UCLouvain

ldats2310

2021

Data science for insurance and finance

3 credits	15.0 h	Q1

Teacher(s)	Hainaut Donatien ;
Language :	English
Place of the course	Louvain-la-Neuve
Prerequisites	A first course in probability and statistics is required e.g. : LBIR1203 Probabilités et statistiques I and LBIR1304 Probabilités et statistiques II (or equivalent modules). A good knowledge of linear regression models (LSTAT2120 Linear models) is an asset. *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 module aims to introduce recent developments in the field of statistical learning, applied to the insurance and financial sectors. Statistical methods are used in the insurance industry to assess the risk profile of an insured. This profile presents two sides: one is the frequency of claims and the other is the size of the claim caused by the insured. Both aspects are studied carefully by insurers so as to propose the best price for an insurance coverage. In the financial industry, advanced statistical methods are needed to evaluate the credit risk of a lender. As for an insurance contract, this risk has two sides. The first one is the probability that the lender will not repay is debt (the default risk). The second aspect is the size of the loss when the lender do not redeem is loan. This module present the common tools to study these risks: generalized linear models, additive models, Regression/classification trees. Some new aspects will also be developed among them we quote shrinkage methods (Lasso, Ridge) and random forests that reveals to be powerful tools to explore massive data.
Aims	At the end of this course, students will be able: 1 • To explain and motivate the choice of a statistical method to analyze insurance or financial data • To use Generalized Linear and Additive models to propose a grid of insurance premium or to propose a model to evaluate the default risk of a counterparty • To use Regression Tree and random forest on insurance or credit datasets. • To adapt the previous cited methods to include constraints of sparsity in the calibration (Lasso Ridge) • To understand the interests of bootstrapping methods and to implement them. - - - - *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.** Students will prepare an individual report in which they compare the GLM and regression tree procedures, to propose a grid of insurance premiums (motor insurance). The dataset is proposed by the lecturer. Notice that the lecture keeps the right to orally question the student on the content of his report.
Teaching methods	**Due to the COVID-19 crisis, the information in this section is particularly likely to change.** • Lectures based on readings • Programs in R • Case studies
Content	1. Introduction to Non-Life Insurance Pricing • Data science and non-life insurance pricing • The compound Poisson model applied to - non-life insurance - credit risk 2. Generalized Linear Models • Claims frequency regression problem • Claims size regression problem • Inference and prediction • The overdispersed Poisson case for claims count modeling - Deviance statistics and parameter reduction - Example in moto insurance pricing • The Gamma case for claims size modeling

	- Example in moto insurance pricing 3. Cross validation and model selection • Cross validation and model selection - Leave-one-out cross-validation - K-fold cross-validation - Stratified K-fold cross-validation 4. Generalized additive models (GAMs) • GAMs for Poisson Regression - Natural cubic splines - Example in moto insurance pricing - Multivariate adaptative regression splines 5. Shrinkage methods for GLM • Sparcity - Lasso GLM - Ridge GLM - Elastic net GLM 6. Classification and Regression trees • Poisson regression tree in insurance and credit risk (CART) - Example in moto insurance pricing - Example in credit risk • Sparse regression trees 7. Bootstrapping • Bootstrap method - Non-Parametric bootstrap - Parametric bootstrap - Illustration • Bagging - Bagging for Poisson regression trees 8. Random forests • Parametric Poisson rand. forests • Non-parametric Poisson rand. forests 9. Boosting machine • Gradient boosting machine • Poisson deviance tree boosting machine • adaBoost algorithm
Inline resources	Moodle website
Bibliography	Slides available on moodle are based on the following references • Data Analytics for Non-Life Insurance Pricing. Lecture notes, M. Wüthrich, Risklab Switzerland, ETH Zurich. • Non-life Insurance pricing with Generalized Linear models. E. Ohlsson, B. Johansson, Springer eds (2010). • The elements of statistical learning: Data mining, Inference, Prediction. T. Hastie, R. Tibshirani, J. Friedman, Second edition, Springer 2008.
Faculty or entity in charge	LSBA

Programmes containing this learning unit (UE)

Program title	Acronym	Credits	Prerequisite	Aims
Master [120] in Statistics: General	STAT2M	3		
Master [120] in Actuarial Science	ACTU2M	3	LACTU2110	
Certificat d'université : Statistique et sciences des données (15/30 crédits)	STAT2FC	3		
Master [120] in Data Science : Statistic	DATS2M	3		