Course program and reading list

Semester 2 Year 2022

School: Efi Arazi School of Computer Science M.Sc.

Advanced Statistics for Data Science

Lecturer:

Dr. Alon Kipnis alon.kipnis@idc.ac.il

Teaching Assistant:

Mr. Ben Galili ben.galili@post.idc.ac.il

Course No.:	Course Type :	Weekly Hours :	Credit:
3676	Elective	3	3

Course Requirements :	Group Code :	Language:
Final Exam	222367600	English

Prerequisites

Prerequisite:

52 - Calculus I
53 - Calculus II
54 - Linear Algebra I
55 - Linear Algebra II
56 - Discrete Mathematics
59 - Data Structures
69 - Logic And Set Theory
109 - Introduction To Probability
417 - Introduction To Computer Science

Course Description

Overview

The course covers core topics in modern statistics at the forefront of research and applications in data science and machine learning. It combines theoretical background with hands-on exercises in data analysis and statistical learning. The course's main units are regression models, hypothesis testing, and variable selection. Specific topics include linear regression, model selection, Bayesian inference, hypothesis testing in one- and two-samples, ANOVA, sparse regression, multiple hypothesis testing, and false discovery-rate controlling.

Course Goals

The course's main objective is to provide theoretical and practical knowledge for learning from data.

Below are the main skills the students are expected to gain in each unit:

- Regression Models:
  - Fitting model parameters to data
  - Selecting the best model out of a family of models
- Multiple Hypothesis Testing:
  - Making decisions based on multiple features and experiments
- Variable Selection
  - Identifying useful features out of possibly many

Grading

The final grade will be based on the homework assignments (60%) and the final exam (40%).

(there will be 6 regular biweekly homework assignments. Homework will include analysis of datasets, theoretical problems, and programming assignments)

Lecturer Office Hours

Monday 15:00-16:00

Tutor Office Hours

TBD

Teaching Assistant

TBD

Reading List

The course does not follow any specific textbook; it borrows materials from:

- Introduction to Linear Regression Analysis. D. Montgomery, E. Peck.
- Modern Applied Statistics with S. D. Venables, B. Ripley.
- Elements of Statistical Learning by Hastie, Tibshirani, and Friedman (2008) (second edition), Springer, NY.
- Statistical Learning with Sparsity by Hastie, Tibshirani and Wainwright 2015 (free pdf)
- An introduction to the bootstrap. Efron, Bradley, and Robert J. Tibshirani. CRC Press, 1994.