

Academic Year: (2023 / 2024)

Review date: 26-04-2023

Department assigned to the subject: Statistics Department

Coordinating teacher: AUSIN OLIVERA, MARIA CONCEPCION

Type: Electives ECTS Credits : 6.0

Year : Semester :

DESCRIPTION OF CONTENTS: PROGRAMME

1. Introduction.
2. Data collection, sampling and preprocessing.
 - 2.1. Types of data.
 - 2.2. Sampling.
 - 2.3. Data visualization tools.
 - 2.4. Missing values.
 - 2.5. Outlier detection and treatment.
 - 2.6. Data transformations.
 - 2.7. Dimension reduction.
 - 2.8. Application: Risk management in the stock market.
3. Supervised learning: regression.
 - 3.1. Linear and polynomial regression.
 - 3.2. Cross-validation.
 - 3.3. Model selection and regularization methods (ridge and lasso).
 - 3.4. Nonlinear models, splines and generalized additive models.
 - 3.5. Application: credit-scoring prediction.
4. Supervised learning: classification.
 - 4.1. Bayes classifiers
 - 4.2. Logistic regression.
 - 4.3. K-nearest neighbors.
 - 4.4. Random forest.
 - 4.5. Support-vector machines.
 - 4.6. Boosting.
 - 4.7. Application: Credit risk.
 - 4.8. Application: Fraud detection.
 - 4.9. Application: Bankruptcy prediction

LEARNING ACTIVITIES AND METHODOLOGY**THEORETICAL PRACTICAL CLASSES**

Knowledge and concepts students must acquire. Student receive course notes and will have basic reference texts to facilitate following the classes and carrying out follow up work. Students partake in exercises to resolve practical problems and participate in workshops and an evaluation tests, all geared towards acquiring the necessary capabilities. Subjects with 6 ECTS are 44 hours as a general rule/ 100% classroom instruction.

TUTORING SESSIONS

Individualized attendance (individual tutoring) or in-group (group tutoring) for students with a teacher. Subjects with 6 credits have 4 hours of tutoring/ 100% on-site attendance.

STUDENT INDIVIDUAL WORK OR GROUP WORK

Subjects with 6 credits have 98 hours/0% on-site.

ASSESSMENT SYSTEM

Final exam (60%). Presentation of exercises in class and recording explanatory videos (40%).

% end-of-term-examination:	60
% of continuous assessment (assignments, laboratory, practicals...):	40

BASIC BIBLIOGRAPHY

- Bradley Efron, Trevor Hastie. Computer Age Statistical Inference: Algorithms, Evidence and Data Science., Cambridge University Press, 2016
- James, G., Witten, D., Hastie, T., Tibshirani, R. An Introduction to Statistical Learning with Applications in R, Springer, 2013