01 - Course Overview

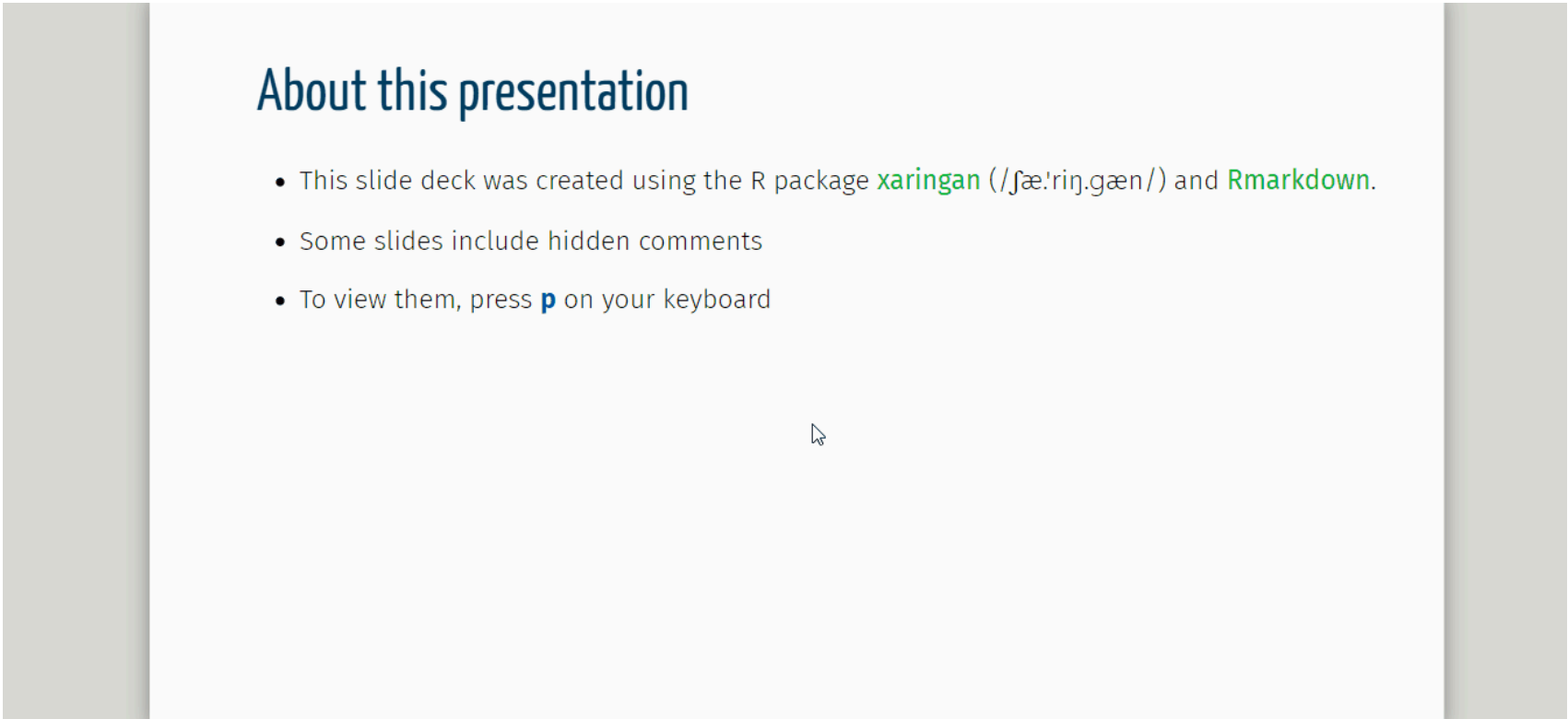
ml4econ, HUJI 2025

Itamar Caspi

March 23, 2025 (updated: 2025-03-23)

An aside: about the structure of these slides

- The course's slide decks are created using the xaringan (/ʃæ.'riŋ.gæn/) R package and Rmarkdown.
- Some slides include hidden comments. To view them, press **p** on your keyboard



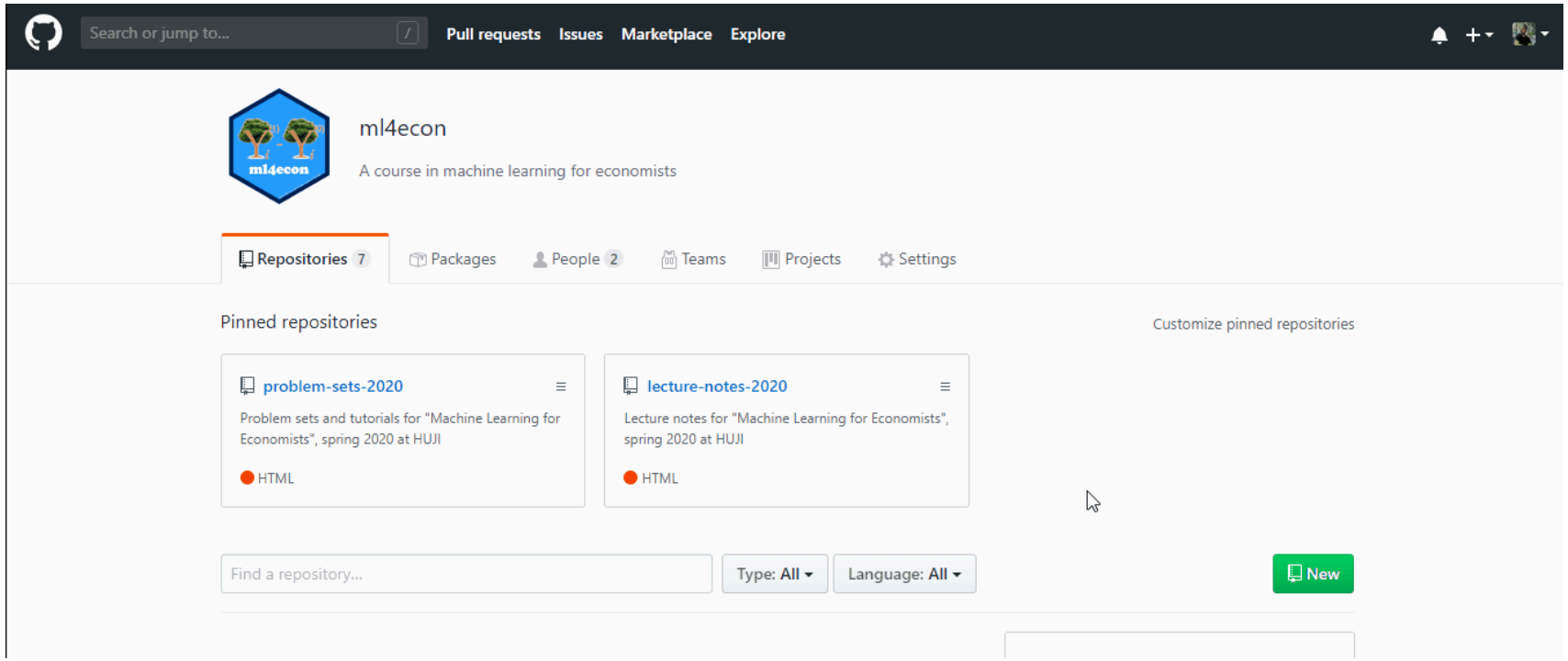
Outline

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

Logistics

ml4econ GitHub repository

The class's GitHub repository: https://github.com/ml4econ



Posit Cloud workspace

Posit Cloud is a hosted version of RStudio in the cloud that will make it easy for R and Python novices to learn data science and machine learning using R and Python.

People

- **Itamar Caspi**
  - email: caspi.itamar@gmail.com
  - homepage: itamarcaspi.rbind.io
- **Inbar Avni (TA)**
  - email: inbar.avni@mail.huji.ac.il
- Meeting hours: after class/zoom, on demand.

Feedback

Your continuous feedback is important!

Please feel free to contact us by

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

About the Course

Prerequisites

Econometrics

- Linear regression models and OLS estimation
- Statistical inference and hypothesis testing
- Endogeneity issues and instrumental variables

Mathematical Foundation

- Derivatives, partial derivatives, and basic optimization
- Matrices, vectors, eigenvalues/eigenvectors
- Random variables, probability distributions,expectations, Sampling distributions

Programming

- Some experience with R (or another programming language) is a plus
- Basic data manipulation and visualization skills
- Familiarity with statistical software packages

Learning Objectives

By the end of this course, students will be able to:

- Apply machine learning methods to economic research problems, including prediction policy tasks and causal inference questions
- Implement and evaluate ML workflows with proper cross-validation, regularization, and model selection techniques
- Interpret ML results in economic contexts and understand when ML approaches complement or improve upon traditional econometric methods

What This Course Is NOT About

This course deliberately excludes:

- Technical implementation details of advanced ML algorithms (we focus on application rather than algorithmic development)
- Deep theoretical foundations of ML methods (we emphasize practical understanding over mathematical proofs)
- Cutting-edge techniques like generative AI, reinforcement learning, or computer vision (we concentrate on established methods with proven economic applications)

Tentative schedule

Week	Topic
1	Course Overview & ML Basics
2	Reproducibility and ML Workflow
3	Regression and Regularization
4	Classification
5	Non-parametrics
6	Unsupervised Learning
7	Text analysis
8	Causal Inference
9	Lasso and Average Treatment Effects
10	Trees and Heterogeneous Treatment Effects
11	Text as Data
11	Deep Learning
12	Large Language Models

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

Where to Start?

- Big Data: New Tricks for Econometrics by Varian (2014) *Journal of Economic Perspectives*, 28(2), 3-28.
- The impact of machine learning on economics by Athey (2018)
  In *The Economics of Artificial Intelligence: An Agenda*.
  University of Chicago Press.
- Machine learning: an applied econometric approach by Mullainathan and Spiess (2017) *Journal of Economic Perspectives*, 31(2), 87-106.
- Machine Learning Methods That Economists Should Know About by Athey and Imbens (2019) *Annual Review of Economics*, 11(1), 685-725.
- Deep Learning for Economists by Dell (2025) *Journal of Economic Literature*, 60(1), 3-63.

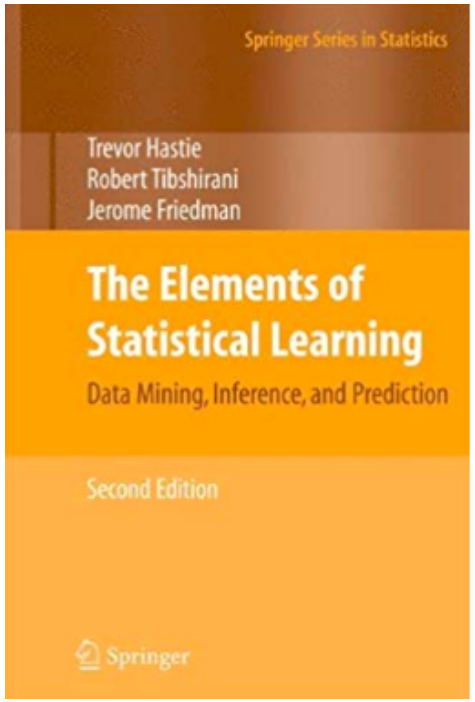
Readings on ML

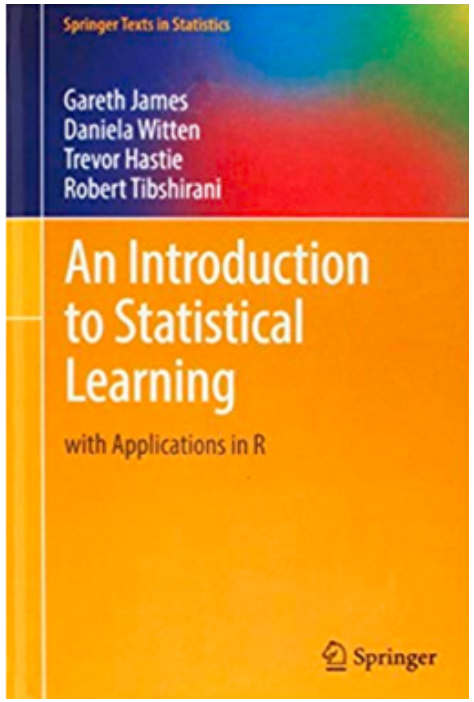
> All materials and lecture notes will be available on the course repo.

There are **no** required textbooks. A couple of suggestions:

- An Introduction to Statistical Learning with Applications in R/Python (ISLR), 2 ed.
  James et al. (2023) **PDF available online**

- The Elements of Statistical Learning (ELS)
  Hastie, Tibshirani, and Friedman (2009)
  PDF available online





Readings on ML and Econometrics

> All materials and lecture notes will be available on the course repo.

There are **no** required textbooks.

A couple of suggestions:

- Applied Causal Inference Powered by ML and AI, by Chernozhukov et al. (2024) **PDF available online**
- Econometrics, Ch. 29, by Hansen (2023) **PDF available online**
- Bussiness Data Science by Taddy (2019) *Available on Amazon*

More resources

Can be found at our GitHub repo:

https://github.com/ml4econ/lecture-notes-2025/blob/master/resources.md

Programming

- Two of the most popular open-source programming languages for data science:
  - R
  - Python
- This course: Mostly R. Why?
  - Widely used in academia.
  - Great for data manipulation and visualization.
  - Excellent for econometrics.
  - Great for machine learning.
  - Great for reproducibility.
- We do encourage you to try out Python. However, I will only be able to provide limited support for Python users. Inbar on the other hand, will be able to provide more support.

Catching up with R

If you are new to R, here are some resources to get you started:

- R for Data Science is a great book to learn R and data science.
- Posit Recipes is a great (old school) resource to learn R.
- DataCamp is a great resource to learn R (and Python).

GenAI

We encourage you to use AI (Claude, ChatGPT, Gemini, etc.) in this course, as it is an **essential skill to acquire**.

It is important you understand the (current) limitations of AI and its implications for economic research:

- Prompt engineering is necessary for quality outcomes.
- Always assume that it is wrong.
- Acknowledge its use in assignments and explain what prompts were used.

A useful resources:

- Read "Generative AI for Economic Research: Use Cases and Implications for Economists" by Korinek (2023 JEL). Check for the latest version.

Grading

Assignments:

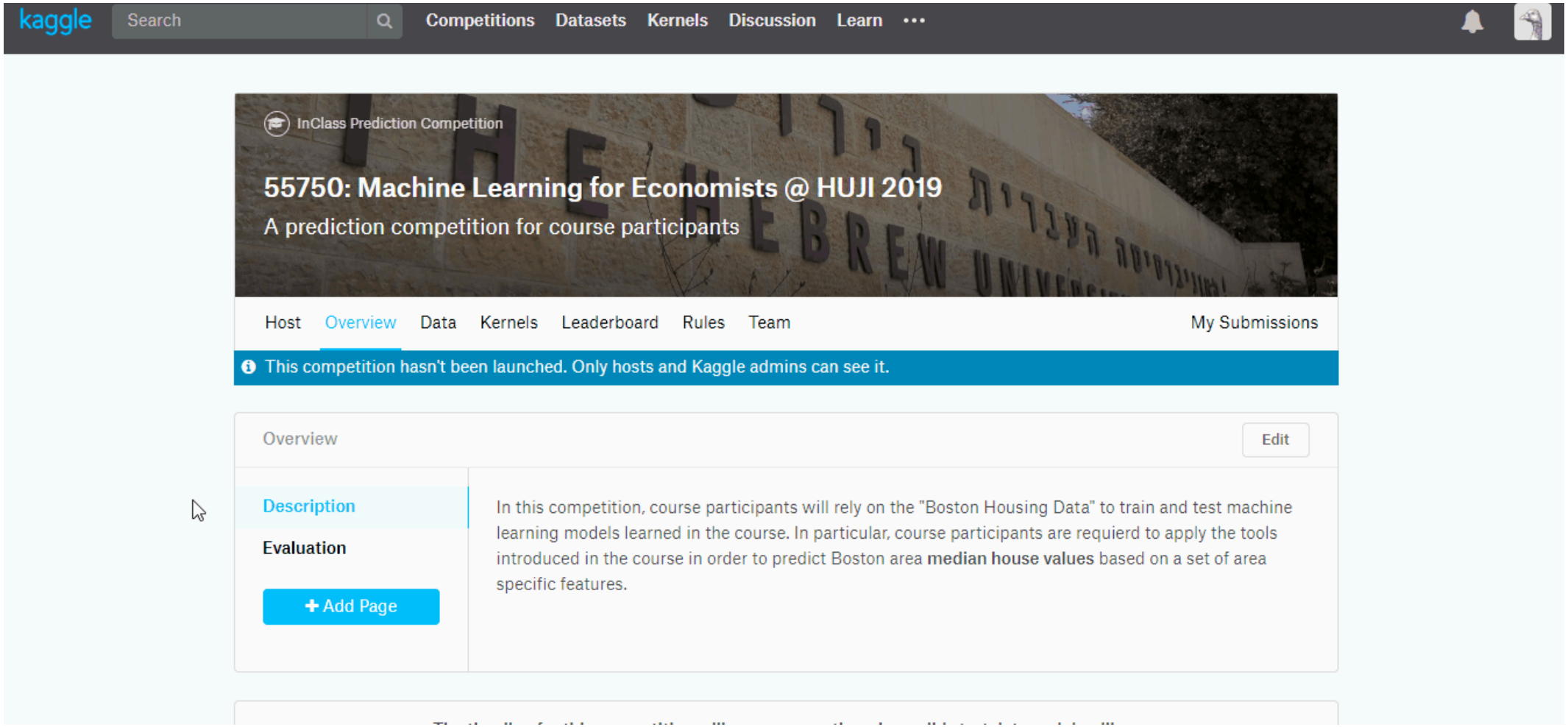
- Submit 4 out of a total of 6 Problem sets.

Two projects:

- Kaggle prediction competition.
- Conduct a replication study based on one of the datasets included in the experimentdatar package, or a paper of your choice.

GRADING: Assingments **20%**, kaggle **30%**, project **50%**.

Kaggle


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
Edit
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 requierd 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

Homework

☑ Download and install Git.

- Git is a version control system that will help you keep track of your code and collaborate with others.

☑ Download and install R and RStudio.

- R is a programming language and environment for statistical computing and graphics. RStudio is an integrated development environment (IDE) for R.

☑ Create an account on GitHub

- GitHub is a platform for hosting and collaborating on code.

☑ Download and install GitHub Desktop.

- GitHub Desktop is a graphical user interface for GitHub.

```
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.