

# Course Overview

Itamar Caspi

March 10, 2019 (updated: 2019-03-10)

# 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.'riŋ.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.'riŋ.gæn/) and **Rmarkdown**.
- Some slides include hidden comments
- To view them, press **p** on your keyboard

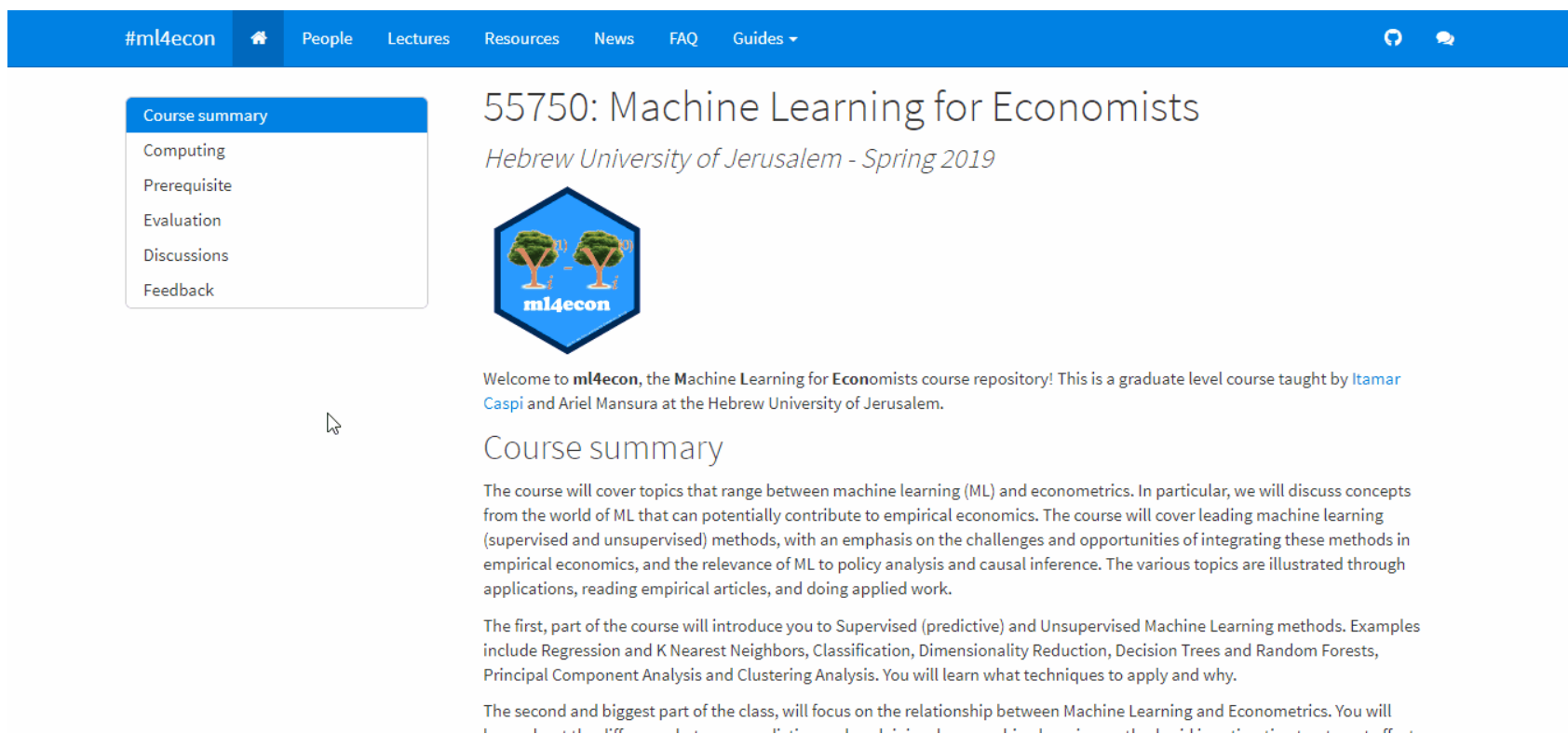
# Outline

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

# Logistics

# Class website

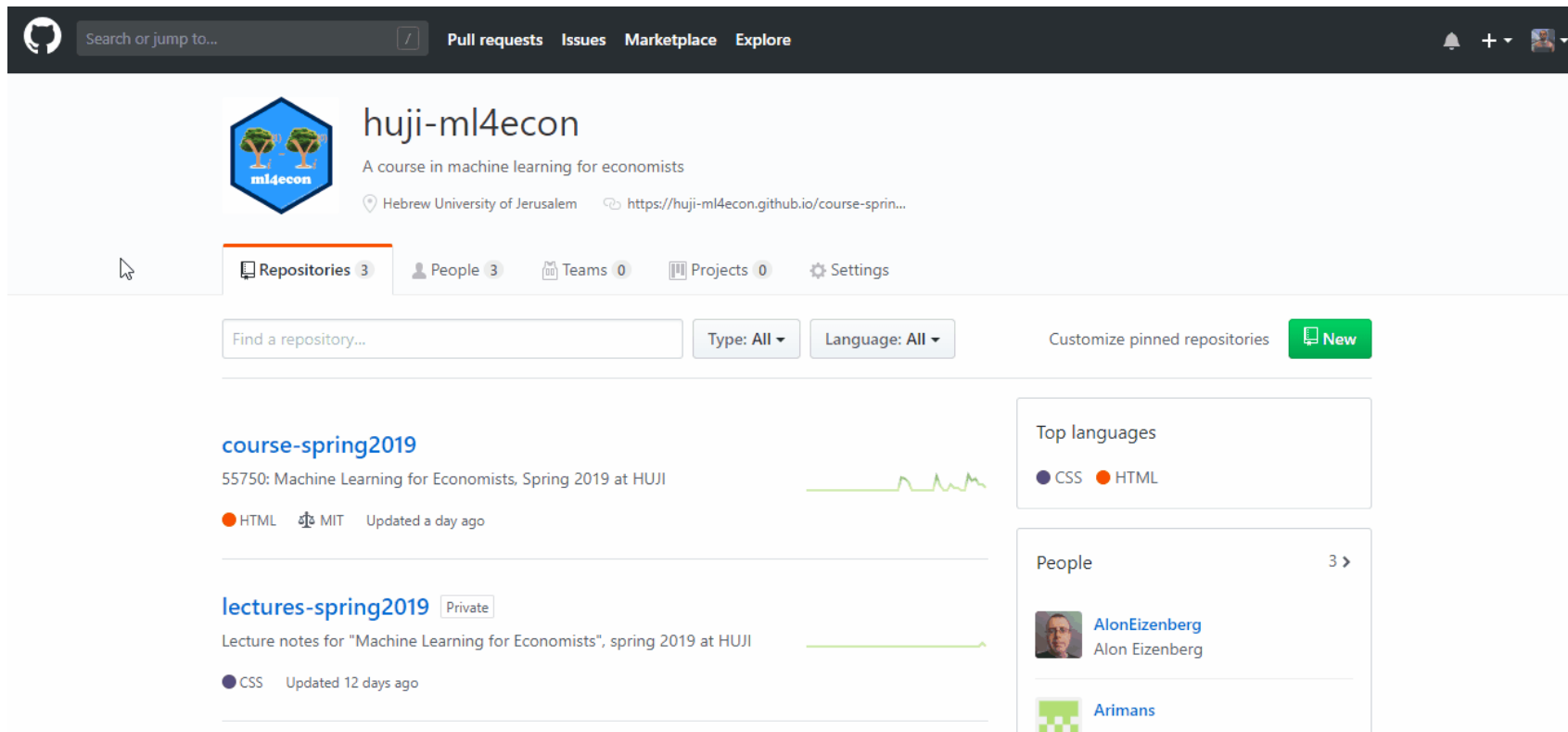
The (unofficial) class website: <https://ml4econ.github.io/course-spring2019>



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, stating: '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: '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.' This is followed by two more paragraphs: '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.' and 'The second and biggest part of the class, will focus on the relationship between Machine Learning and Econometrics. You will'.

# Discussion forum

We will use a **GitHub discussion repository**. To use it, you'll need to create a GitHub account and ask for an invitations from Itamar.



The screenshot shows the GitHub profile page for the organization **huji-ml4econ**. The header includes the GitHub logo, a search bar, and navigation links for Pull requests, Issues, Marketplace, and Explore. The profile section features the organization's logo (a blue hexagon with two trees and the text 'ml4econ'), the name 'huji-ml4econ', and a description: 'A course in machine learning for economists'. It also lists the location 'Hebrew University of Jerusalem' and a website link 'https://huji-ml4econ.github.io/course-sprin...'. Below the profile information, there are tabs for Repositories (3), People (3), Teams (0), Projects (0), and Settings. A search bar for repositories is present, along with filters for Type (All) and Language (All). The main content area displays two repositories: 'course-spring2019' and 'lectures-spring2019'. The 'course-spring2019' repository is described as '55750: Machine Learning for Economists, Spring 2019 at HUJI' and is tagged with HTML and MIT. The 'lectures-spring2019' repository is described as 'Lecture notes for "Machine Learning for Economists", spring 2019 at HUJI' and is tagged with CSS. On the right side, there are sections for 'Top languages' (CSS and HTML) and 'People' (AlonEizenberg and Arimans).

Search or jump to... Pull requests Issues Marketplace Explore

**huji-ml4econ**  
A course in machine learning for economists  
Hebrew University of Jerusalem <https://huji-ml4econ.github.io/course-sprin...>

Repositories 3 People 3 Teams 0 Projects 0 Settings

Find a repository... Type: All Language: All Customize pinned repositories New

**course-spring2019**  
55750: Machine Learning for Economists, Spring 2019 at HUJI  
HTML MIT Updated a day ago

**lectures-spring2019** Private  
Lecture notes for "Machine Learning for Economists", spring 2019 at HUJI  
CSS Updated 12 days ago

Top languages  
CSS HTML

People 3 >  
AlonEizenberg  
Alon Eizenberg  
Arimans

# People

- **Itamar Caspi**

- Head of Monetary Analysis Unit, Research Department, Bank of Israel.
- email: [caspi.itamar@gmail.com](mailto:caspi.itamar@gmail.com)
- homepage and blog: <https://itamarcaspi.rbind.io/>

- **Ariel Mansura**

- Head of Statistical Methodology Unit, Information and Statistics Department, Bank of Israel.
- email: [ariel.mansura@boi.org.il](mailto:ariel.mansura@boi.org.il)

- Meeting hours: after class, on demand.



# Feedback

This is the first time we run this course  $\Rightarrow$  your continuous feedback is important!

Please contact us by

- email
- in person
- or open an issue in our discussion forum

# About the Course

# Prerequisites

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

# This course is

## About

How and when to apply ML methods in economics

- estimate treatment effects
- prediction policy
- generate new 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., Hadoop, Spark)

# Syllabus

Week	Who	Topic
1	Itamar	Course Overview
2	Itamar	R, Rstudio, git, and GitHub
3	Ariel	Regression and K Nearest Neighbors
4	Ariel	Classification
5	Ariel	Dimensionality Reduction
6	Ariel	Decision Trees and Random Forests
7	Ariel	Unsupervised Learning
8	Itamar	Prediction in Aid of Estimation I
9	Itamar	Prediction in Aid of Estimation II
10	Itamar	Prediction Policy Problems
11	Itamar	Text as Data
12	Itamar	TBA

**NOTE:** This schedule can (and probably will) go through changes!

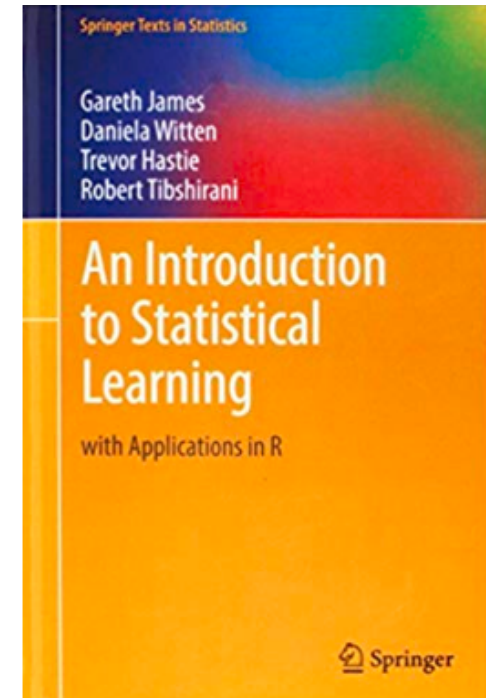
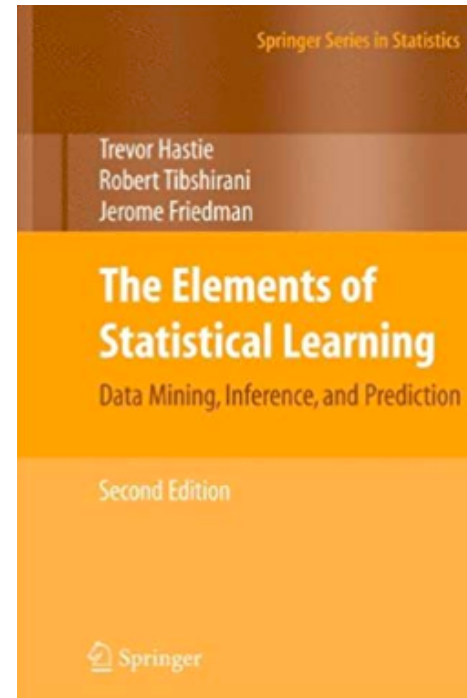
# Readings on ML

All materials and lecture notes will be available on the [class website](#).

There are **no** required textbooks.

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 ML for economists



All materials and lecture notes will be available on the [class website](#).

Read these excellent surveys:

- **The impact of machine learning on economics** Athey (2018)  
*In The Economics of Artificial Intelligence: An Agenda.*  
University of Chicago Press.
- **Machine learning: an applied econometric approach** Mullainathan and Spiess (2017)  
*Journal of Economic Perspectives*, 31(2), 87-106.




# Programming language

- Two of the most popular open-source programming languages for data science:
  - 
  -  Python
- Our recommendation: R.
- Why R? See presentation notes and the [FAQ section](#) of our class website.
- We do encourage you to try out Python. However, we will be able to provide limited support for Python users.



# DataCamp in the classroom

 DataCamp

Course Outline →

2

Exercise

## How it works

In the editor on the right you should type R code to solve the exercises. When you hit the 'Submit Answer' button, every line of code is interpreted and executed by R and you get a message whether or not your code was correct. The output of your R code is shown in the console in the lower right corner.


R makes use of the `#` sign to add comments, so that you and others can understand what the R code is about. Just like Twitter! Comments are not run as R code, so they will not influence your result. For example, *Calculate 3 + 4* in the editor on the right is a comment.

You can also execute R commands straight in the console. This is a good way to experiment with R code, as your submission is not checked for correctness.

Instructions 100 XP

script.R

```
1 # Calculate 3 + 4
2 3 + 4
3
4 # Calculate 6 + 12
5
```

 Run Code Submit Answer

R Console

> |

# Grading

## Assignments:





- **DataCamp Classroom**: you will be assigned with specific courses that will teach you essential R programming skills.
- Problem sets.


## Projects:

- **Kaggle** prediction competition: predict median house value in Boston area.
- Conduct a replication study based on one of the datasets included in the **experimentdata** package.

**GRADING:** Assignments **20%**, project **40%**, final exam **40%**.

# 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.

# experimentdatar

We will also make use of the `experimentdatar` data package that contains publicly available datasets that were used in Susan Athey and Guido Imbens' course "[Machine Learning and Econometrics](#)" (AEA continuing Education, 2018).

- You can install the **development** version from [GitHub](#)

```
# install.packages("devtools")  
devtools::install_github("itamarcaspi/experimentdatar")
```

- **EXAMPLE:** Load the `experimentdatar` package and the `social` dataset:

```
library(experimentdatar)  
data(social)
```

- Tips:
  1. Running `?social` provides variable definitions.
  2. Running `dataDetails("social")` will open a link to the paper associated with `social`.

# To Do List

# Homework\*

- ✓ Download and install [Git](#).
- ✓ Download and install [R](#) and [RStudio](#).
- ✓ Create an account on [GitHub](#)
- ✓ Download and install [GitHub Desktop](#).
- ✓ Create an account on [DataCamp](#) and ask Itamar to invite you to [DataCamp Classroom](#).
- ✓ Create an account on [Kaggle](#) and ask Itamar to invite you the course's [prediction competition](#).

[\*] Please consult the [Guides](#) section in our course's website.

```
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, Feb. 2009. 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.