






5.00 crédits	30.0 h + 7.5 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 le cadre 'modélisation en grande dimension' et sur les techniques permettant l'estimation des paramètres, la sélection de modèles et les procédures inférentielles valides pour les modèles de grande dimension en statistique.
Acquis d'apprentissage	<b>A la fin de cette unité d'enseignement, l'étudiant est capable de :</b> 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
Modes d'évaluation des acquis des étudiants	Session de janvier : 1. Pendant le quadrimestre l'étudiant doit rendre 2 devoirs obligatoires (courts, à 2-3 pages maximum par devoir), comptant pour 1 point de la note finale (chaque devoir = 0.5 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 (14 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 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 <2.5/5 (sans arrondi) ou si la note à l'examen est <7.5/14 (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 LSTAT2450 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 <2.5/5 (sans arrondi) ou si la note à l'examen est <7.5/15 (sans arrondi).
Méthodes d'enseignement	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>	<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 débiaisage, 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>
<p>Ressources en ligne</p>	<p>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></p>
<p>Bibliographie</p>	<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>
<p>Faculté ou entité en charge:</p>	<p>LSBA</p>

<b>Programmes / formations proposant cette unité d'enseignement (UE)</b>				
Intitulé du programme	Sigle	Crédits	Prérequis	Acquis d'apprentissage
Master [120] en statistique, orientation biostatistiques	BSTA2M	5		
Master [120] en sciences mathématiques	MATH2M	5		
Master [120] en statistique, orientation générale	STAT2M	5		
Master [120] : ingénieur civil en science des données	DATE2M	5		
Certificat d'université : Statistique et science des données (15/30 crédits)	STAT2FC	5		
Master [120] en science des données, orientation technologies de l'information	DATI2M	5		