Federal State Autonomous Educational Institution of Higher Education "Moscow Institute of Physics and Technology (National Research University)"

APPROVED
Head of the Phystech School of Applied Mathematics and Informatics
A.M. Raygorodskiy

Work program of the course (training module)

course: Machine Learning/Машинное обучение
major: Information Science and Computer Engineering
specialization: Computer Science/Информатика
Phystech School of Applied Mathematics and Informatics
Chair of Machine Learning and Digital Humanities
term: 4
qualification: Bachelor

Semesters, forms of interim assessment:
7 (fall) - Grading test
8 (spring) - Exam

Academic hours: 120 AH in total, including:
lectures: 60 AH.
seminars: 0 AH.
laboratory practical: 60 AH.

Independent work: 255 AH.

Exam preparation: 30 AH.

In total: 405 AH, credits in total: 9

Author of the program: V.V. Yakovlev, candidate of physics and mathematical sciences

The program was discussed at the Chair of Machine Learning and Digital Humanities 04.06.2020

Annotation

To learn data analysis concepts and methods

1. Study objective

Purpose of the course

To learn data analysis concepts and methods.

Tasks of the course

- Learn mathematical basics of data analysis;
- learn programming tools for data processing and analysis.

2. List of the planned results of the course (training module), correlated with the planned results of the mastering the educational program

Mastering the discipline is aimed at the formation of the following competencies:

Code and the name of the competence	Competency indicators
Gen.Pro.C-3 Write scientific and/or technical (technological, innovative) reports (publications, projects)	Gen.Pro.C-3.1 Meet the general criteria for submission of manuscripts, scientific and technical documentation and use relevant software applications
	Gen.Pro.C-3.2 Employ practical methodologies for preparing scientific and technical reports (projects)
	Gen.Pro.C-3.3 Visually and graphically present scientific (scientific and technical, innovative technological) outcomes in the form of reports, scientific publications
Gen.Pro.C-4 Collect and process scientific and technical and/or technological data for fundamental and applied problem-solving	Gen.Pro.C-4.1 Apply scientific research and intellectual analysis methods for professional problem-solving
	Gen.Pro.C-4.2 Search for primary sources of scientific and technical and/or technological information in professional settings
	Gen.Pro.C-4.3 Prepare abstracts, reports, bibliographies, and reviews of information in professional settings
	Gen.Pro.C-4.4 Use computer and network skills to obtain, store, and process scientific (technical, technological) information
Pro.C-2 Conduct scientific research and testing independently or as a member (leader) of a small research team	Pro.C-2.1 Apply the principles of scientific work, methods of collecting and analyzing the obtained data and ways of argumentation
	Pro.C-2.2 Conduct scientific research independently or as a member (leader) of a small research team
	Pro.C-2.3 Present research results through scientific publications and participation in conferences

3. List of the planned results of the course (training module)

As a result of studying the course the student should:

know:

- mathematical foundations of data analysis.

be able to:

- To prepare source data$;
- to find key features;
- to choose proper method to process.

master:

- Data processing methods and tools;
- data analysis methods and tools.

4. Content of the course (training module), structured by topics (sections), indicating the number of allocated academic hours and types of training sessions

4.1. The sections of the course (training module) and the complexity of the types of training sessions

№	Topic (section) of the course	Types of training sessions, including independent work			
		Lectures	Seminars	Laboratory practical	Independent work
1	Basic concepts. Python	10		10	40
2	Quality of classification. Generalizing ability	10		10	40
3	Linear models	10		10	40
4	Regression. Linear regression	15		15	67
5	Artificial neural networks	15		15	68
AH in total		60		60	255
Exam preparation		30 AH.			
Total complexity		405 AH., credits in total 9			

4.2. Content of the course (training module), structured by topics (sections)

Semester: 7 (Fall)

1. Basic concepts. Python

Basic concepts: objects, features, answers, supervised and unsupervised learning.
Python Classification. Simplest classification algorithms
Python. Scikit-learn. Matplotlib

2. Quality of classification. Generalizing ability

Quality of classification. Generalizing ability
Overfitting and underfitting. Feature selection
Decision tree. Decision tree ensemble

3. Linear models

Linear models
Regularization
Support Vector Machine

Semester: 8 (Spring)

4. Regression. Linear regression

Regression. Linear regression
Quality of regression. Regularization
Dimensionality reduction techniques

5. Artificial neural networks

Artificial neural networks

Unsupervised learning. Clustering
Quality of clustering

5. Description of the material and technical facilities that are necessary for the implementation of the educational process of the course (training module)

- A projector with the ability to connect via HDMI and / or VGA);
- blackboard with chalk or whiteboard with felt-tip pens;
- computer class equipped with a PC.

6. List of the main and additional literature, that is necessary for the course (training module) mastering

Main literature

Additional literature

7. List of web resources that are necessary for the course (training module) mastering

Not used

8. List of information technologies used for implementation of the educational process, including a list of software and information reference systems (if necessary)

The lecture classes use multimedia technology, including the presentation of presentations. Electronic books are used.

9. Guidelines for students to master the course

A student studying a discipline must, on the one hand, master the general conceptual apparatus, and on the other hand, must learn to put theoretical knowledge into practice.

As a result of studying the discipline, the student must know the basic definitions, concepts, axioms.

Successful development of the course requires intense independent work of the student. The course program provides the minimum necessary time for the student to work on the topic. Independent work includes:

- reading and taking notes of recommended literature;
- study of educational material (according to lecture notes, educational and scientific literature), preparation of answers to questions intended for independent study, proof of individual statements, properties;
- laboratory work to understand the connections between theory and practical skills;
- preparation for exam.

The management and control of the student's independent work is carried out in the form of individual consultations.

It is important to gain an understanding of the material being studied, and not its mechanical memorization. If it is difficult to study individual topics, questions, you should consult a lecturer.

SUPPLEMENT

Assessment funds for course (training module)

major:	Information Science and Computer Engineering
specialization:	Computer Science/Информатика
	Phystech School of Applied Mathematics and Informatics
	Chair of Machine Learning and Digital Humanities
term:	4
qualification:	Bachelor

Semesters, forms of interim assessment:

7 (fall) - Grading test

8 (spring) - Exam

Author: V.V. Yakovlev, candidate of physics and mathematical sciences

1. Competencies formed during the process of studying the course

Code and the name of the competence	Competency indicators
Gen.Pro.C-3 Write scientific and/or technical (technological, innovative) reports (publications, projects)	Gen.Pro.C-3.1 Meet the general criteria for submission of manuscripts, scientific and technical documentation and use relevant software applications
	Gen.Pro.C-3.2 Employ practical methodologies for preparing scientific and technical reports (projects)
	Gen.Pro.C-3.3 Visually and graphically present scientific (scientific and technical, innovative technological) outcomes in the form of reports, scientific publications
Gen.Pro.C-4 Collect and process scientific and technical and/or technological data for fundamental and applied problem-solving	Gen.Pro.C-4.1 Apply scientific research and intellectual analysis methods for professional problem-solving
	Gen.Pro.C-4.2 Search for primary sources of scientific and technical and/or technological information in professional settings
	Gen.Pro.C-4.3 Prepare abstracts, reports, bibliographies, and reviews of information in professional settings
	Gen.Pro.C-4.4 Use computer and network skills to obtain, store, and process scientific (technical, technological) information
Pro.C-2 Conduct scientific research and testing independently or as a member (leader) of a small research team	Pro.C-2.1 Apply the principles of scientific work, methods of collecting and analyzing the obtained data and ways of argumentation
	Pro.C-2.2 Conduct scientific research independently or as a member (leader) of a small research team
	Pro.C-2.3 Present research results through scientific publications and participation in conferences

2. Competency assessment indicators

As a result of studying the course the student should:

know:

- mathematical foundations of data analysis.

be able to:

- To prepare source data$;
- to find key features;
- to choose proper method to process.

master:

- Data processing methods and tools;
- data analysis methods and tools.

3. List of typical control tasks used to evaluate knowledge and skills

Not provided.

4. Evaluation criteria

For differential credit:

1. Basic concepts: objects, features, answers, supervised and unsupervised learning.
2. Python Classification. Simplest classification algorithms
3. Python. Scikit-learn. Matplotlib
4. Quality of classification. Generalizing ability
5. Overfitting and underfitting. Feature selection

6. Decision tree. Decision tree ensemble
7. Linear models
8. Regularization
9. Support Vector Machine
10. Regression. Linear regression
11. Quality of regression. Regularization
12. Dimensionality reduction techniques
13. Artificial neural networks
14. Unsupervised learning. Clustering
15. Quality of clustering

For exam:
1. Vector representations of words. Classic approaches. Word2Vec.
2. Machine translation without parallel text corpora (Unsupervised translation).
3. RNN in word processing: main ideas and existing limitations.
4. CNN in word processing: main ideas and existing limitations.
5. The mechanism of attention: principles of work.
6. Self-attention mechanism, differences from the usual attention mechanism.
7. Transformer architecture: description, work ideas.
8. Assessment of quality in machine translation.
9. Pretrained embeddings, ELMo.
10. BERT: principle of operation, architecture and teaching features.
11. Statement of the problem of reinforcement learning.
12. Basic concepts of reinforcement learning.
13. Method of cross entropy.
14. Genetic algorithms.
15. The value iteration and policy iteration methods.
16. Model of the environment. Value function, q-function.
17. Q-learning in the case of a continuous action space. Autocorrelation problem.
18. ALGORITHM REINFORCE. Log-derivative trick. Policy gradient.
19. Improvements to the policy gradient method. Baseline concept. Advantage Actor Critic.
20. Self-critical Sequence training in the text generation task.
21. The main tasks of computer vision.
22. Assessment of quality in tasks of computer vision.
23. Basic open datasets of labeled data for the problem of computer vision.
24. The main features of the YOLO architecture (1, 2, 3, 4).
25. Main features of R-CNN, Fast (er) R-CNN architectures.
26. Methods of increasing the resolution (upsampling).
27. Variational autoencoder (VAE): structure, differences from the usual autoencoder (AE).
28. Generative adversarial networks (GAN): basic ideas, structure, teaching methods.

Ticket 1.
1. RNN in word processing: main ideas and existing limitations.
2. The main tasks of computer vision.
3. The REINFORCE algorithm. Log-derivative trick. Policy gradient.

Assessment “excellent (10)” is given to a student who has displayed comprehensive, systematic and deep knowledge of the educational program material, has independently performed all the tasks stipulated by the program, has deeply studied the basic and additional literature recommended by the program, has been actively working in the classroom, and understands the basic scientific concepts on studied discipline, who showed creativity and scientific approach in understanding and presenting educational program material, whose answer is characterized by using rich and adequate terms, and by the consistent and logical presentation of the material;

Assessment “excellent (9)” is given to a student who has displayed comprehensive, systematic knowledge of the educational program material, has independently performed all the tasks provided by the program, has deeply mastered the basic literature and is familiar with the additional literature recommended by the program, has been actively working in the classroom, has shown the systematic nature of knowledge on discipline sufficient for further study, as well as the ability to amplify it on one’s own, whose answer is distinguished by the accuracy of the terms used, and the presentation of the material in it is consistent and logical;

Assessment “excellent (8)” is given to a student who has displayed complete knowledge of the educational program material, does not allow significant inaccuracies in his answer, has independently performed all the tasks stipulated by the program, studied the basic literature recommended by the program, worked actively in the classroom, showed systematic character of his knowledge of the discipline, which is sufficient for further study, as well as the ability to amplify it on his own;

Assessment “good (7)” is given to a student who has displayed a sufficiently complete knowledge of the educational program material, does not allow significant inaccuracies in the answer, has independently performed all the tasks provided by the program, studied the basic literature recommended by the program, worked actively in the classroom, showed systematic character of his knowledge of the discipline, which is sufficient for further study, as well as the ability to amplify it on his own;

Assessment “good (6)” is given to a student who has displayed a sufficiently complete knowledge of the educational program material, does not allow significant inaccuracies in his answer, has independently carried out the main tasks stipulated by the program, studied the basic literature recommended by the program, showed systematic character of his knowledge of the discipline, which is sufficient for further study;

Assessment “good (5)” is given to a student who has displayed knowledge of the basic educational program material in the amount necessary for further study and future work in the profession, who while not being sufficiently active in the classroom, has nevertheless independently carried out the main tasks stipulated by the program, mastered the basic literature recommended by the program, made some errors in their implementation and in his answer during the test, but has the necessary knowledge for correcting these errors by himself;

Assessment “satisfactory (4)” is given to a student who has discovered knowledge of the basic educational program material in the amount necessary for further study and future work in the profession, who while not being sufficiently active in the classroom, has nevertheless independently carried out the main tasks stipulated by the program, learned the main literature but allowed some errors in their implementation and in his answer during the test, but has the necessary knowledge for correcting these errors under the guidance of a teacher;

Assessment “satisfactory (3)” is given to a student who has displayed knowledge of the basic educational program material in the amount necessary for further study and future work in the profession, not showed activity in the classroom, independently fulfilled the main tasks envisaged by the program, but allowed errors in their implementation and in the answer during the test, but possessing necessary knowledge for elimination under the guidance of the teacher of the most essential errors;

Assessment “unsatisfactory (2)” is given to a student who showed gaps in knowledge or lack of knowledge on a significant part of the basic educational program material, who has not performed independently the main tasks demanded by the program, made fundamental errors in the fulfillment of the tasks stipulated by the program, who is not able to continue his studies or start professional activities without additional training in the discipline in question;

Assessment “unsatisfactory (1)” is given to a student when there is no answer (refusal to answer), or when the submitted answer does not correspond at all to the essence of the questions contained in the task.

5. Methodological materials defining the procedures for the assessment of knowledge, skills, abilities and/or experience

The exam time is 2 academic hours. During the differential test, students can use the discipline program and source texts.