| UCLouvain | ldats2450
2025 | Statistical learning. Estimation, selection and inference |
|---|---|---|

The version you're consulting is not final. This course description may change. The final version will be published on 1st June.

5.00 credits	30.0 h + 7.5 h	Q2

Language :	English
Place of the course	Louvain-la-Neuve
Prerequisites	Concepts and tools equivalent to those taught in teaching units
LSTAT2020   Logiciels et programmation statistique de base
LSTAT2120   Linear models
LSTAT2100   Modèles linéaires généralisés et données discrêtes
Main themes	The course focuses on high-dimensional settings and on techniques to that allow for parameter estimation, model selection and valid inferential procedures for high-dimensional models in statistics.
Learning outcomes	**At the end of this learning unit, the student is able to :**
1   With regard to the AA reference framework of the Master's programme in Statistics, general orientation, this activity contributes to the development and acquisition of the following AAs, as a matter of priority : 1.4, 1.5, 2.4, 4.3, 6.1, 6.2
Evaluation methods	January Session:
 • During the semester the student must submit 2 compulsory assignments (short, 2-3 pages maximum per assignment), counting for 1 point of the final grade (each assignment = 0.5 points). The assignments are to be solved individually or in groups of 2. A mark will be assigned per group. Assignments arriving after the deadline are not considered.
 • A project (written in French/English in min 6 and max 12 pages in the template on Moodle, appendices not included) which will illustrate the methods of the course for 5 points. This (written) project will be submitted before the exam session and discussed with the teacher during the exam session. The evaluation of the project is done on the basis of the written report and on the basis of the answers in an oral discussion (without slides) on the results and methodology used for the report, during the exam session. The project is to be solved individually or in groups of 2. A score will be awarded per group. Projects arriving after the deadline are not considered.
 • An oral exam (~45min), in which the teacher will assess knowledge about the material covered in class (14 points), the quality of the project and the homework.
  The final grade for the LSTAT2450 course in January is given by the points obtained for the assignments + the points obtained for the project + the points obtained for knowledge on the subject.
 Attention: To validate the course, the student needs a final mark of 10 or more.
 August session:
 • A project (written in French/English in min 6 and max 12 pages in the template on Moodle, appendices not included) which will illustrate the methods of the course for 5 points. This (written) project will be submitted before the exam session and discussed with the teacher during the exam session. The evaluation of the project is done on the basis of the written report and on the basis of the answers in an oral discussion (without slides) on the results and methodology used for the report, during the exam session. The project is to be solved individually or in groups of 2. A score will be awarded per group. Projects arriving after the deadline are not considered.
 • An oral exam (~45min), in which the teacher will assess the knowledge about the material covered in class (15 points) and the quality of the project.
  The final grade for the LSTAT2450 course in August is given by the points obtained for the project + the points obtained for knowledge on the subject.
 Attention: To validate the course, the student needs a final mark of 10 or more.
 Attention: Any usage of artificial intelligence software for producing part of text, code, figures or equations that are included in the final project or homework is strictly forbidden. All projects and homework will be analyzied with specialized software.
Teaching methods	The class consists of lectures (30h) and exercises sessions (7.5h).
The classes and the TP are intended to be face to face.
Teaching language: English.

Content	The class is focused on the presentation of key concepts of statistical learning and high-dimensional models such as: • Statistical learning • Challenges concerning high-dimensional models and differences from low-dimensional models • Classical variable selection techniques for linear regression models: R2, adj.R2, Cp • Information criteria selection: KL divergence, AIC/TIC/BIC derivation • Cross-validation based selection: Leave-one-out and K-fold • Under- and overfitting or the bias-variance trade-off • Ridge shrinkage: theoretical properties, bias/variance trade-off, GCV • Lasso shrinkage: regularization paths, LARS, coordinate descent algorithm, prediction error bounds, degrees of freedom for lasso, support recovery, stability selection, knock-offs; inference by debiasing, post-selection inference, Bayesian inference • Extensions of Lasso: elastic net, group lasso, adaptive lasso, fused lasso • Other techniques: sparse graphical models, sparse PCA, sparse Disriminant Analysis
Inline resources	Moodle website of the class : LSTAT2450 - Statistical learning. Estimation, selection and inference. https://moodleucl.uclouvain.be/course/view.php?id=14890
Bibliography	• Hastie, T., Tibshirani, R. and Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer. • James, G., Witten, D., Hastie, T., and Tibshirani, R. (2014). An Introduction to Statistical Learning: With Applications in R. Springer • Hastie, T., Tibshirani, R. and Wainwright, M. J. (2015). Statistical Learning with Sparsity: The Lasso and Generalizations. Chapman and Hall/CRC. • Wainwright, M. J. (2019). High-Dimensional Statistics: A Non-Asymptotic Viewpoint. Cambridge University Press. • Bühlmann, P., van de Geer, S. (2011). Statistics for High-Dimensional Data. Springer.
Faculty or entity in charge	LSBA

Programmes containing this learning unit (UE)

Program title	Acronym	Credits	Prerequisite	Learning outcomes
Master [120] in Data Science : Statistic	DATS2M	5		🔍
Master [120] in Statistics: Biostatistics	BSTA2M	5		🔍
Master [120] in Mathematics	MATH2M	5		🔍
Master [120] in Statistics: General	STAT2M	5		🔍
Master [120] in Mathematical Engineering	MAP2M	5		🔍
Master [120] in Data Science Engineering	DATE2M	5		🔍
Certificat d'université : Statistique et science des données (15/30 crédits)	STAT2FC	5		🔍
Master [120] in Data Science: Information Technology	DATI2M	5		🔍