UCLouvain

Istat2450

2022

Statistical learning. Estimation, selection and inference

5.00 credits 30.0 h + 7.5 h Q1

Teacher(s)	Pircalabelu Eugen ;					
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: 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	Session de janvier : 1. Pendant le quadrimestre l'étudiant doit rendre 2 devoirs obligatoires (courts, à 2 pages maximum par devoir), comptant pour 2 points de la note finale (chaque devoir = 1 points). Les devoirs sont à résoudre individuellement ou en groupe de 2. Une note sera attribuée par groupe. Les devoirs qui arrivent après la date limite ne sont pas prises en compte. 2. Un projet (récigé en français/anglais en min 6 et max 12 pages dans le template sur Moodle, annexes non-incluses) qui illustrera les méthodes du cours pour 5 points. Ce projet (écrit) sera remis avant la session d'examen et discuté avec le professeur lors de la session d'examen. L'évaluation du projet se fait sur base du rapport écrit et sur base des réponses dans une discussion orale (sans diapositives) sur les résultats et méthodologie utilisée pour le rapport, pendant la session d'examens. Le projet est à résoudre individuellement ou en groupe de 2. Une note sera attribuée par groupe. Les projets qui arrivent après la date limite ne sont pas prises en compte. 3. Un examen oral (~45min), auquel le professeur évaluera la maîtrise de la matière vue au cours (13 points), la qualité du projet et des devoirs. La note finale pour le cours LSTAT2450 en janvier est donné par les points obtenues pour les devoirs + les points obtenues pour le projet + les points obtenues pour le maîtrise de la matière. Attention : Pour valider le cours, l'étudiant a besoin d'une note finale de 10 ou plus. L'étudiant ne peut pas valider le cours si la note au projet est <3/5 (sans arrondi) ou si la note à l'examen est <6/13 (sans arrondi). Session d'août : 1. Un projet (récigé en français/anglais en min 6 et max 12 pages dans le template sur Moodle, annexes non-incluses) qui illustrera les méthodes du cours pour 5 points. Ce projet (écrit) sera remis avant la session d'examen et discuté avec le professeur lors de la session d'examen. L'évaluation du projet se fait sur base du rapport écrit et sur base des réponses dans une discussion orale (sans diapositi					
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 Statistics: Biostatistics	BSTA2M	5		٩		
Master [120] in Mathematics	MATH2M	5		٩		
Master [120] in Statistics: General	STAT2M	5		٩		
Master [120] in Data Science Engineering	DATE2M	5		٩		
Certificat d'université : Statistique et science des données (15/30 crédits)	STAT2FC	5		٩		
Master [120] in Data Science: Information Technology	DATI2M	5		٩		