MGT-502

Data science and machine learning

Vlachos Michail

Cursus	Sem.	Type
Managmt, dur et tech	MA2	Obl.

Language of teaching	English
Credits	5
Withdrawal	Unauthorized
Session	Summer
Semester	Spring
Exam	During the semester
Workload	150h
Weeks	14
Hours	**5 weekly**
Courses	3 weekly
Exercises	2 weekly
Number of positions	**40**

It is not allowed to withdraw from this subject after the registration deadline.

Summary

Hands-on introduction to data science and machine learning. We explore recommender systems, generative AI, chatbots, graphs, as well as regression, classification, clustering, dimensionality reduction, text analytics, neural networks. The course consists of lectures and coding sessions using Python.

Content

- Basic regression methods. Predicting numeric values.
- Basic classification methods. Predicting categorical values: logistic regression, k-NN classification, decision trees, random forests
- Fundamental concepts: cost function and optimization, gradient descent, K-fold cross-validation, overfitting, model calibration, confusion matrix
- Unsupervised learning: Dimensionality reduction, Clustering, Association Rules
- Neural networks and Deep Learning
- Text analytics: text representation, sentiment classification, similarity search
- Recommender Systems
- Graph Analytics: Pagerank, centrality, 6-degrees of separation
- Generative AI and chatBots.

Keywords

AI, Machine learning, Data Science Algorithms, Recommender Systems, Graphs, Generative AI, Regression, Classification, Dimensionality reduction, Clustering, Neural networks, Text analytics, Python

Learning Prerequisites

Required courses

Statistics and data science (MGT-499)

Important concepts to start the course

- Basic Probability and Statistics knowledge (random variables, expectation, mean, conditional and joint distribution, independence, Bayes rule, central limit theorem)
- Basic linear algebra (matrix/vector multiplication, system of linear equations)
- Multivariate calculus (derivative w.r.t. vector and matrix variables)
- Basic programming skills (Python)

Learning Outcomes

By the end of the course, the student must be able to:

- Describe the principal types of machine learning algorithms
- Investigate data, data types, and problems with the data
- Choose an appropriate Machine Learning method for a given task
- Implement Machine Learning algorithms in Python
- Optimize the main tradeoffs such as overfitting and computational cost vs accuracy
- Conduct a Data Science project

Transversal skills

- Plan and carry out activities in a way which makes optimal use of available time and other resources.
- Demonstrate the capacity for critical thinking
- Access and evaluate appropriate sources of information.
- Use a work methodology appropriate to the task.

Teaching methods

- Lectures
- Lab sessions: coding exercices
- Data Science projects

Expected student activities

The students are expected to:

- attend lectures and lab sessions;
- work on the weekly theory and coding exercises;
- complete assignments (graded);
- conduct data science projects making use of the theory learned during lectures and code developed during lab sessions (graded)

Assessment methods

- Personal coding assignments: 25%
- Group Project: 25%
- Final Quiz: 50%

Supervision

Office hours	No

Assistant.e.s	Yes
Forum	No
Others	Slack channel

Resources

Virtual desktop infrastructure (VDI)

No

Bibliography

-

Ressources en bibliothèque

- Find the references at the Library

Notes/Handbook

Slides will be made available on the course Moodle page. Notebooks will be made available in a GitHub repository.

Moodle Link

- https://go.epfl.ch/MGT-502