

MGMTMFE 431:

*Data Analytics and Machine Learning*

Topics 10:  
Final Thoughts

Spring 2019

Professor Lars A. Lochstoer

# Agenda

---

- a. Remaining group presentations
- b. Brief class review
- c. The Final Exam

*”What information consumes is rather obvious: it consumes the attention of its recipients. Hence, a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.”*

- Herbert Simon,  
Nobel Laureate in Economics



## d. Class ‘review’

---

This class serves as an **INTRODUCTION** to the many topics under the umbrella “Machine Learning and Data Analytics”

Not enough time to go through everything in detail

- But (I hope!) a great starting point for your *continued learning*
- Code snippets and homeworks give you a starting point, which can help you when you are faced with your own (more complicated) problems
- The packages I have used and references are well-known and we have only discussed some of what they are capable of – keep reading!

## d. The Course Outline

---

### 1) **Introductory Concepts**

- Examples of data analytics in finance
- Prediction and the value of models
- Data analytics and R, useful tools for data management
- Visualization as a model specification and communication tool
- ggplot2: a useful graphics package in R
- Forecasting signals based on data visualization

### 2) **Big data and panel regressions**

- Introduction to panel regressions
- Clustering, fixed effects
- Multiple regressions and marginal effects, omitted variables
- Fama-MacBeth (review): from signal to trading strategy

## d. The Course Outline

---

### **3) Logistic Regressions, Credit Data and Sample Selection**

- Models for discrete outcomes
- Default prediction and interest rates using loan/credit card databases
- Sample selection and propensity scores

### **4) Machine Learning I: Model selection**

- Bayesian shrinkage
- Lasso and ridge regressions
- Sparsity principle
- Cross-validation, best subset
- The data-mining problem in finance

## d. The Course Outline

---

### 5) **Unstructured data and textual analysis**

- Introduction to text analysis in R
- Mapping unstructured data to numeric signals
- Similarity metrics, investor inattention
- Text as Big Data: 21<sup>st</sup> century reading of financial reports (web-scraping, EDGAR database)

### 6) **Machine Learning II: Nonlinear methods**

- High-dimensional data; model selection and regularization revisited
- Classification: k-means clustering, asymptotic PCA
- Flexible prediction models: Decision trees, bagging (Random Forests), boosting (XGBoost)
- Support vector machines, neural networks (deep learning)

### **Student presentations**

## d. Major goal of class

---

- Ability to generate informative (meaning marginally; in addition to standard, existing) signals for:
  1. Alpha (return forecasting)
  2. Risk (variance and covariance forecasting)
  3. Any other decision that must be made in investment process
  
- Generate these signals from a wide variety of data sources
  1. Returns, of course
  2. Accounting data
  3. Text data
  4. Credit card data
  5. Any other data source



## d. Traditional vs. Machine Learning Techniques

---

- Machine learning has a current element of “hype”
  - Yep.
- We are still learning what works and what does not
- Don’t underestimate:
  1. Value of traditional methods of analysis (e.g., regressions) and economics
  2. Investor appetite for the “newest” topic (tech, private equity, big data) (yes, this may seem to contradict 1., but the key is the right balance)
  3. Value of having your current solid foundation, built in this class and Empirical Methods, in ability to quickly learn new methods
  4. Yourself. You guys are well-trained and really good!

## d. JP Morgan Report: Main Points

---

1. Banks will need to hire excellent data scientists who also understand how markets work
  2. Machines are best equipped to make trading decisions in the short and medium term
  3. An army of people will be need to acquire, clean, and assess the data
  4. There are different kinds of machine learning and they are used for different purposes
  5. Supervised learning will be used to make trend-based predictions using sample data
  6. Unsupervised learning will be used to identify relationships between a large number of variables
  7. Deep learning systems will undertake tasks that are hard for people to define but easy to perform
  8. Reinforcement learning will be used to choose a successive course of actions to maximize the final reward
  9. *You won't need to be a machine learning expert, you will need to be an excellent quant and an excellent programmer*
- **From report:** If you're only planning to learn one coding language related to machine learning, J.P. Morgan suggests you choose **R**

## d. Barclays' report on ML in Hedge Funds

---

Summary of main points:

- See: <https://news.efinancialcareers.com/us-en/287300/barclays-machine-learning-ai-and-big-data-jobs-in-hedge-funds>
1. Not all hedge funds apply these tools just yet (54% are investing in big data, 62% in machine learning)
  2. Finding and cleaning data is a major part of the job
  3. Data mining and 'black box' nature of machine learning is a big concern
  4. Move towards "quantamental" hedge funds. ("Quantamental" is a mix of quant strategies and classic fundamental analysis). Need data specialists that can communicate efficiently with more classically trained analysts and portfolio managers

## e. The Final Exam

---

- Monday, June 10: 11:30am – 2:30pm, B301
- The format will be like last year's exam, posted on CCLE
- Bring calculator, no computer
- Bring 1, letter-size, cheat sheet. You may write on both sides.
- Questions?

## f. Concluding...

---

- It's been a real pleasure having you in both Empirical Methods and this class
- Keep learning!
- Please, do stay in touch. We need to maintain a strong UCLA MFE network and I want to help with that.

**I wish you the best of luck!**