Faculty: Science and Technology

Course: **Statistics and Data Analysis**

Program: Study Abroad in Engineering

Semester: 2 - Spring

ECTS credits: 9

Duration: 67,5 hours

Language of instruction: English

Instructor: Rafel Martí i Nora Silva

Course Description

This course provides a comprehensive exploration of multidimensional data analysis and machine learning techniques. The syllabus is divided into two main modules – one focusing on statistical analysis and the other on machine learning.

Students will gain theoretical knowledge and practical skills in utilizing mathematical tools, parametric statistics, linear models, principal component analysis (PCA), and classic machine learning algorithms. Practical implementation will be emphasized using Python (Scikit-Learn) and R software.

Prerequisites

- Matrix calculation (diagonalization)
- Basic Probabilities (multidimensional probability laws, moments)
- Basic Statistics (moment and maximum likelihood methods)
- Descriptive statistics
- Basic knowledge of Python language and R software

Attendance policy
Attendance is mandatory for all classes, including study visits. Any exams, tests, presentations, or other work missed due to student absences can only be rescheduled in cases of certified medical or family emergencies.
Regular attendance is expected for both online lectures and face-to-face practical tutorials.
Participation in discussions and collaborative work is essential for a comprehensive understanding of the material.

Learning outcomes

By the end of the course, students should be able to:

- Modelize and organize a dataset.
- Analyze and organize real-world data.
- Master statistical processing software.
- Understand different types of machine learning: supervised, unsupervised, and reinforcement.
- Master theoretical concepts of classic machine learning algorithms.
- Implement machine learning algorithms in Python from scratch.
- Choose appropriate algorithms for specific use cases.
- Use Python libraries for implementing machine learning methods.
- Interpret the results of algorithms and identify their limits.
- Evaluate and compare multiple machine learning models for a particular application.

Method of presentation
- The course will be presented through a combination of online lectures, practical tutorials, and face-to-face sessions. Real-world datasets will be utilized for hands-on implementation, and tools like Numpy, Pandas, Matplotlib, Seaborn, and machine learning libraries will be covered.

Required work and assessment methods
- Practical Exercises and implementation.
- Online lectures
- Practical tutorials
- Validation test

Unit One: Inferential statistics using RStudio and Introducing Python

Week 1. Population and Sample. (P1)

Week 2. Probability Distribution (P2)

Week 3. Hypothesis testing .

Week 4. Confidence intervals (P3).

Week 5. Parametric and Nonparametric test.

Week 6. Sample Exercises and Practices.

Unit Two: Regression using RStudio and Introducing Python

Week 7. Simple linear Regression

Week 8. Multi Linear Regression. Model Evaluation and Selection. (P4)

Week 9. Python Basics for Regression.

Week 10. Polynomial Regression and Regularization. (P5)

Unit Three: Data Analysis.

Week 11. Introduction to Data Analysis.

Week 12. Advance Data Analysis with RStudio. Regression Analysis and Data Visualization. (P6)

Week 13. Python for Data Analysis.

Week 14. Advance Data Analysis with Python. Python Techniques and Regression Analysis. (P7)

Week 15. Final Test (TEST)

Activities weight. Ordinary evaluation

	P1	P2	P3	P4	P5	P6	P7	TEST
%	10	10	10	10	10	10	10	30

Retake exams and activities
The student must present the activities pending to delivery (P1 from P7).
If the student passed the test TEST, it is not necessary to take the retake test. It is only mandatory to take the retake test if the student fails it.

The activities weight in the retake evaluation is the same as activities weight in the ordinary evaluation (see the percentages table), but the maximum grade is 5.

Bibliography:
A. Géron, Machine Learning avec Scikit-Learn (2e édition), 2019
L. Bellanger & R. Tomassone, Exploration de données et méthodes statistiques, 2014
F. Husson, S. Lê & J. Pagès, Analyse de données avec R (2e édition), 2016
P-A Cornillon & al., R pour la statistique et la science des données, 2018