am EST

WEBINAR SERIES
A Conversation With...
Michael Oliver Weinberg
Managing Dir., Head of Hedge Funds & Alt. Alpha APG
“Autonomous Learning Investment Strategies”
March 25th – 4pm EST

WEBINAR SERIES
A Conversation With...
George Mussalli, CFA
Mike Chen, Ph.D.
PanAgora
“An integrative Approach to Quantitative ESG Investing”
April 1, 2020 @ 12noon
In Closing

- Registration for the October 26 – November 8th exam opens May 10th
- For a recent candidate webinar go to www.fdpinstitute.org/webinars

Learn more about the FDP Institute at www.fdpinstitute.org