



5.00 crédits	30.0 h + 7.5 h	Q1
--------------	----------------	----

Enseignants	Pircalabelu Eugen ;
Langue d'enseignement	Anglais
Lieu du cours	Louvain-la-Neuve
Préalables	LSTAT2011 Éléments de mathématiques pour la statistique LSTAT2013 - Concepts de base en statistique inférentielle LSTAT2120 Linear models LSTAT2020 Logiciels et programmation statistique de base
Thèmes abordés	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.
Acquis d'apprentissage	<p>A la fin de cette unité d'enseignement, l'étudiant est capable de :</p> <p>1 Eu égard au référentiel AA du programme de master en statistique, orientation générale, cette activité contribue au développement et à l'acquisition des AA suivants, de manière prioritaire : 1.4, 1.5, 2.4, 4.3, 6.1, 6.2</p>
Modes d'évaluation des acquis des étudiants	<p>L'évaluation pour ce cours consiste en trois parties:</p> <ul style="list-style-type: none"> • Pendant le quadrimestre l'étudiant doit rendre 2 devoirs obligatoires (courts, 1 à 2 pages maximum par devoir), comptant pour 20% de la note finale. Les devoirs sont à résoudre individuellement ou en groupe de 2. Une note sera attribuée par groupe. • Un projet (rédigé en français/anglais en min 5 et max 9 pages dans le template sur Moodle, annexes non-incluses) qui illustrera les méthodes d'apprentissage statistique dans un cas concret (30% des points). L'évaluation du projet se fait sur base du rapport écrit. Le projet est à résoudre individuellement ou en groupe de 2. Une note sera attribuée par groupe. • Un examen oral (~45 min.), auquel le titulaire évaluera la maîtrise de la matière vue au cours (50% des points). Si nécessaire le titulaire posera aussi des questions par rapport aux résultats et la méthodologie utilisée pour le rapport et pour les devoirs. <p>Les modalités exactes d'évaluation pourraient être adaptées en fonction des contraintes liées aux conditions sanitaires en vigueur au moment des sessions d'examens.</p>
Méthodes d'enseignement	<p>Le cours comprend des exposés magistraux (30h) et des séances d'exercices (7.5h). Les cours et les TP sont destinés à être en présentiel. Langue d'enseignement: Anglais.</p>
Contenu	<p>Le cours est centré sur la présentation des concepts clés de l'apprentissage statistique et des modèles à haute dimension tels que:</p> <ul style="list-style-type: none"> • Apprentissage statistique • Défis concernant les modèles à haute dimension et différences par rapport aux modèles à faible dimension • Techniques classiques de sélection de variables pour les modèles de régression linéaire: R^2, adj.R^2, C_p • Sélection basée sur des critères d'information: divergence KL, dérivation AIC / TIC / BIC • Sélection basée sur la validation croisée: Leave-one-out et K-fold • Sous-ajustement et surajustement ou le compromis biais-variance • Ridge: propriétés théoriques, compromis biais / variance, GCV • Lasso: chemins de régularisation, LARS, algorithme de descente de coordonnées, limites d'erreur de prédiction, degrés de liberté pour le lasso, récupération du support, sélection de stabilité, knock-offs; inférence par débiaising, inférence post-sélection, inférence bayésienne • Extensions de Lasso: élastique net, lasso de groupe, lasso adaptatif, lasso fusionné • Autres techniques: modèles graphiques, PCA sparse, analyse discriminante sparse
Ressources en ligne	<p>Site Moodle du cours : LSTAT2450 - Statistical learning. Estimation, selection and inference. https://moodleucl.uclouvain.be/course/view.php?id=14890</p>

Bibliographie	<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.
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 statistique, orientation générale	STAT2M	5		
Certificat d'université : Statistique et sciences des données (15/30 crédits)	STAT2FC	5		
Master [120] en science des données, orientation statistique	DATS2M	5		