Syllabus
Course Program

Methods of Mathematical Modeling and Data Analysis

Specialty
113 – Applied Mathematics

Institute
Institute of Computer Modeling, Applied Physics and Mathematics

Educational program
Computer and Mathematical Modeling

Department
Mathematical Modeling and Intelligent Computing in Engineering (161)

Level of education
Master's level (1 year 4 months)

Course type
Special (professional), Mandatory

Semester
1

Language of instruction
English

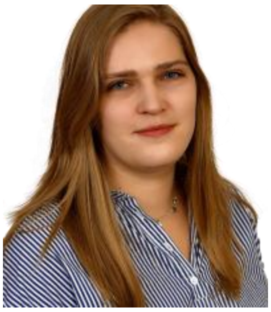
Lecturers and course developers

Oleksiy Larin (responsible lecturer)

Oleksiy.Larin@khpi.edu.ua
Doctor of Engineering Science, Professor,
Director of the Institute of Computer Modeling, Applied Physics and Mathematics

Specialist in computational modelling, in particular probabilistic, and reliability prognosis. The main focus of scientific works deals with the development of models, methods and algorithms of computer simulations based on classical numerical and ML-based approaches with applications to engineering systems. Author of more than 150 scientific and methodical works

General information, number of publications, main courses, etc.
More about the lecturer on the department's website



Kseniia Potopalska (assistant)

Kseniia.Potopalska@khpi.edu.ua
PhD in Engineering, Masters' in Compute Science,
Assoc. professor of the department of Mathematical modeling and intelligent computing in engineering

General information, number of publications, main courses, etc.
More about the lecturer on the department's website

General information

Summary

The course offers a comprehensive journey into the world of data analysis and machine learning. The course's structure includes coding workshops and in-depth theoretical insights in equal proportion. Students' progress is assessed through individual programming assignments and comprehensive reports, reflecting their competence in implementing machine learning algorithms and mathematical modeling techniques and will end with an oral exam. The course prepares individuals to excel in data-driven

careers by fostering a deep knowledge of mathematical techniques and hands-on skills in machine learning applications.

Course objectives and goals

The course aims to equip students with a strong mathematical foundation of the modern method for data analysis and modeling, with a specific focus on machine learning techniques.
The course objectivities are:
1) to enhance the understanding of the mathematical foundations of the main ML methods and algorithms
2) to improve the skills in algorithm realization (coding) and the ability to choose a better computational approach for real-world practical issues.
3) to improve the understanding of the different ML algorithm limitations of usage and the directions of the challenges overcoming

Format of classes

Lectures, laboratory classes, consultations, self-study. Final control in the form of an exam.

Competencies

GC3. Ability to master modern knowledge, formulate and solve problems.
GC5. Ability to conduct professional activities, in particular in the international environment.
GC7. Ability to think abstractly, analyse and synthesise.
PC2. Ability to conduct scientific research aimed to develop new and adapt existing mathematical and computer models to study various processes, phenomena and systems, conduct appropriate experiments and analyse the results.
PC4. Ability to develop and research mathematical and computer models, conduct computational experiments and solve formalised problems using specialised software.
PC5. Ability to build and research models for decision-making using intelligent systems.
PC6. Ability to apply artificial intelligence methods, develop and implement machine learning algorithms in practice.
PC7. Ability to design and develop software to solve formalised problems, including systems with large amounts of data.
PC8. Ability to formalise and build data or knowledge models, obtain relevant knowledge from large amounts of data, choose data mining methods to solve problems.
PC9. The ability to mathematically formalise the formulation of scientific and practical problems, to choose a mathematical analytical or numerical method of its solution, which ensures the required accuracy and reliability of the result.

Learning outcomes

LO1. Communicate within the scope of professional competences in one of the EU languages.
LO5. Justify and, if necessary, develop new algorithms and software tools for solving scientific and applied problems, apply, modify and investigate analytical and computational methods for solving them.
LO6. Apply procedures for formal description of systems, checking their adequacy for the study of socio-economic, technical, natural and other systems.
LO8. Develop and implement algorithms for solving applied problems, system and application software of information systems and technologies.
LO9. Be able to analyse and design systems with large amounts of data, apply and adapt methods of knowledge acquisition, methods of evaluation and interpretation of the found patterns.
LO10. Develop and apply modern concepts of machine learning and data mining.
LO11. Possess skills of abstract thinking, analysis and synthesis.

Student workload

The total volume of the course is 150 hours (5 ECTS credits): lectures - 32 hours, laboratory classes - 32 hours, self-study - 86 hours.

Course prerequisites

For a general understanding of lectures, it is assumed that students have a background in the following disciplines:

Mathematical Analysis (Differential calculus and Series);
Linear Algebra (matrix, vector algebra),
Numerical Methods and the Theory of Optimization basics;
Theory of Probability and Mathematical Statistics.

Skills in basic programming is required;
Experience in Python/MatLab coding is welcomed but not mandatory.

Features of the course, teaching and learning methods, and technologies

The course , teaching and learning methods include traditional lectures for theoretical foundations, hands-on workshops for practical coding and problem-solving, self-study for reinforcement, and individual report defenses in workshops. These approaches ensure students receive a well-rounded education in mathematical modeling and computer science, promoting both theoretical knowledge and practical skills.

Program of the course

Topics of the lectures

Topic 1. General mathematical formulation of the machine learning (ML) problem. Basic concepts and definitions

a. What is a description of features? Types of objects features in machine learning tasks.
b. Basic types of machine learning tasks (variants of the output)
c. The main stages of solving the problem of machine learning

Topic 2. The problem of machine learning - examples of practical applications

a. Problems in solving the problem of machine learning
b. The main elements of the CRISP-DM standard

Topic 3. Problem of overfitting. Validation of the models and their types.

a. Means of quality control in machine learning.
b. Hold-out validation (The meaning of the procedure. Disadvantages / limitations).
c. Leave-one-out validation (The meaning of the procedure. Disadvantages / limitations).
d. Cross-validation (The meaning of the procedure. Disadvantages / limitations).

Topic 4. Metric methods of classification (recognition)

a. Compactness hypothesis
b. The concept of distance, types and examples.

Topic 5. The method of nearest neighbors kNN

a. General statement of the algorithm
b. Problems and challenged/limitations.
c. Parzen window method as a modification of kNN. Types of kernel-functions.

Topic 6. Nonparametric regression

a. The meaning of the problem and its mathematical formulation
b. The Nadaraya-Watson formula
c. The idea of a locally weighted scatterplot smoothing (LOWESS) procedure

Topic 7. Linear classifier. General mathematical formulation.

a. How mathematically the prediction model is set on
b. The loss-function
c. The concept of margin
d. The problem of the loss function as a function of the margin. Some basic of variants of its analytical formulation of this function.

Topic 8. Stochastic gradient algorithm. Robinson-Monroe Procedure (SGM)

Topic 9. Support vector machine (SVM). Basic statement.

a. Geometric interpretation.

b. Basic formulations and proofness.
c. Extension for the case of linearly non-separable dataset. The concept of soft margin.

Topic 10. Support vector machine (SVM). Advanced issues.
a. Duality of interpretation SVM as linear classifier.
b. Generalization to a nonlinear classifier.

Topic 11. Logistic Regression

Topic 12. Assessment of the quality of metric methods of classification
a. Accuracy. Interpretation and limitations for this quality criterion
b. Types of errors in the classifier problem. Matrix of Errors
c. Precision
d. Recall
e. Averaging by quality criteria. F-measure

Topic 13. The decision trees.
a. General statement
b. Learn ID3 algorithm
c. Criteria of dataset division (information gain, information gain ratio)
d. Entropy and information growth. The role of information growth in the algorithm
e. Advantages/ disadvantages and limitations the Learn ID3 algorithm

Topic 14. Major modifications of the Learn ID algorithm. Algorithm C4.5.
a. Reducing the "greediness" of the algorithm
b. Generalization in case of continuous attributes
c. Solve the problem of missing data

Topic 15. Trees pruning
a. Pre-pruning
b. Post-pruning

Topic 16. Random forest
a. General statement
b. Boosting and bagging

Topics of the workshops

Topic 1. Introduction to Python. Data preparation in Pandas

Topic 2. Data engineering

Topic 3. kNN-method.

Topic 4. Linear Regression

Topic 5. Support vector machine (SVM). Part 1
The 1st part of assignment in SVM topic is assumed to limits the students usage of sklearn library.

Topic 6. Support vector machine (SVM). Part 2
The 2nd part of assignment in SVM topic shows the possibilities of sklearn library.

Topic 7. Learn ID3 method.
The assignments restricted for the manual coding of the algorithm from the sketch

Topic 8. C4.5 method.

Topic 9. Logistic regression.

Topics of the laboratory classes

In this course the laboratory classes are given as a programming (coding) assessment in Python language within a topics that provided in the workshops section of this syllabus.

Self-study

There are no special individual or group assignments for self-studying in the course, just the mentioned list of programming assignments. The self-study time should be spent on the repetition of the theoretical materials given in the lectures and on the preparation of the programming (coding) assignments. Here it should be claimed that the workshop time is mainly given for consultations with course assistants and for individual defending of the reports for each programming assignment, but not for their completion which is assumed to be done within a self-study.

Course materials and recommended reading

[1] Andreas C. Müller and Sarah Guido Introduction to Machine Learning with Python. A Guide for Data Scientists. O'Reilly Media Inc., 2016, 385 p.
[2] Wes McKinney Python for Data Analysis. O'Reilly Media Inc., 2013, 470 p.
[3] Kevin P. Murphy Machine Learning: A Probabilistic Perspective. MIT Press, 2012, 1096 p.
[4] Trevor Hastie, Robert Tibshirani, and Jerome Friedman The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, 2016, 767 p.
[5] Sebastian Raschka, Vahid Mirjalili "Python Machine Learning" 2019, 772p.
[6] Christoph Molnar "Interpretable Machine Learning" 2023, 626p.

Assessment and grading

Criteria for assessment of student performance, and the final score structure

Theoretical part:
30 points in the form of an oral exam: 2 questions in the exam paper (12 points each)+ additional ones.
Practical part of the course:
70 points. The assessment is a recommendation given by course assistant form the results of programming (codding) assessment, i.e. 9 Programing practice should be done in total with the different weight of each (from 5 to 15 point). The maximum points for the Programing practice is 65.

5 extra points the student could obtain by some advanced achievement either on the Practice or during the Exam.

Grading scale

Total points	National	ECTS
90–100	Excellent	A
82–89	Good	B
75–81	Good	C
64–74	Satisfactory	D
60–63	Satisfactory	E
35–59	Unsatisfactory (requires additional learning)	FX
1–34	Unsatisfactory (requires repetition of the course)	F

Norms of academic integrity and course policy

The student must adhere to the Code of Ethics of Academic Relations and Integrity of NTU "KhPI": to demonstrate discipline, good manners, kindness, honesty, and responsibility. Conflict situations should be openly discussed in academic groups with a lecturer, and if it is impossible to resolve the conflict, they should be brought to the attention of the Institute's management.
Regulatory and legal documents related to the implementation of the principles of academic integrity at NTU "KhPI" are available on the website: http://blogs.kpi.kharkov.ua/v2/nv/akademichna-dobrochesnist/

Approval

Approved by	Date	
	Date August 30, 2023	Head of the department Oleksii VODKA
	Date August 30, 2023	Guarantor of the educational and professional program (1 year 4 months) Oleksiy LARIN