UCLouvain

lepl1109

2020

Statistics and data sciences

Due to the COVID-19 crisis, the information below is subject to change, in particular that concerning the teaching mode (presential, distance or in a comodal or hybrid format).

5 credits	30.0 h + 30.0 h	Q1

Teacher(s)	Hainaut Donatien ;Jacques Laurent ;SOMEBODY ;
Language :	English
Place of the course	Louvain-la-Neuve
Prerequisites	To follow this course the student must have a basic knowledge of probabilities such as taught in courses **LEPL1108** or **LBIR1212**. *The prerequisite(s) for this Teaching Unit (Unité d'enseignement – UE) for the programmes/courses that offer this Teaching Unit are specified at the end of this sheet.*
Main themes	This course presents the fundamental statistical concepts in an engineering context (exploratory analysis, inference, simulation) as well as basis method for analysing multivariate databases (like the linear regression, the principal component analysis and the classification).
Aims	1 • Explore datasets of small and big sizes with few or many dimensions • Infer features of a population from a sample using techniques of inference, estimation, confidence intervals and statistical tests. • To connect the deductive approach from the probability theory to the statistical inductive approach, and to identify the probabilistic models used in statistical inference. • To translate the textual formulation of a problem of statistical inference into an accurate, statistical and mathematical formalism, while recognizing the adequate models and corresponding estimation methods. • To solve an applied problem by following a logical approach based on a correct use of models and statistical inference. • To use techniques of Monte-Carlo simulations, K-fold cross validation and bootstrapping in order to estimate models and validate results. • To analyse multivariate data with fundamental methods of linear regressions, of principal component analysis and of classification/clustering. • To use statistical tools to validate the conclusions from a model e.g. like the linear regression. • To make the link between the mathematical objectives of a method of data mining and its practical purposes. - - - - *The contribution of this Teaching Unit to the development and command of the skills and learning outcomes of the programme(s) can be accessed at the end of this sheet, in the section entitled "Programmes/courses offering this Teaching Unit".*
Evaluation methods	**Due to the COVID-19 crisis, the information in this section is particularly likely to change.** Written individual exam to evaluate the understanding of concepts and techniques The hackathons represents 20% of the final mark. Lecturers keep the right to orally question students about their exam and hackathons.
Teaching methods	**Due to the COVID-19 crisis, the information in this section is particularly likely to change.** The course is composed of: - 9 lectures on the topics listed in the course content; - 7 practical sessions, both classical and numerical; - 4 hackathons, representing 2 x 2 hours each, associated with small Python projects realized in group on subjects discovered both in the lectures and in the practical sessions.
Content	- Exploratory analysis and sampling - Introduction to multivariate data analysis - Parametric estimate (methods of moments and log-likelihood maximization) and properties of estimators (bias, variance, mean-squared error). - Statistical inference (confidence intervals and significance tests): comparison of means of two or several normal populations, proportions, variance testing. - Linear regression, including the analysis of coefficients and significance tests. - Panorama of learning techniques, supervised and unsupervised learning methods - Links between objectives of data analysis methods and their mathematical representation.

	- Regression and classification methods (such as linear models and least square, k-nearest neighbors, logistic regression) - Training, test error and generalization error, the Bias-Variance tradeoff, and elements of statistical decision theory - Resampling techniques for model selection/evaluation (e.g., validation set, K-fold cross validation, bootstrap) - Unsupervised learning: reduction of dimension (principal component analysis) and methods of clustering (K-means).
Inline resources	The totality of teaching material is available on the companion moodle website of the course. The schedule of course is subject to modification due to sanitary conditions, please consult the Moodle website of the course for additional information.
Other infos	To follow this course the student must have a basic knowledge of probabilities such as taught in courses **LEPL1108** or **LBIR1212**. The schedule of course is subject to modifications due to sanitary conditions. Please check the Moodle website for more details.
Faculty or entity in charge	EPL

Force majeure

Teaching methods	- the lectures, the practical sessions as well as the hackathons are organized remotely
Evaluation methods	The evaluation is based on the continuous assessment and on a written exam, that respectively count for 20% and 80% of the final mark. If sanitary conditions permit it, a closed book examination will be held on site, during the session. Failing this, a distance open book examination will be organized.

Programmes containing this learning unit (UE)

Program title	Acronym	Credits	Prerequisite	Aims
Bachelor in Computer Science	SINF1BA	5	LBIR1212	🔍
Bachelor in Engineering	FSA1BA	5	LEPL1108	🔍
Master [120] in Environmental Science and Management	ENVI2M	5		🔍