The Hebrew University of Jerusalem

Syllabus

Statistical learning and data analysis - 52525

Last update 04-03-2019

HU Credits: 4

Degree/Cycle: 1st degree (Bachelor)

Responsible Department: Statistics

Academic year: 0

Semester: 2nd Semester

Teaching Languages: Hebrew

Campus: Mt. Scopus

Course/Module Coordinator: Yuval Benjamini

Coordinator Email: yuval.benjamini@mail.huji.ac.il

Coordinator Office Hours:

Teaching Staff:

Dr. Yuval Benjamini
Mr. boaz matan

Course/Module description:
The course deals with the statistical analysis of large modern datasets.
We will discuss statistical and computational challenges, and learn general
principles as well as specific methods of analysis.

In particular, we focus on exploratory data analysis, and on prediction models.

Labs (hand-in assignments) are an important part of the course. Students will
analyze real data sets including voting patterns, gene expression data, and neural
(fMRI) responses. They will also compare methods on real and simulated data.

Course/Module aims:
The course aims to present modern data analysis techniques. The goal is also for
the students to learn and practice research work in data analysis and statistical
method development.

Learning outcomes - On successful completion of this module, students should be
able to:
At the end of this course, students will be able to:
- Examine a data set and display its features
- Postulate a research interest as a prediction problem, and understand the
advantages and disadvantages of the prediction paradigm compared to other types
of inference
- Construct a prediction model (categorical or continuous)
- Quantify the success of the model, and compare different methods or models.
Estimate the error and uncertainty.
- Communicate the analysis in writing.

Attendance requirements(%):
0

Teaching arrangement and method of instruction: Lecture and discussion section,

Course/Module Content:
1. Cleaning and exploring data
2. PCA
3. Representation and distances

4. Clustering
5. Stability and Bootstrap
6. Introduction to supervised learning, bias vs. variance
7. Regression: expanding basis + wavelets
8. Regularized regression: Ridge, Lasso, Elastic Net
9. Regression trees
10. Classification: Generative models
11. Discriminative analysis
12. Boosting

Required Reading:
 None

 Additional Reading Material:
  Advanced Data Analysis from an Elementary Point of View, Cosma Rohillla Shalizi
http://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/

The Elements of Statistical Learning – Data mining, inference and prediction
(Tibshirani, Hastie and Friedman)
http://www-stat.stanford.edu/~tibs/ElemStatLearn/

 Course/Module evaluation:
 End of year written/oral examination 0 %
 Presentation 0 %
 Participation in Tutorials 0 %
 Project work 24 %
 Assignments 76 %
 Reports 0 %
 Research project 0 %
 Quizzes 0 %
 Other 0 %

Additional information: