

En raison de la crise du COVID-19, les informations ci-dessous sont susceptibles d'être modifiées, notamment celles qui concernent le mode d'enseignement (en présentiel, en distanciel ou sous un format comodal ou hybride).

5 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>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> <p>-----</p> <p><i>La contribution de cette UE au développement et à la maîtrise des compétences et acquis du (des) programme(s) est accessible à la fin de cette fiche, dans la partie « Programmes/formations proposant cette unité d'enseignement (UE) ».</i></p>
Modes d'évaluation des acquis des étudiants	<p><b>En raison de la crise du COVID-19, les informations de cette rubrique sont particulièrement susceptibles d'être modifiées.</b></p> <p>Un examen oral, auquel le professeur évaluera :</p> <ul style="list-style-type: none"> <li>• la maîtrise de la matière vue au cours (50% des points);</li> <li>• la qualité d'un projet (rédigé en français/anglais en min 5 et max 8 pages dans le template sur Moodle, annexes non-incluses), qui illustrera les méthodes du statistical learning dans un cas concret (50% des 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 reponses dans une présentation orale sur les resultats et méthodologie utilisée pour le rapport, pendant la session d'examens.</li> </ul> <p>L'échec de l'une des deux parties entraîne l'échec automatique de l'examen!</p> <p>Pour être autorisé à passer l'examen, l'étudiant doit rendre 3 devoirs obligatoires (courts, 1 à 2 pages maximum par devoir). Les devoirs ne sont pas cotés car ils ne font pas partie de l'évaluation.</p> <p>La remise de moins de 3 devoirs entraîne l'échec du cours!</p>
Méthodes d'enseignement	<p><b>En raison de la crise du COVID-19, les informations de cette rubrique sont particulièrement susceptibles d'être modifiées.</b></p> <p>Le cours comprend des exposés magistraux (30h) et des séances d'exercices (7.5h)</p> <p>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"> <li>• Apprentissage statistique</li> <li>• Défis concernant les modèles à haute dimension et différences par rapport aux modèles à faible dimension</li> <li>• Techniques classiques de sélection de variables pour les modèles de régression linéaire: <math>R^2</math>, adj.<math>R^2</math>, <math>C_p</math></li> <li>• Sélection basée sur des critères d'information: divergence KL, dérivation AIC / TIC / BIC</li> <li>• Sélection basée sur la validation croisée: Leave-one-out et K-fold</li> <li>• Sous-ajustement et surajustement ou le compromis biais-variance</li> <li>• Ridge: propriétés théoriques, compromis biais / variance, GCV</li> <li>• 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</li> <li>• Extensions de Lasso: élastique net, lasso de groupe, lasso adaptatif, lasso fusionné</li> <li>• Autres techniques: modèles graphiques, PCA sparse, analyse discriminante sparse</li> </ul>

Ressources en ligne	Site Moodle du cours : LSTAT2450 - Statistical learning. Estimation, selection and inference. <a href="https://moodleucl.uclouvain.be/course/view.php?id=14890">https://moodleucl.uclouvain.be/course/view.php?id=14890</a>
Bibliographie	<ul style="list-style-type: none"> <li>• Hastie, T., Tibshirani, R. and Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.</li> <li>• James, G., Witten, D., Hastie, T., and Tibshirani, R. (2014). An Introduction to Statistical Learning: With Applications in R. Springer</li> <li>• Hastie, T., Tibshirani, R. and Wainwright, M. J. (2015). Statistical Learning with Sparsity: The Lasso and Generalizations. Chapman and Hall/CRC.</li> <li>• Wainwright, M. J. (2019). High-Dimensional Statistics: A Non-Asymptotic Viewpoint. Cambridge University Press.</li> <li>• Bühlmann, P., van de Geer, S. (2011). Statistics for High-Dimensional Data. Springer.</li> </ul>
Faculté ou entité en charge:	LSBA

### Force majeure

Modes d'évaluation des acquis des étudiants	<p>La crise sanitaire implique des incertitudes quant aux modalités d'évaluation en particulier pour la session de janvier. Deux options sont envisagées selon la sévérité des contraintes liées à la crise sanitaire.</p> <p>Un plan A en présentiel :</p> <ul style="list-style-type: none"> <li>• Examen oral</li> </ul> <p>Un plan B en distanciel :</p> <ul style="list-style-type: none"> <li>• Examen oral sur Teams</li> </ul>
---	--

<b>Programmes / formations proposant cette unité d'enseignement (UE)</b>				
Intitulé du programme	Sigle	Crédits	Prérequis	Acquis d'apprentissage
Master [120] en science des données, orientation statistique	DATS2M	3		
Certificat d'université : Statistique et sciences des données (15/30 crédits)	STAT2FC	3		
Master [120] en statistique, orientation générale	STAT2M	5		