470SM - INTRODUCTION TO MACHINE LEARNING

Teaching objectives

Knowledge and understanding. Know main kinds of problems which can be tackled with ML and EC and those ones concerning text and natural language. Know main ML techniques; know the high-level working scheme of EAs. Know design, development, and assessment phases of a ML system; know main assessment metrics and procedures suitable for a ML system. Applying knowledge and understanding. Formulate a formal problem statement for simple practical problems in order to tackle them with ML or EC techniques. Develop simple end-to-end ML DM systems. Experimentally assess a simple end-to-end ML DM system.

Teaching methods

Frontal lessons with blackboard and slide projection; exercises, under teacher supervision, in dealing with simple problems with ML techniques.

Examination

The exam consists of a project and a written test. The final grade is the average of the two grades: the exam is considered failed if at least one of the two grades is 25 are automatically registered; in the remaining cases, the student may repeat the exam.
Written test: questions on theory and application with short open answers.
Project (home assignment): the student chooses a problem among a closed, teacher-defined set of problems and proposes a solution based on ML or EC techniques. The expected outcome is a written document (few pages) including: the problem statement; one or more performance indexes able to capture any solution ability to solve the problem; a description of the proposed solution from the algorithmic point of view; the results and a discussion about the experimental assessment of the solution with, if applicable, information about used data. Students may form groups for the project: in this case, the document must show, for each student of the group, which activities the student took part in. The project is evaluated according to clarity (~ 50%), technical soundness (~ 33%), and results (~ 17%).

Course Program (preliminar)

Definition of Machine Learning; examples of applications of ML; taxonomy of ML problems; phases of design, development, and assessment of a ML system; terminology and mathematical notation.
Introduction to the software/language R; elements of data visualization.
Supervised learning.
- Tree-based methods. Decision and regression trees: learning and prediction; role of the parameter and overfitting. Trees aggregation: bagging, Random Forest, boosting. Supervised learning system assessment: cross-fold validation; accuracy and other metrics; metrics for binary classification (FPR, FNR, EER, AUC) and ROC.
- Support Vector Machines (SVM). Separating hyperplane: maximal margin classifier; support vectors; learning as an optimization problem; maximal margin classifier limitations. Soft margin classifier: learning, role of the parameter C. Non linearly separable problems; kernel: brief background and main options (linear, polynomial, radial); intuition behind radial kernel; SVM. Multiclass classification with SVM.
- Naive-Bayes classification.
- The K-nearest neighbors classifier.

Unsupervised learning
- Cluster analysis: hierarchical methods, partitional methods (k-means algorithm)

Text and natural language applications (text mining)

- Sentiment analysis; features for text mining; common pre-processing steps; topic modeling.