University of Milano-Bicocca
Department of Economics, Management and Statistics (DEMS)
Phd Programme in Economics and Statistics (ECOSTAT)

Statistical Learning and Big Data

Milano, October 7-18, 2019

Lecturer

Saharon Rosset is Professor of Statistics at Tel Aviv University. His research interests are in statistical learning theory and methodology, data mining and statistical genetics. Prior to his tenure at Tel Aviv, he received his PhD from Stanford University in 2003 under the supervision of Professors Jerome Friedman and Trevor Hastie and spent four years as a Research Staff Member at IBM Research in New York. His work has been published in the most prestigious journals in the fields of statistics and machine learning. He is a five-time winner of major data mining competitions, including KDD Cup (four times) and INFORMS Data Mining Challenge, and two time winner of the best paper award at KDD (ACM SIGKDD International Conference on Knowledge Discovery and Data Mining).
Website: https://m.tau.ac.il/~saharon/

Course objectives

The goal of this course is to gain familiarity with the basic ideas and methodologies of statistical (machine) learning. The focus is on supervised learning and predictive modeling in regression and classification. We will start by thinking about some of the simpler, but still highly effective methods, like nearest neighbors and linear regression, and gradually learn about more complex and "modern" methods and their close relationships with the simpler ones. We will also cover one or more industrial "case studies" where we track the process from problem definition, through development of appropriate methodology and its implementation, to deployment of the solution and examination of its success in practice. The course places emphasis on the theoretical contents of statistical learning: please check the Course prerequisites section.

Course programme

1st week: October 7-11, 2019
Module I (14 hours)
- Introduction, basic concepts: predictive modeling, decision theory, local non-parametric (kNN) vs global parametric (linear regression) methods, curse of dimensionality, error (bias-variance) decompositions
- Linear regression as predictive modeling: inference, regularization, variable selection, PCA
- Linear methods for classification: linear and logistic regression, LDA, SVM classification
- Case studies

2nd week: October 14-18, 2019
Module II (14 hours)
- Trees and Random Forest
- Boosting
- Deep learning
- Model evaluation and selection
- Case studies

We offer the possibility to enrol in single modules of 14 hours (*Module I* - 1st week or *Module II* - 2nd week) but we strongly recommend to take the full course because its structure is intended as a single block of 28 hours.

Course homepage

http://www.tau.ac.il/~saharon/StatLearn-Milan.html

Course references

Textbook:
Elements of Statistical Learning by Hastie, Tibshirani & Friedman
Book home page (including downloadable PDF of the book, data and errata)
Other recommended books:
Computer Age Statistical Inference by Efron and Hastie
Modern Applied Statistics with Splus by Venables and Ripley
Neural Networks for Pattern Recognition by Bishop
All of Statistics and *All of Nonparametric Statistics* by Wasserman
Online resources:
Data Mining and Statistics by Jerry Friedman
Statistical Modeling: The Two Cultures by the late, great Leo Breiman
Course on Machine Learning from Stanford's Coursera
The Netflix Prize competition is now over, but will still play a substantial role in our course

Course prerequisites

The course places emphasis on the theoretical contents of statistical learning rather than on programming and practical applications. It is highly recommended to meet and acknowledge each of the following prerequisites:
- Math basics: calculus and linear algebra at undergraduate level
- Probability basics (equivalent of 1-2 courses)
- Statistics: a course in regression, a course in statistical theory
- Programming: general familiarity, R familiarity an advantage

Examples of specific assumed background:
- Calculus: Integration and differentiation; Lagrange multiplier theory; Change of variable; L'Hospital's rule
- Algebra: Matrices and vectors; Linear spaces; Matrix inversion and spectral decompositions (SVD, Eigen decomposition); Optimization with matrix and vector valued objects; Trace calculations; Least squares calculations and geometric interpretations (projections)
- Probability: Multivariate normal distribution and its properties; Conditional distributions and expectations: Laws of Total Variation and Iterated Expectations
- Statistics: Basic definitions of statistical inference and hypothesis testing; Maximum likelihood estimation; Neyman-Pearson Lemma; Least squares regression and its standard inference; Logistic regression and GLMs; Quantiles and summary statistics;
- Optimization: linear and quadratic programming; Lagrange multipliers and optimality conditions

Venue and timetable

DEMS, University of Milano-Bicocca
Piazza dell'Ateneo Nuovo 1, 20126 Milano, Italy

Lectures will be in English, with the following schedule:
Monday to Friday: Lectures 14:30-17:00
Friday: Case studies 17:00-18:30

Fees and enrollment

- Non-academics: Euro 1200
- Academics (PhD students, post-docs, assistant professors, professors): Euro 600
- Academics (PhD students, post-docs, assistant professors, professors) of the University of Milano-Bicocca and of the Catholic University of Milano: no fee
- Participants attending one module only: 50% of the corresponding full fee

Application period

September 16, 2019 – September 23, 2019

Applications should be addressed via e-mail to the ECOSTAT Administration Office (Dr. Silvia Locatelli, e-mail: silvia.locatelli@unimib.it) and should include the following documents as attachments: i) filled-out application form; ii) updated curriculum vitae. Admissions are conditional on place availability. The ECOSTAT Administration Office will contact non-admitted candidates only. The application form can be downloaded from the following website (section Events): https://www.dems.unimib.it/en/research/phd-programme

Fees payment period

September 25, 2019 – October 1, 2019

Payments should be made online only. Detailed information will be published shortly at the following website (section Events): https://www.dems.unimib.it/en/research/phd-programme

Contacts

For more information:
Prof. Matteo Manera, Coordinator of ECOSTAT, e-mail: matteo.manera@unimib.it
Prof. Aldo Solari, ECOSTAT faculty member, e-mail: aldo.solari@unimib.it
For administrative issues:
Dr. Silvia Locatelli, ECOSTAT Administration Office, e-mail: silvia.locatelli@unimib.it