



5.00 credits

30.0 h + 7.5 h

Q1

Teacher(s)	Pircalabelu Eugen ;
Language :	English
Place of the course	Louvain-la-Neuve
Prerequisites	LSTAT2011 Éléments de mathématiques pour la statistique LSTAT2013 - Concepts de base en statistique inférentielle LSTAT2120 Linear models LSTAT2020 Logiciels et programmation statistique de base
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	<p>At the end of this learning unit, the student is able to :</p> <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 1.4, 1.5, 2.4, 4.3, 6.1, 6.2</p>
Evaluation methods	<p>The evaluation for this course consists of three parts:</p> <ul style="list-style-type: none"> • During the semester, the student must hand-in 2 compulsory assignments (short, 1 to 2 pages maximum per assignment), counting for 20% of the final grade. The homework is to be solved individually or in groups of 2. A grade will be awarded per group. • A project (written in French / English in min 5 and max 9 pages in the template on Moodle, annexes not included) which will illustrate statistical learning methods in a concrete case (30% of the points). The project is evaluated on the basis of the written report. The project is to be solved individually or in groups of 2. A score will be awarded per group. • An oral exam (~ 45 min.) at which the lecturer will assess the knowledge of the student with respect to the materials covered during the class (50% of the points). If necessary the lecturer will also ask questions about the results and the methodology used for the report and for the homework. <p>The exact evaluation methods could be adapted according to the constraints linked to the sanitary conditions in force at the time of the exam sessions.</p>
Teaching methods	<p>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.</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	Learning outcomes
Master [120] in Statistics: General	STAT2M	5		
Certificat d'université : Statistique et sciences des données (15/30 crédits)	STAT2FC	5		
Master [120] in Data Science : Statistic	DATS2M	5		