



Due to the COVID-19 crisis, the information below is subject to change, in particular that concerning the teaching mode (presential, distance or in a comodal or hybrid format).

5 credits	30.0 h + 7.5 h	Q1
-----------	----------------	----

Teacher(s)	Pircalabelu Eugen ;
Language :	English
Place of the course	Louvain-la-Neuve
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.
Aims	<p>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</p> <p>-----</p> <p><i>The contribution of this Teaching Unit to the development and command of the skills and learning outcomes of the programme(s) can be accessed at the end of this sheet, in the section entitled "Programmes/courses offering this Teaching Unit".</i></p>
Evaluation methods	<p>Due to the COVID-19 crisis, the information in this section is particularly likely to change.</p> <p>An oral examination, where the instructors evaluate:</p> <ul style="list-style-type: none"> • knowledge about the concepts seen in class throughout the semester (50% des points); • the quality of a project (written in French / English in min 5 and max 8 pages in the template on Moodle, annexes not included) of data analysis/simulation that illustrates the statistical learning methods in a concrete case (50% des points). This written project will be handed in before the exam session and discussed with the instructors during the exam session. The evaluation of the project is based on the written manuscript and responses to questions in an oral discussion about the results and the methodology used for the report. <p>The failure of one of the two parts results in the automatic failure of the course!</p> <p>To be allowed to take part in the examination the student has to submit 3 compulsory homeworks (short, 1-2 pages maximum per homework). The homeworks are not graded as they are not part of the evaluation.</p> <p>Submission of less than 3 homework results in failure of the course!</p>
Teaching methods	<p>Due to the COVID-19 crisis, the information in this section is particularly likely to change.</p> <p>The class consists of lectures (30h) and exercises sessions (7.5h).</p> <p>Teaching language: English.</p>
Content	<p>The class is focused on the presentation of key concepts of statistical learning and high-dimensional models such as:</p> <ul style="list-style-type: none"> • Statistical learning • Challenges concerning high-dimensional models and differences from low-dimensional models • Classical variable selection techniques for linear regression models: R^2, adj.R^2, C_p • 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 Discriminant Analysis
Inline resources	<p>Moodle website of the class : LSTAT2450 - Statistical learning. Estimation, selection and inference. https://moodleucl.uclouvain.be/course/view.php?id=14890</p>

Bibliography	<ul style="list-style-type: none">• 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	Aims
Master [120] in Data Science : Statistic	DATS2M	3		
Certificat d'université : Statistique et sciences des données (15/30 crédits)	STAT2FC	3		
Master [120] in Statistic: General	STAT2M	5		