

Course Overview

Itamar Caspi

November 25, 2019 (updated: 2019-11-26)

10-Year Challenge

2009: ML = Maximum Likelihood

2019: ML = Machine Learning

An Aside: About the Structure of These Slides

- This slide deck was created using the R package **xaringan** (/ʃæ.'rɪŋ.gæn/) and **Rmarkdown**.
- Some slides include hidden comments. To view them, press **p** on your keyboard

About this presentation

- This slide deck was created using the R package **xaringan** (/ʃæ.'rɪŋ.gæn/) and **Rmarkdown**.
- Some slides include hidden comments
- To view them, press **p** on your keyboard

Outline

1. Logistics
2. About the Course
3. To Do List

Logistics

Discussion Forum

- Google and [StackExchange](#) are typically the first places to look for answers for computational related questions. It is safe to always assume that you are not the first one to encounter your problem.
- We will also use the BoI's "Success Factors" forum. Yael Dreyfus will provide more details later.

People

Lectures:

- **Ariel Mansura**, Head of Statistical Methodology Unit, Information and Statistics Department.
- **Itamar Caspi**, Head of Monetary Analysis Unit, Research Department.

Practical sessions:

- **Jonathan Rosen**, Statistical Methodology Unit, Information and Statistics Department.
- Meeting hours: after class, on demand.

Feedback

This is the second time we run this course. Nevertheless, your continuous feedback is important!

Please contact us by

- email
 - in person
 - open an issue on our forum (Best option)
-

About the Course

Prerequisites

- Advanced course in econometrics / statistics.
- Some experience with R (or another programming language) are a major plus.

This Course Is

About

Tools of trade: R, Tidyverse, Git...

How and when to apply ML methods

- Generate good quality predictions
- Classify data objects
- Organize unstructured data

To do that we will need to understand

- what is ML?
- How it relates to stuff we already know?
- How it differs?

Not about

- Cutting-edge ML techniques (e.g., deep learning)
- Computational aspects (e.g., gradient descent)
- Data wrangling (a.k.a. "feature engineering")
- Distributed file systems (e.g., Spark)

Schedule

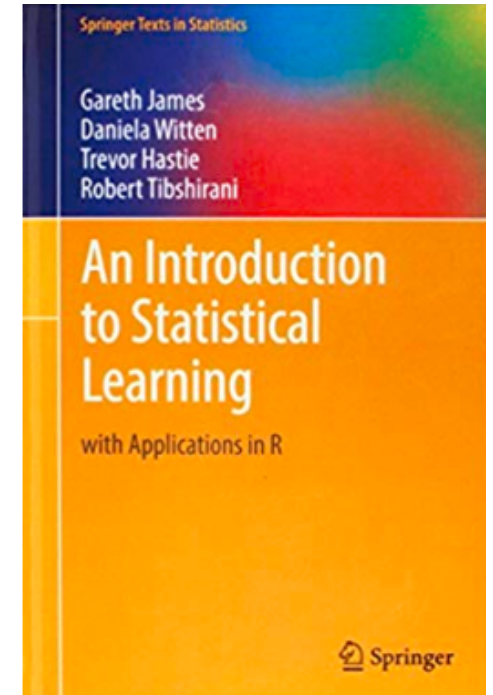
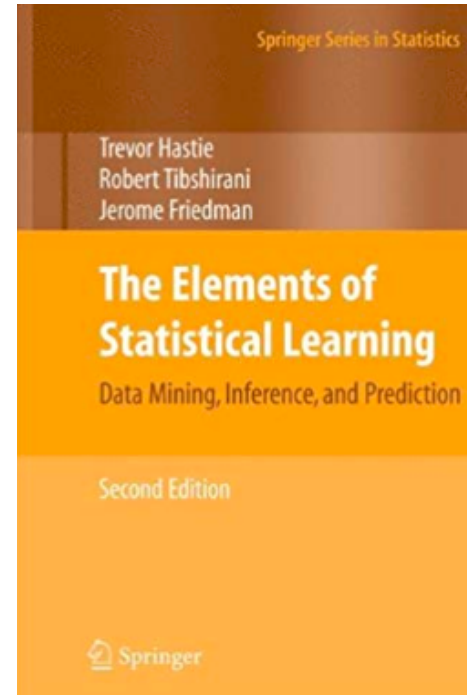
Week	Who	Topic
[1]	Itamar	R, RStudio, git, and GitHub
[2]	Itamar	Basic Concepts
[3]	Ariel	Regression and K Nearest Neighbors
[4]	Ariel	Regularization
[5]	Ariel	Classification
[6]	Ariel	Decision Trees and Random Forests
[7]	Ariel	Unsupervised Learning
[8]	Itamar	Text as Data

NOTE: This schedule might go through changes.

Readings on ML

A couple of suggestions:

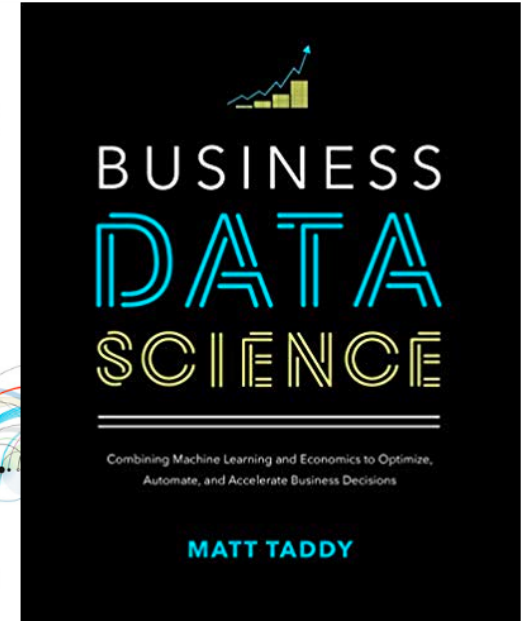
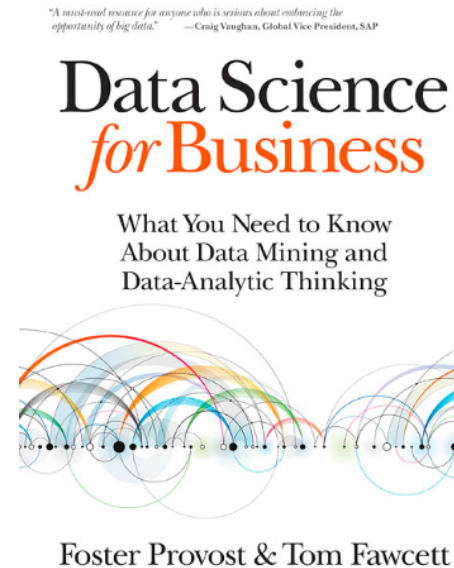
- **An Introduction to Statistical Learning with Applications in R (ISLR)**
James, Hastie, Witten, et al. (2013)
PDF available online
- **The Elements of Statistical Learning (ELS)**
Hastie, Tibshirani, and Friedman (2009)
PDF available online



Readings on Applied ML

A couple of suggestions:

- **Business Data Science: Combining Machine Learning and Economics to Optimize, Automate, and Accelerate Business Decisions**
Taddy (2019)
- **Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking**
Provost and Fawcett (2013)



Readings on ML and Economics

- **Machine Learning: An Applied Econometric Approach** Mullainathan and Spiess (2017) *Journal of Economic Perspectives*, 31(2), 87-106.
- **Machine Learning Methods Economists Should Know About** Athey and Imbens (2019) *Annual Review of Economics*, 11, 625-725.



More Resources

The (unofficial) class website: <https://ml4econ.github.io/course-spring2019>




The screenshot shows the homepage of the ml4econ website. The top navigation bar is blue with white text for '#ml4econ', 'People', 'Lectures', 'Resources', 'News', 'FAQ', and 'Guides'. On the left, a sidebar menu lists 'Course summary' (highlighted), 'Computing', 'Prerequisite', 'Evaluation', 'Discussions', and 'Feedback'. The main content area features the title '55750: Machine Learning for Economists' and the subtitle 'Hebrew University of Jerusalem - Spring 2019'. Below this is the ml4econ logo, which consists of a blue hexagon containing two stylized trees and the text 'ml4econ'. A welcome message follows: 'Welcome to ml4econ, the Machine Learning for Economists course repository! This is a graduate level course taught by Itamar Caspi and Ariel Mansura at the Hebrew University of Jerusalem.' The 'Course summary' section begins with a paragraph about the course's focus on machine learning and econometrics, followed by two paragraphs detailing the course structure and topics.

#ml4econ People Lectures Resources News FAQ Guides

Course summary
Computing
Prerequisite
Evaluation
Discussions
Feedback

55750: Machine Learning for Economists

Hebrew University of Jerusalem - Spring 2019



Welcome to **ml4econ**, the **M**achine **L**earning for **E**conomists course repository! This is a graduate level course taught by [Itamar Caspi](#) and Ariel Mansura at the Hebrew University of Jerusalem.



Course summary

The course will cover topics that range between machine learning (ML) and econometrics. In particular, we will discuss concepts from the world of ML that can potentially contribute to empirical economics. The course will cover leading machine learning (supervised and unsupervised) methods, with an emphasis on the challenges and opportunities of integrating these methods in empirical economics, and the relevance of ML to policy analysis and causal inference. The various topics are illustrated through applications, reading empirical articles, and doing applied work.

The first, part of the course will introduce you to Supervised (predictive) and Unsupervised Machine Learning methods. Examples include Regression and K Nearest Neighbors, Classification, Dimensionality Reduction, Decision Trees and Random Forests, Principal Component Analysis and Clustering Analysis. You will learn what techniques to apply and why.

The second and biggest part of the class, will focus on the relationship between Machine Learning and Econometrics. You will

Programming Language

- Two of the most popular open-source programming languages for data science:
 - 
 -  Python
- We choose R.
- Why R? A matter of taste...
- We do encourage you to try out Python. However, we will be able to provide limited support for Python users.
- (Pro tip: You can use both! see the [reticulate](#) package.)

RStudio Cloud



[Log In](#)

[Sign Up](#)




Welcome to RStudio Cloud^{alpha}

Do, share, teach and learn data science with R.







[Get Started](#)

If you already have an RStudio shinyapps.io account, you can log in using your existing credentials.


RStudio Projects

☰ Bol Data Science Course 2019 / mastering-the-tidyverse ⚙️ ⋮  Itamar Caspi

File Edit Code View Plots Session Build Debug Profile Tools Help






     Go to file/function  Addins R 3.6.0


Console Terminal x Jobs x

/cloud/project/ 

```
> |
```







Environment History Connections


   Import Dataset  List 


Global Environment 

Environment is empty

Files Plots Packages Help Viewer

 New Folder  Upload  Delete  Rename  More 





 Cloud > project ...


	Name	Size	Modified
	..		
<input type="checkbox"/>	01-Introduction		
<input type="checkbox"/>	02-Visualize		
<input type="checkbox"/>	03-Transform		
<input type="checkbox"/>	04-Model		
<input type="checkbox"/>	05-Report		
<input type="checkbox"/>	install-instructions.md	1013 B	Jul 31, 2019, 9:05 PM
<input type="checkbox"/>	LICENSE.txt	19.7 KB	Jul 26, 2019, 6:19 AM

Hands-on

- DIY: [RStudio Primers](#), [DataCamp](#), etc.
- In-class code sessions
- [Kaggle](#) prediction competition: predict median house value in Boston area.

Kaggle


 Search  Competitions Datasets Kernels Discussion Learn ...  


 InClass Prediction Competition

55750: Machine Learning for Economists @ HUJI 2019

A prediction competition for course participants

[Host](#) [Overview](#) [Data](#) [Kernels](#) [Leaderboard](#) [Rules](#) [Team](#) [My Submissions](#)

 This competition hasn't been launched. Only hosts and Kaggle admins can see it.

Overview 

[Description](#)
[Evaluation](#)
[+ Add Page](#)

In this competition, course participants will rely on the "Boston Housing Data" to train and test machine learning models learned in the course. In particular, course participants are required to apply the tools introduced in the course in order to predict Boston area **median house values** based on a set of area specific features.

To Do List

Officework

Cloud:

- ✓ Create an **RStudio Cloud** account and send your username (email) to Itamar.
- ✓ Create an account on **Kaggle** and ask Itamar to invite you the course's **prediction competition**.

Office:

- ✓ Download and install **Git**.
 - ✓ Download and install **R** and **RStudio**.
 - ✓ Create an account on **GitHub**
 - ✓ Download and install **GitHub Desktop**.
-

```
slides %>% end()
```

 [Source code](#)

References

- [1] S. Athey. "The impact of machine learning on economics". In: *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press, 2018.
- [2] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer, 2009. פבר. ISBN: 9780387848570.
- [3] G. James, T. Hastie, D. Witten, et al. *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics. Springer London, Limited, 2013. ISBN: 9781461471370.
- [4] S. Mullainathan and J. Spiess. "Machine learning: an applied econometric approach". In: *Journal of Economic Perspectives* 31.2 (2017), pp. 87-106.