Machine learning in data mining

Academic Year: (2022 / 2023)   **Review date:** 20/05/2022 18:51:20

Department assigned to the subject: Computer Science and Engineering Department

Coordinating teacher: ALER MUR, RICARDO

Type: Electives ECTS Credits : 6.0

Year : 4 Semester :

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

Programming II

OBJECTIVES

1.) OF KNOWLEDGE:
- Know the different tasks that can be solved with machine learning
- Know machine learning techniques and their typology
- Know the methodology of machine learning and the phases it entails
- Know tools available for machine learning
2.) UNDERSTANDING:
- Understand the fundamentals and motivations of machine learning
- Understand the work methodology and the different phases of machine learning
- Understand the usefulness of different machine learning techniques
- Understand the relationship between model complexity, amount of data, problem characteristics and overlearning
3.) APPLICATION:
- Analyze the domains and design knowledge extraction processes according to the problem.
- Evaluate the performance and efficiency of the different machine learning methods
- Work on specific domains and contrast different techniques to check their performance in machine learning
4.) CRITICISM OR ASSESSMENT
- Selection of algorithms, selection of models and adjustment of parameters.
- Consider the relationship between computational cost and marginal improvement of different solutions
- Assessment of whether the results obtained are adequate, compared with chance or basic algorithms

DESCRIPTION OF CONTENTS: PROGRAMME

1. Introduction to Machine Learning
2. Basic methods for classification and regression:
 2.1. Nearest neighbour (KNN)
 2.2. Trees and Rules
3. Machine Learning pipeline
 3.1. Training
 3.2. Hyper-parameter tuning
 3.3. Evaluation
 3.4. Preprocessing and feature selection
4. Advanced methods for classification and regression
 4.1. Ensembles: bagging, random forests, boosting
 4.2. Support Vector Machines

LEARNING ACTIVITIES AND METHODOLOGY

Theory: Lectures will be focused on teaching all concepts related to machine learning. They will be carried out in synchronous on-line mode.

Practical computer Sessions: The practical classes will be developed so that, in a supervised way, students learn to solve real problems with machine learning. The practices will be carried out in groups of 2 students. There are several assignments related to topics in the course.

There will be tutorials to help the understanding both of theory and practice.

ASSESSMENT SYSTEM

% end-of-term-examination/test:	30
% of continuous assessment (assigments, laboratory, practicals…):	70

A)Final exam, focusing on the theoretical side: 30% (2 points)
B)Assignments, to be carried out during the semester: 70% (8 points)

The final course grade is calculated by adding the marks obtained in A and B. The final exam has no cutoff grade. A+B>=5 implies passing this course.

BASIC BIBLIOGRAPHY

 - Aurélien Géron Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, O'Reilly Media, 2019

 - Brett Lantz Machine Learning with R, Packt Publishing, 2019

 - Max Kuhn Applied Predictive Modeling, Springer, 2013

ADDITIONAL BIBLIOGRAPHY

 - Hadley Wickham, Garrett Grolemund, R for Data Science: Import, Tidy, Transform, Visualize, and Model Data, O'Reilly Media, 2016

BASIC ELECTRONIC RESOURCES

 - MLR . Machine Learning in R: https://mlr3.mlr-org.com/

 - Scikit-learn team . Scikit-learn: machine learning in Python: https://scikit-learn.org/