Machine learning in data mining

Academic Year: (2021 / 2022)   **Review date: 10/06/2021 17:47:14**

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.) KNOWLEDGE:
- To know the basics of extracting knowledge from data
- To know the different tasks that can be solved with machine learning
- To know the different techniques of machine learning and their characteristics
- To know the methodology of knowledge extraction and the phases involved
- To know tools available for extracting knowledge
2.) UNDERSTANDING:
- To understand the basic concepts of knowledge extraction
- To understand the basics and motivations of data mining
- To understand the methodology and the different phases of knowledge extraction
- To understand the usefulness of different techniques for extracting knowledge
- To understand the differences of different representations: propositional and relational
- To understand the relationship between model complexity, amount of data, characteristics of the problem and overfitting
3.) APPLICATION:
- Analyze the domain and design knowledge extraction processes adapted to the problem.
- Evaluate the performance and efficiency of different methods of extracting knowledge
- Work on specific domains and compare different techniques to verify their performance in extracting knowledge
4.) EVALUATION
- Selection of algorithms, model selection and parameter adjustment.
- Consider the relationship between computational cost and marginal improvement of different solutions
- Assessment of whether the results are adequate, compared to random 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. Machine learning with R and with the MLR (mlr3) library.
5. Advanced methods for classification and regression
  5.1. Ensembles: bagging, random forests, boosting
  5.2. Support Vector Machines
6. Brief introduction to machine learning in Python (scikit-learn library)

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:	20
% of continuous assessment (assigments, laboratory, practicals…):	80

A)Final exam, focusing on the theoretical side: 20% (2 points)
B)Assignments, to be carried out during the semester: 80% (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.

The main tool for practical assignments is the R language and the MLR (mlr3) library. There will also be a brief introduction to machine learning using Python and its scikit-learn library.

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

- Hefin I. Rhy Machine Learning with R, the tidyverse, and mlr, Manning Publications, 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: https://scikit-learn.org/