The Hebrew University of Jerusalem

Syllabus

Mining big data - 52019

Last update 29-12-2023

HU Credits: 3

Degree/Cycle: 2nd degree (Master)

Responsible Department: Statistics

Academic year: 0

Semester: 1st Semester

Teaching Languages: Hebrew

Campus: Mt. Scopus

Course/Module Coordinator: Or Zuk

Coordinator Email: or.zuk@mail.huji.ac.il

Coordinator Office Hours: Monday 10:30-11:30

Teaching Staff:

Dr. Or Zuk

Course/Module description:
We will learn and apply methods for analyzing big datasets

Course/Module aims:
Acquiring statistical and computational tools for performing statistics on large-scale data

Learning outcomes - On successful completion of this module, students should be able to:
Analyze datasets with millions of records and thousands of variables. Use in an efficient manner programs with parallel/cloud computing. Extract and analyze data from the web.

Attendance requirements(%):
70

Teaching arrangement and method of instruction: Lectures, hands-on demonstrations on the computer

Course/Module Content:
Working remotely in a cluster environment and/or cloud computing.
Database (SQL), information extraction from the web.
Finding similarities: Hash functions, nearest neighbours
Distributed computing
Analyzing network data: finding communities, sampling large graphs
Streaming data: online algorithms

Additional subjects as time permits

Required Reading:
None

Additional Reading Material:
Leskovec, Rajaraman&Ullman (2014). Mining of massive datasets, Cambridge University Press

Tan, Steinbach, Karpatne and Kumar (2005). Introduction to Data Mining. Pearson Addison Wesley

Liu (2011). Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (Data-Centric Systems and Applications). Springer

White (2015). Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale. O'Reilly Media

<u>Grading Scheme:</u>
  Essay / Project / Final Assignment / Home Exam / Referat 75 %
  Submission assignments during the semester: Exercises / Essays / Audits / Reports / Forum / Simulation / others 25 %

<u>Additional information:</u>
  There will be a mid-term project and a few short check-list assignments during the semester that will comprise together 25% of the course grade.
After the end of the semester there will be given a final project that will comprise 75% of the course grade.
In addition, there will be a few exercises for self-practice (not for grade).