Computational Machine Learning II

Winter Term - 3 ECTS

Mandatory Course

Prof. Joan Verd€

Prerequisites to Enroll

Computational Machine Learning I

Computing for Data Science

Knowledge of Python, Jupyter notebooks, and algebra.

Overview and Objectives

This is an extended course to machine learning based on a hands-on approach using Jupyter notebooks plus extra theoretical lessons for deeper understanding.
Computational Machine Learning II is the continuation of an introduction to some of the basic techniques of machine learning required for data science. It provides a solid training in computational algorithms for supervised problems (classification and regression), such as decision trees and forests, support vector machines or nearest neighbors. There is a hands-on part that focuses on the use of scientific scripting languages and special attention is devoted to Python language and working in a Jupyter Notebooks environment. This practical part is followed by more in-depth explanations of the techniques.

All material is motivated by specific information retrieval and data analysis questions and each thematic unit concludes with a small project.
The course will be delivered by Joan Verdú, Head of Consulting and Knowledge Transfer at the BGSE Data Science Center, in collaboration with Data Scientists affiliated to the Data Science Center.

Course Outline

The course covers the following list of topics:

O. Introduction

Introduction to the jupyter notebook programming environment
Introduction to Supervised learning (classification algorithms, hyperparameter fine-tuning, training and

testing)
Handling missing data

A. Nearest neighbor methods

k-Nearest-Neighbor Classifiers
Adaptive Nearest-Neighbor Methods
Computational Considerations
Introduction to k-NearestNeighor with python (project)

B. Support vector machines

Introduction to linear decision boundaries for classification
Linear model: Optimum hyperplane in separable case
Linear model: Optimum hyperplane in non-separable case
Non-linear models: Support Vector Machine with Kernels
Introduction to SVM with python (project)

C. Decision trees algorithms

Introduction to decision tree
Complexity, accuracy and overfitting
Ensemble methods: forests, bagging and boosting
Introduction to decision trees with Python (project)

D. Model ensembling

Stacking and blending
Feature propagation

E. Model interpretability

LIME

Shaply and SHAP values

F. Uncertainty estimation in ML

Required Activities and evaluation

Attendance at classes, and submission of homework.

2 or 3 projects in total

- The in-class projects will be individual
- The home projects will be in groups of 2
- Will be given at the last class of each part
- Students will be given 2 weeks to submit their project.

The grade will be the average of these projects, plus eventually some extra tasks to be done in class.

Competences

☒ Construct a global vision of the situation of the problem based on knowledge of the synergies between advanced statistical methods, computing and business analysis to generate added value.

☒ Modeling and predicting high-dimensional data with advanced statistical methods in the field of data science in order to improve strategic decision making.

☒ Solve the real problems that arise in the fields of study through the accurate analysis of the data.

☒ That the students know to communicate their conclusions and the knowledge and last reasons that sustain them to specialized and non-specialized publics in a clear and unambiguous way.

☒ That students have the learning skills that allow them to continue studying in a way that will be largely self-directed or autonomous.

Computational Machine Learning II

Winter Term - 3 ECTS

Mandatory Course

Prof. Joan Verd€

Learning Outcomes

☒ Elaborate and estimate probabilistic prediction models based on certain data.

☒ Apply supervised and semi-supervised learning algorithms.

☒ Predicting information needs based on decisions that must be made.

Materials

Ian H. Witten, Eibe Frank, Mark A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3rd Edition, Elsevier Inc, 2011

Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006

∃Software for Data Analysis: Programming with R&, John Chambers, Springer 2008

Hastie, T., R. Tibshirani, and J. H. Friedman (2009). The elements of statistical learning: data mining, inference, and prediction (Second ed.). Berlin; New York: SpringerVerlag Inc

VanderPlas,J., 2016. *Python Data Science Handbook*. O'Reilly