Course guide
230338 - APATP - Machine Learning: from Theory to Practice

Last modified: 17/06/2024

Unit in charge: Barcelona School of Telecommunications Engineering
Teaching unit: 739 - TSC - Department of Signal Theory and Communications.

Degree: BACHELOR'S DEGREE IN TELECOMMUNICATIONS TECHNOLOGIES AND SERVICES ENGINEERING (Syllabus 2015). (Optional subject).
BACHELOR'S DEGREE IN ELECTRONIC ENGINEERING AND TELECOMMUNICATION (Syllabus 2018). (Optional subject).

Academic year: 2024 **ECTS Credits:** 2.0 **Languages:** Catalan, Spanish

LECTURER

Coordinating lecturer: Josep Vidal

Others: Josep Vidal, Veronica Vilaplana

PRIOR SKILLS

Estadística

REQUIREMENTS

BACHELOR'S DEGREE IN TELECOMMUNICATIONS TECHNOLOGIES AND SERVICES ENGINEERING (GRETST)
PROBABILITY AND STATISTICS - Prerequisite
BACHELOR'S DEGREE IN ELECTRONIC ENGINEERING AND TELECOMMUNICATION (GREELEC)
PROBABILITY AND STOCHASTIC PROCESSES

TEACHING METHODOLOGY

Time will be shared by theory lectures and labs in Python.

LEARNING OBJECTIVES OF THE SUBJECT

This subject reviews the theory of classification and regression based on data from a mathematical perspective and at the same time applied to different areas. It is intended to give an overview of Bayesian decision theory, parametric classification techniques (logistic regression, neural networks, decision trees, combination of classifiers, etc.) and unsupervised learning with simultaneous experimentation in different applications, among which we highlight decision on biomedical data, image classification, event detection, spam detection, etc...
The credits of the subject are divided into theoretical credits and practical credits (Python laboratory) in order to parallel the theoretical knowledge with the development of applications.
In the last weeks of the course, all students will participate in an internal machine learning competition proposed by the teachers.

STUDY LOAD

Type	Hours	Percentage
Hours small group	10,0	20.00
Self study	30,0	60.00
Hours large group	10,0	20.00

Total learning time: 50 h

CONTENTS

1. Introduction to machine learning

Description:
Introduction to machine learning. Supervised and unsupervised learning. Regression and classification

Related activities:
Lab 1. Introduction to Python, Google Colab and Scikit-Learn. Reading of dataframes and exploratory data analysis using the Iris dataset and the Breast Cancer dataset. Simple examples showing how to solve supervised (classification) and unsupervised (dimensionality reduction and clustering) tasks using the Iris dataset

Full-or-part-time: 4h 15m
Theory classes: 2h
Guided activities: 2h
Self study : 0h 15m

2. Regression

Description:
Regression. Models, cost functions and training. Linear and polynomial regression. Gradient descent training. Overfitting, underfitting, and generalization. Regularization: lasso, ridge and elastic net. Validation. Performance measures: MSE, MAP, R-coefficient, Pearson correlation. Challenges in learning associated to the quality of data.

Related activities:
Lab 2. Linear regression and regularization using the MTCars dataset. Closed form solution and solution using gradient descent. Polynomial regression.

Full-or-part-time: 4h 15m
Theory classes: 2h
Guided activities: 2h
Self study : 0h 15m

3. Classification

Description:
Two-class, multi-class and multi-label classification. Logistic regression. Decision boundaries. Cost function and gradient training. Performance measures: accuracy, precision, recall, F-score, confusion matrix, ROC. Regularization.

Related activities:
Lab 3. Logistic regression using the Iris dataset and the Breast Cancer dataset. Confusion matrix and classification metrics. Train-test splits. Presentation of the machine learning competition.

Full-or-part-time: 4h 15m
Theory classes: 2h
Guided activities: 2h
Self study : 0h 15m

4. Neural networks

Description:
Neural networks. Structures. Regression/classification. Cost functions. Backpropagation algorithm. Gradient descent techniques. Batch training. Overfitting and generalization. Regularization and overfit avoidance. RNN and DNN. Rules of a machine learning competition. K-folding validation.

Related activities:
Lab 4. Multi-layer neural network classifier. Sklearn pipelines. Hyperparameter search using GridSearchCV. Regularization. Application to Iris and MNIST datasets.

Full-or-part-time: 4h 15m
Laboratory classes: 2h
Guided activities: 2h
Self study : 0h 15m

5. Decission trees

Description:
Decision trees. Training, overfitting and pruning. Performance in terms of bias and variance. Random forest and ensamble learning.

Related activities:
Lab 5. Decision trees: training, visualization, regularization. Ensamble classifiers: random forest, bagging, boosting. Voting classifiers (soft and hard). Application to Iris and MNIST datasets.

Full-or-part-time: 4h 15m
Theory classes: 2h
Guided activities: 2h
Self study : 0h 15m

6. Unsupervised learning

Description:
Unsupervised machine learning. K-means and Gaussian Mixture Models. Expectation-Maximization algorithm. Selection of model order: BIC and silhouette methods.

Full-or-part-time: 2h 15m
Theory classes: 2h
Self study : 0h 15m

GRADING SYSTEM

Quizzes: 50%
Competition: 50%

EXAMINATION RULES.

The use of calculators, mobile phones or class notes is not allowed.

BIBLIOGRAPHY

Basic:
- Aurelien Geron. Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. 2. UK: O'Reilly Media, 2022. ISBN 978-1098125974.

Complementary:
- Bishop, C.M. Pattern recognition and machine learning. New York: Springer, 2006. ISBN 0387310738.
- Kuncheva, L.I. Combining pattern classifiers: methods and algorithms [on line]. 2nd ed. Hoboken (NJ): J. Wiley & Sons, 2014 [Consultation: 21/09/2018]. Available on: https://onlinelibrary.wiley.com/doi/book/10.1002/9781118914564. ISBN 9781118914564.
- Hastie, T.; Tibshirani, R.; Friedman, J.H. The elements of statistical learning: data mining, inference and prediction [on line]. 2nd ed. New york [etc.]: Springer, 2009 [Consultation: 18/07/2023]. Available on: https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-0-387-84858-7. ISBN 9780387848570.

RESOURCES

Other resources:
Recommended on-line courses
https://es.coursera.org/learn/machine-learning? />Open ML competitions
http://www.kdd.org/kdd-cup />http://www.kaggle.com/competitions />https://datahack.analyticsvidhya.com/contest/all/ />