

**Introduction to machine learning and artificial intelligence for
material science, 7.5 credits**

Introduktion till maskininlärning och artificiell intelligens för
materialvetenskap, 7.5 hp

Third-cycle education course

6FIFMA6

Department of Physics, Chemistry and Biology

Valid from: Second half-year 2025

Approved by

Approved

Registration number

Entry requirements

- Completion of 240 credits in required courses, including at least 60 second-cycle credits, or equivalent knowledge acquired through other means.
- Basic knowledge of mathematics, including topics such as statistics, linear algebra, and probability theory.
- Some basic familiarity with programming. A significant part of the course involves Python, wherefore previous experience in Python programming is advantageous. However, the course begins with a module that provides a crash course in Python and scientific programming in Python, ensuring that all students get the opportunity to acquire the necessary foundation to follow the rest of the course.

Specific information

The course is designed for PhD students (and potentially master's students) with a background in chemistry, materials science, physics, or related fields, but with little to no exposure to machine learning and artificial intelligence. The aim is to provide a strong introduction to these fields, equipping students with the knowledge and confidence needed to begin applying ML/AI tools to solve problems in their research. The course combines theory with hands-on coding exercises and covers best practices for using Python to tackle real-world data science problems through ML/AI. While the content is broadly applicable, illustrative examples will, when relevant, be drawn from materials science and chemistry.

Learning outcomes

At the end of the course, the students should:

- Be familiar with common classical machine learning techniques.
- Be familiar with various types of neural networks.
- Be familiar with a few more advanced concepts, such as Bayesian optimization and large language models.
- Be familiar with how to access materials data from open online repositories.
- Based on a given problem, be able to make an intelligible choice of which techniques/algorithms to use.
- Be able to write python code for using common machine learning algorithms to solve regression, classification, and clustering problems.
- Be able to use Python to solve machine learning problems relevant to material science related research.

Contents

The course is structured into four modules.

1. Introduction to Python.

This short module provides a crash course in Python, covering how to set up your environment and the programming concepts needed to start using machine learning. For students with no prior Python experience, this module offers the essential foundation required to follow the rest of the course.

****2) Classical machine learning. ****

This module introduces fundamental machine learning algorithms for regression, classification, clustering, and dimensionality reduction. Algorithms covered include linear regression, logistic regression, SoftMax, k-nearest neighbors, support vector machines, naïve Bayes, decision trees, random forests, principal component analysis, K-means, DBSCAN, and more. Alongside algorithmic discussions, we will explore data handling, model training, cost functions, regularization, and model evaluation. Python frameworks such as NumPy, Pandas, SciPy, and Scikit-learn will be introduced in this module.

****3) Neural networks. ****

This module focuses on the basics of neural networks, including components, activation functions, training processes, network design, and regularization techniques. Common network architectures such as feedforward networks, convolutional networks, and autoencoders will be explored. In this module, we will introduce Python frameworks like TensorFlow and Keras.

****4) Advanced topics. ****

The final module introduces a few more advanced topics such as Bayesian optimization, graph neural networks, explainable AI (including SHAP analysis), generative models, transformers, and large language models. The module also covers materials informatics and feature engineering with a focus on materials and molecular features. Python frameworks related to materials informatics, such as RDKit and Pymatgen, will be introduced in this section.

Educational methods

The course is structured to include:

- 15 lectures
- 4 computer lab sessions with hands-on exercises to implement concepts from the lectures using Python
- 1 mandatory end-of-course seminar for presentation of the course final project.

Examination

The examination consists of two components:

1. **Programming Assignments:** Students will submit assignments based on the programming exercises completed during the course.
2. **Final Project:** Students will choose a project topic in consultation with the course instructor, ideally related to their own research. The project will be evaluated through a written report and an oral presentation.

Grading

Two-grade scale

Course literature

The main book will be: Hands-On- Machine Learning with Scikit-Learn, Keras, and Tensorflow, second edition, by Aurélien Géron

In addition to the main book, various complementary sources will be recommended during the course.

General information

The course is planned and carried out according to what is stated in this syllabus.

Course evaluation, analysis and suggestions for improvement should be fed back to the Research and PhD studies Committee (FUN) by the course coordinator.

If the course is withdrawn or is subject to major changes, examination according to this syllabus is normally offered at three occasions within/in close connection to the two following semesters.