

Course Overview

Itamar Caspi

ml4e @ HUJI, 2018

April 26, 2018

LOGISTICS

Course Google Drive:

`https:`

`//drive.google.com/open?id=1IiZroSBQOV4I6bYafVehLrvXZ75vICXc`

- ▶ Syllabus
- ▶ Slides
- ▶ Textbooks and background papers
- ▶ R tutorial and code
- ▶ Assignment

ASSIGNMENT

- ▶ Your course assignment is a Kaggle competition:
<https://www.kaggle.com/t/38a084622e714eee98b70c7574781060>
- ▶ where you will be required to predict the median value of a house in Boston districts using one or more of the machine learning algorithms you will learn in class

PEOPLE

Instructors involved in providing this course:

- ▶ Ariel Mansura
- ▶ Shir Kamenetsky
- ▶ Itamar Caspi
- ▶ Igor Rochlin (GSTAT, igorochlin@gmail.com)

PREREQUISITES

- ▶ Advanced level (M.A.) course in Statistics/Econometrics
- ▶ Working knowledge of R

RESOURCES

Main texts:

- ▶ *An Introduction to Statistical Learning, with Applications in R* (ISLR) by James, Witten, Hastie and Tibshirani (Springer, 2013)
- ▶ *The Elements of Statistical Learning - Data Mining, Inference, and prediction* (ESL) by Friedman, Tibshirani, and Hastie (Springer, 2008)

COURSE GOALS

What this course is about

- ▶ learn tools that will enable you to work with big data of the type you are familiar with
- ▶ learn how to implement these tools using R and
 - ▶ produce quality prediction
 - ▶ classify unstructured data

What this course is NOT about

- ▶ state-of-the-art machine learning methods (gradient boosting, deep learning, etc.)
- ▶ causal inference
- ▶ efficient computation
- ▶ data querying (SQL)
- ▶ complex data structures

SYLLABUS

Here is a list of the main topics we intend to cover during the course:

- ▶ Basic concepts (Itamar)
- ▶ Regression and K -nearest neighbors (Ariel)
- ▶ Classification methods (Igor)
- ▶ Support vector machines (SVM) (Igor)
- ▶ Neural networks (Igor)
- ▶ LASSO, Ridge and principal component regression (Ariel)
- ▶ Unsupervised learning (Igor)

Bonus lecture: **"Machine learning applications in the Israeli fintech industry"**
(tentative title) by Ido Mintz, lead data scientist @ Intuit.

FEEDBACK

Your feedback is important! Please feel free to share with us your comments, concerns and suggestions on the course in person, email or anonymously here:

<http://www.admonymous.com/boibigdata>