GHENT UNIVERSITY

Course Specifications

Valid as from the academic year 2024-2025

Machine Learning for Life Sciences (I002932)

Course size *(nominal values; actual values may depend on programme)*

Credits 5.0 **Study time 150 h**

Course offerings in academic year 2024-2025

A (semester 1) English Gent

Lecturers in academic year 2024-2025

Waegeman, Willem	LA26	lecturer-in-charge

Offered in the following programmes in 2024-2025	crdts	offering
Master of Science in Bioinformatics(main subject Bioscience Engineering)	5	A
Master of Science in Bioinformatics(main subject Systems Biology)	5	A
Master of Science in Bioscience Engineering: Cell and Gene Biotechnology	5	A
Master of Science in Bioscience Engineering: Environmental Technology	5	A
Master of Science in Bioscience Engineering: Forest and Nature Management	5	A
Master of Science in Bioscience Engineering: Land, Water and Climate	5	A
Master of Science in Pharmaceutical Engineering	5	A
Exchange Programme in Bioinformatics (master's level)	5	A
Exchange Programme in Bioscience Engineering: Agricultural Sciences (master's level)	5	A
Exchange Programme in Bioscience Engineering: Cell and Gene Biotechnology (master's level)	5	A
Exchange Programme in Bioscience Engineering: Chemistry and Bioprocess Technology (master's level)	5	A
Exchange Programme in Bioscience Engineering: Environmental Technology (master's level)	5	A
Exchange Programme in Bioscience Engineering: Food Science and Nutrition (master's level)	5	A
Exchange Programme in Bioscience Engineering: Land and Forest management (master's level)	5	A

Teaching languages

English

Keywords

Classification, regression, advanced prediction problems, applications in the life sciences

Position of the course

This course gives the students an introduction to the field of predictive modelling (aka machine learning), aiming at the recognition and prediction of complex patterns in data. Predictive computational and statistical models are needed in many applications in bioinformatics and the life sciences in general. Think in this context at the prediction of diseases from genetic data, the forecast of natural events from climate data, the modelling of ecological migration from geographic data, etc.
Both classification methods (output = class label) and regression methods (output = real value) will be discussed thoroughly. Moreover, more involved prediction problems are touched upon as well, such as learning structured objects, multi-label classification, ordinal regression, etc.
The students will apply the considered machine learning methods by means of existing software on several case studies that are situated in the life sciences. In the theory part as well, the course instructors are specifically focusing on predictive methods for the life sciences, by dealing with subjects such as predicting from high-dimensional and non-vectorial data.

Contents

The course focuses on introducing machine learning principles, techniques and applications.

Every method will be practiced during computer exercise sessions on real biological problems.

1 Introduction to machine learning (classification vs. regression, training vs. test data, overfitting, cross validation, performance measures, modelling high-dimensional and non-vector data, complexity control and regularization, relation with bias-variance decomposition.
2 Overfitting, the bias-variance trade-off, nearest neighbour methods
3 Linear regression revisited (parameter fitting of least-squares systems, assumptions, shortcomings)
4 Basic classification methods (least squares classification, logistic regression, linear discriminant analysis).
5 Evaluating machine learning methods (cross-validation and bootstrap, bias in experimental design, etc.)
6 Linear model selection methods (variable selection, ridge regression, lasso)
7 Feature expansions, polynomial regression and kernel ridge regression
8 Tree-based methods (decision trees, random forests)
9 Support vector machines and modelling of structured and non-vectorial data (text, DNA sequences, graphs, images, spectral data, etc.)
10 Unsupervised learning (principal component analysis and clustering)
11 Neural networks and the backpropagation algorithm
12 Neural network architectures (convolutional neural networks, recurrent neural networks, LSTMs, transformer models)
13 Advanced prediction problems (multidimensional prediction, structured prediction, graph learning, semi-supervised learning, etc.)

Initial competences

It is important that the students have already hands-on experience with programming (Matlab, Python, R, etc.). We will use Python in the PC-classes.
Basic knowledge of mathematics, informatics, probability and statistics is recommended, in particular the following topics:
- vector and matrix algebra
- least squares problems
- singular value decomposition
- quadratic forms
- the gradient and partial derivatives
- extreme values of functions
- Lagrange multipliers
- Bayes' rule
- probability distributions (in particular the Gaussian distribution)
- linear regression

Final competences

1 The student must be able to:
1. select the most appropriate method for a given classification or regression problem;

2 2. apply these methods and interpret the results;

3 3. understand recent literature in machine learning, process and apply the methods presented in these articles.

Conditions for credit contract

Access to this course unit via a credit contract is determined after successful competences assessment

Conditions for exam contract

This course unit cannot be taken via an exam contract

Teaching methods

Seminar, Lecture, Independent work, Peer teaching

Study material

Type: Handbook

Name: Introduction to Statistical Learning, with applications in Python
Indicative price: Free or paid by faculty
Optional: yes
Language : English
Author : James Gareth
Number of Pages : 300

Oldest Usable Edition : 2023
Online Available : Yes
Available in the Library : No
Available through Student Association : No
Usability and Lifetime within the Course Unit : intensive
Usability and Lifetime within the Study Programme : intensive
Usability and Lifetime after the Study Programme : regularly

References

James et al. An introduction to statistical learning, Springer 2013.

Course content-related study coaching

1. The lecturer announces office hours for problems related to the theory.
2. The teaching assistant guides the practical exercises.

Assessment moments

end-of-term and continuous assessment

Examination methods in case of periodic assessment during the first examination period

Oral assessment, Participation, Assignment

Examination methods in case of periodic assessment during the second examination period

Oral assessment, Participation, Assignment

Examination methods in case of permanent assessment

Oral assessment

Possibilities of retake in case of permanent assessment

examination during the second examination period is possible in modified form

Extra information on the examination methods

Project 1: Weight 50%
Building and evaluating machine learning models on classical tabular datasets. Report + Oral defense
Project 2: Weight 50%
Building and evaluating machine learning models on non-tabular datasets. Report + Oral defense

Calculation of the examination mark

Students who eschew period aligned and/or non-period aligned evaluations for this course unit may be failed by the examinator.