Statistical foundations of machine learning

Lecturer
Gianluca BONTEMPI (Coordinator)

Course mnemonic
INFO-F422

ECTS credits
5 credits

Language(s) of instruction
English

Course period
Second term

Campus
Plaine

Course content

(1) Foundations of statistical modelling, (2) parametric estimation, (3) nonparametric estimation and resampling, (4) supervised learning (model selection, variable selection), (5) algorithms for regression (neural networks, local learning, (6) classification algorithms (KNN, Naive- Bayes, SVM), (vii) applications of machine learning (data mining, text mining, web mining)

Objectives (and/or specific learning outcomes)

Statistical machine learning is the discipline which aims at extracting knowledge and inferring predictive models from observed data. The course will focus on the statistical notions (like bias, variance, regression, validation) which are necessary to create, identify and assess a predictive model. This course aims to find a good balance between theory and practice by situating most of the theoretical notions in a real context with the help of illustrative case studies (from biology, finance, medicine) and real datasets.

Pre-requisits and co-requisits

Courses having this one as co-requisit

INFO-F439 | Methods in Bioinformatics | 5 crédits, INFO-H515 | Big Data: Distributed Data Management and Scalable Analytics | 5 crédits and INFO-Y099 | Multicore programming | 6 crédits

Required knowledge and skills

- Basic notions of probability and estimation (bias, variance)
- Linear algebra and numerical analysis (linear systems, eigenvalues)
- Least-squares
- Programming

Teaching method and learning activities

Lectures, exercises and practical assignments. All the assignments involve practical work using the software package R.

Contribution to the teaching profile

- Analysis and mathematical modelling of information
- Collect, analyse, discuss and interpret data
- Learning of new concepts
- Design a modelling procedure
- Critical analysis of the results wrt state-of-the-art
- Operation knowledge of English
- Conceive a structural solution and algorithms to solve a problem
- Implement a prototype
- Learning of R statistical software

References, bibliography and recommended reading

- G. Bontempi "Statistical foundations of machine learning: the handbook [https://www.researchgate.net/publication/242692234_Statistical_foundations_of_machine_learning_"
- L. Wasserman (2004) All of statistics: a coincise course in statistical inference. Springer.
- R. O. Duda, P.E. Hart, D. G. Stork (2001) Pattern Classification. Wiley.
- T. Hastie, R. Tibshirani, J. Friedman (2001) The elements of statistical learning: data mining, inference, and prediction. Springer.

Course notes

Syllabus and Université virtuelle

Other information

Place(s) of teaching

Plaine

Contact(s)

Email: Gianluca.Bontempi@ulb.be

Office: Campus La Plaine, NO8-107

Postal address: Département d'Informatique, Bld de Triomphe, CP 212

Evaluation method(s)

Project and written examination

Evaluation method(s) (additional information)

Project (in R language) and written exam on theoretical aspects of the course. The written exam (on the UV platform) will require as well the usage of the R software to answer questions.

Determination of the mark (including the weighting of partial marks)

- 10/20 (project)
- 10/20 (UV written exam about theory requiring the use of the R software)

Main language(s) of evaluation

English

Other language(s) of evaluation, if applicable

French

Programmes

Programmes proposing this course at the faculty of Sciences

MA-BINF | Master in Bio-informatics and Modelling | finalité Research/unit 1, MA-GEOG | Master in Geography : General | finalité territorial Development/unit 2, MA-INFO | Master in Computer science | finalité Professional/unit 1 and finalité Professional/unit 2, MA-STAT | Master in Statistics : General | finalité Research General/unit 1 and MS-BGDA | Specialized Master in data science, Big data | unit U

Programmes proposing this course at the Solvay Brussels School of Economics and Management

MA-ECOE | Master in Economics : Econometrics | finalité Research in Economics and statistics/unit 2 and MS-BGDA | Specialized Master in data science, Big data | unit U

Programmes proposing this course at the Brussels School of Engineering

MA-IREM | Master of science in Electromechanical Engineering | finalité Operations engineering and management/unit 2, MA-IRIF | Master of science in Computer Science and Engineering | finalité Professional/unit 1 and finalité Professional/unit 2 and MS-BGDA | Specialized Master in data science, Big data | unit U