UCLouvain

lstat2185

2022

Numerical Methods for Statistics: Optimization, Simulations and the Bootstrap

5.00 crédits	30.0 h + 15.0 h	Q1

Enseignants	Pircalabelu Eugen ;
Langue d'enseignement	Anglais
Lieu du cours	Louvain-la-Neuve
Préalables	Concepts et outils équivalents à ceux enseignés dans les UEs LSTAT2020 Logiciels et programmation statistique de base LSTAT2120 Linear models LSTAT2100 Modèles linéaires généralisés et données discrètes
Thèmes abordés	Le cours se concentre sur les méthodes numériques et sur les techniques de calcul intensif et itératif qui permettent (i) l'estimation des paramètres et (ii) l'inférence pour les modèles statistiques. Le cours introduit des concepts tels que l'optimisation contrainte et sans contrainte, les algorithmes numériques populaires tels que Newton, les méthodes quasi-Newton et l'algorithme EM (entre autres), la mise en place d'études de simulation pour évaluer les performances de plusieurs concurrents, ainsi que des techniques inférentielles basées sur le principe du Bootstrap.
Acquis d'apprentissage	**A la fin de cette unité d'enseignement, l'étudiant est capable de :** to understand the basic algorithms presented throughout the class and will be able to implement and apply these concepts to perform estimation/inference for classical models in statistics. to set up a simulation study to investigate the performance of multiple competitors and decide on the appropriate performance metrics, depending on the objectives of the study.
Modes d'évaluation des acquis des étudiants	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 (rédigé 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 LSTAT2185 en janvier est donné par les points obtenues pour les devoirs + les points obtenues pour le projet + les points obtenues pour la 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 (rédigé 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. 2. Un examen oral (~45min), auquel le professeur évaluera la maîtrise de la matière vue au cours (15 points) et la qualité du projet. La note finale pour le cours LSTAT2185 en août est donnée par les points obtenues pour le projet + les points obtenues pour la 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 écrit est < 6/15 (sans arrondi).

Méthodes d'enseignement	The class consists of lectures (30h) and exercise sessions (15h)

Contenu	The course outline is as follows:
	• Part I:
	  • Basics of one-dimensional function optimization. Special case: the likelihood function.
	  • Global vs local optima; numerical convergence and approximation errors.
	  • Challenges of optimizing multi-dimensional functions. Special cases: linear and generalized linear models and computing the multivariate normal density (link with LSTAT2120, 2100, 2110, 2040).
	  • Newton's method, Fisher scoring and IRLS.
	  • Non-linear functions and numerical differentiation.
	  • Case study: ?optim(), ?nlm(), ?deriv() and friends.
	  • The EM algorithm. Special cases: missing data, normal mixture models and linear mixed models (link with LSTAT2210).
	• Part II:
	  • Setting up controlled simulation studies: competitor selection, performance metrics and reproducibility.
	  • Sampling from distributions and DGPs.
	  • Case study I: Sample mean (mean, median, trimmed mean from normal and skewed distributions); German tank problem (estimators from slide LSTAT 2040) - mean, bias, MSE, RE.
	  • Case study II: t-test (size, power, coverage and length for CIs) and Binomial test (Tables from slides LSTAT 2040).
	  • Case study III: GLM variable selection (TPR, FPR, FDR).
	  • Parallel computing in R: doParallel, foreach, mclapply and friends to illustrate Case study I-III.
	• Part III:
	  • Bootstrap and resampling methods.
	  • Bias and variance approximation based on resampling.
	  • Bootstrap confidence intervals and hypothesis testing.
	  • Other techniques: Permutation tests and Jacknife

Ressources en ligne	Slides and notes on Moodle. Site Moodle of the class :LSTAT2185 - Numerical Methods for Statistics: Optimization, Simulations and the Bootstrap https://moodle.uclouvain.be/course/view.php?id=5785
Bibliographie	Givens, G.H. and Hoeting. J.A. (2013). Computational Statistics (2nd ed). Wiley. Rizzo, M.L. (2007). Statistical Computing with R (2nd ed). Chapman & Hall /CRC. Gentle, J.E. (2009). Computational Statistics. Springer. Lange, K. (2010). Numerical Analysis for Statisticians (2nd ed). Springer. Peng, R.D. (2020+). Advanced Statistical Computing. Available at https://bookdown.org/rdpeng/advstatcomp/ Chernick, M.R. (2008). Bootstrap methods : a guide for practitioners and researchers, Wiley Series in Probability and Statistics. Davison, A.C. et Hinkley, D.V. (1997). Bootstrap Methods and their Applications, Cambridge University Press. Efron, B. et Tibshirani, R.J. (1993). An Introduction to the Bootstrap, Chapman and Hall. Hall, P. (1992). The Bootstrap and Edgeworth Expansion, Springer. Mammen, E. (1992). When does bootstrap work ? Springer.
Autres infos	Software: R
Faculté ou entité en charge:	LSBA

Programmes / formations proposant cette unité d'enseignement (UE)

Intitulé du programme	Sigle	Crédits	Prérequis	Acquis d'apprentissage
Master [120] en science des données, orientation statistique	DATS2M	5		
Master [120] en statistique, orientation biostatistiques	BSTA2M	5		
Master [120] en statistique, orientation générale	STAT2M	5		
Certificat d'université : Statistique et science des données (15/30 crédits)	STAT2FC	5		