MACHINE LEARNING IN ECONOMICS
(Thursdays, 10.00-12.50)
(2021-22 Spring Semester)

Syllabus

Instructor: Prof. Dr. Hüseyin Taştan
Email: tastan@yildiz.edu.tr
Classnotes: https://github.com/htastan/Machine-Learning-in-Economics
Office: Davutpaşa Campus, IIBF/G2-205
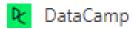**Office Hours**: Tuesdays 14:00-16:30 or by appointment (please send an email)

SCOPE and PURPOSE

The purpose of this course is to teach basic machine learning algorithms and methods relevant for empirical economists. The availability of large scale data sets has led to the development of new methods that are similar to those utilized in econometrics but also distinct in some aspects. This course, therefore, will emphasize the use of these new algorithms by focusing on statistical learning in economics and business. Students will learn the basic concepts, methods, and algorithms used in machine learning and develop skills to apply them in practice.

Econometrics software: We will use R in class and in lab sessions. R is an open-source software for statistical computing and graphics which is widely used by statisticians, researchers, data scientists and econometricians as well as industry professionals. The latest version of R can be downloaded from:
https://www.r-project.org/
And R-studio may be used as an integrated development environment for R:
https://www.rstudio.com/products/RStudio/

This class is supported by www.datacamp.com

PREREQUISITES
- You need to pass Statistics for Economists II (or equivalent). Prior knowledge of econometrics is not necessary but may be beneficial.

TEXTBOOK
- **ISLR**: James, G., D. Witten, T. Hastie, R. Tibshirani (2021), *An Introduction to Statistical Learning with Applications in R*, 2nd ed., Springer. Available at: https://www.statlearning.com/

- Alpaydın, Ethem (2016), *Introduction to Machine Learning*, 3rd ed., MIT Press, Cambridge MA.
  Here is the Turkish translation of this book:
- Alpaydın, Ethem (2018), *Yapay Öğrenme*, 4. Baskı (Ethem Alpaydın, Introduction to Machine Learning, 2. baskıdan çeviri), Boğaziçi Üniversitesi Yayınevi, İstanbul.

EVALUATION
Class attendance and participation: 10%, Homeworks and assignments: 30% , Midterm: 20%
Project and presentation: 40%

CLASS SCHEDULE

Week	Topics (Text: James et al. ISLR, 2nd ed)	Preparation
1 (March 3)	Introduction to machine learning in economics, Concepts and tools in statistical learning theory, Lab: Introduction to R and RStudio Part I	Ch. 1
2 (March 10)	Supervised and unsupervised learning, estimating prediction accuracy, bias-variance trade-off Lab: Introduction to R Part II, R programming, handling data in base R	Ch.1, Ch.2, Ch. 5
3 (March 17)	A modern interpretation of R: Tidy approach to data analysis in R, introduction to the R Tidyverse	Classnotes
4 (March 24)	Introduction supervised learning methods Linear regression	Ch. 3
5 (March 31)	Classification problems, logistic regression, PCA, discriminant analysis	Ch. 4
6 (April 7)	Validation and Cross-Validation, bias-variance trade-off, Bootstrap, data-dependent information criteria	Ch. 5
7 (April 14)	Model Selection and Regularization: shrinkage methods, LASSO, Ridge regression	Ch. 6
8 (April 21)	Midterm	
9 (April 28)	Nonlinear extensions: polynomial regression, splines, local regression, General Additive Model	Ch. 7
10 (May 5)	Tree-based methods, decision trees, regression and classification trees, random forests, bagging, boosting	Ch.8
11 (May 12)	Unsupervised Learning I: Dimension reduction methods, Principal Components Analysis (PCA), Introduction to clustering, K-means clustering	Ch. 12
12 (May 19)	No class	
13 (May 26)	Unsupervised Learning II: Clustering continued, Hierarchical clustering An introduction to the analysis of text data	Ch.12, Classnotes
14 (June 2)	Projects due + presentations	
15	Final exam week (project presentations)	