

Statistical foundations of machine learning

Titulaire

Gianluca BONTEMPI (Coordonnateur)

Mnémonique du cours

INFO-F422

Crédits ECTS

5 crédits

Langue(s) d'enseignement

Anglais

Période du cours

Deuxième quadrimestre

Campus

Plaine

Contenu du cours

(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)

Objectifs (et/ou acquis d'apprentissages spécifiques)

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, parametric and nonparametric estimation, 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.

Pré-requis et co-requis

Cours ayant celui-ci comme co-requis

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

Connaissances et compétences pré-requises

- > Basic notions of probability and estimation (bias, variance)

- > Linear algebra and numerical analysis (linear systems, eigenvalues)
- > Least-squares
- > Programming

Méthodes d'enseignement et activités d'apprentissages

5 ECTS (Th: 3, practicals: 1, project: 1)

Contribution au profil d'enseignement

- > 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

Références, bibliographie et lectures recommandées

- > Bontempi G., (2021) "Statistical foundations of machine learning: the handbook [https://www.researchgate.net/publication/242692234_Statistical_foundations_of_machine_learning_]"
- > T. Hastie, R. Tibshirani, J. Friedman (2002) The Elements of Statistical Learning. Springer.

Support(s) de cours

Syllabus et Université virtuelle

Autres renseignements

Lieu(x) d'enseignement

Plaine

Contact(s)

Email: Gianluca.Bontempi@ulb.be

Office: Campus La Plaine,

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

Méthode(s) d'évaluation

Examen écrit et Projet

Méthode(s) d'évaluation (complément)

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.

Construction de la note (en ce compris, la pondération des notes partielles)

10/20 (project)

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

Langue(s) d'évaluation principale(s)

Anglais

Autre(s) langue(s) d'évaluation éventuelle(s)

Français

Programmes

Programmes proposant ce cours à la faculté des Sciences

MA-BINF | **Master en bioinformatique et modélisation** | finalité Approfondie/bloc 1, MA-GEOG | **Master en sciences géographiques, orientation générale** | finalité Développement territorial/bloc 2, MA-INFO | **Master en sciences informatiques** | finalité Spécialisée/bloc 1 et finalité Spécialisée/bloc 2, MA-STAT | **Master en statistique, orientation générale** | finalité Approfondie/bloc 1 et MS-BGDA | **Master de spécialisation en science des données, Big data** | bloc U

Programmes proposant ce cours à la Solvay Brussels School of Economics and Management

MA-ECOE | **Master en sciences économiques, orientation économétrie** | finalité Research in Economics and Statistics/bloc 2 et MS-BGDA | **Master de spécialisation en science des données, Big data** | bloc U

Programmes proposant ce cours à l'école polytechnique de Bruxelles

MA-IREM | **Master : ingénieur civil électromécanicien** | finalité Operation engineering and management/bloc 2, MA-IRIF | **Master : ingénieur civil en informatique** | finalité Spécialisée/bloc 1 et finalité Spécialisée/bloc 2 et MS-BGDA | **Master de spécialisation en science des données, Big data** | bloc U