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

APPROVED
Head of the Phystech School of Biological and Medical Physics
D.V. Kuzmin

Work program of the course (training module)

course: Deep Learning in Biology/Глубокое обучение в биологии
major: Applied Mathematics and Physics
specialization: Applied Bioinformatics/Прикладная биоинформатика
Phystech School of Biological and Medical Physics
Chair of Bioinformatics and Systems Biology
term: 1
qualification: Master

Semester, form of interim assessment: 2 (spring) - Exam

Academic hours: 60 AH in total, including:
lectures: 30 AH.
seminars: 30 AH.
laboratory practical: 0 AH.

Independent work: 45 AH.

Exam preparation: 30 AH.

In total: 135 AH, credits in total: 3

Author of the program: A.A. Zverkov

The program was discussed at the Chair of Bioinformatics and Systems Biology 04.06.2020

Annotation

The aim of the course is to master the methods and theory of deep learning for solving applied problems, including in the field of computer vision and natural language processing. After mastering the course, the student will understand the theory of machine and deep learning (concepts, axioms, methods), modern problems of the corresponding sections of deep learning, basic libraries and software for the implementation of their own or typical projects using deep learning.

1. Study objective

Purpose of the course

Mastering the methods and theory of deep learning for solving applied problems, including in the field of computer vision and natural language processing.

Tasks of the course

Mastering by students of knowledge (concepts, concepts, methods and models) in the field of deep learning;

Acquisition of theoretical knowledge and practical skills in the field of deep learning for solving applied problems, including the necessary knowledge of software development;

Providing advice and assistance to students in conducting their own research in the field of deep learning.

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
UC-1 Use a systematic approach to critically analyze a problem,and develop an action plan	UC-1.1 Systematically analyze the problem situation, identify its components and the relations between them
	UC-1.2 Search for solutions by using available sources
	UC-1.3 Develop a step-by-step strategy for achieving a goal, foresee the result of each step, evaluate the overall impact on the planned activity and its participants
Gen.Pro.C-1 Gain fundamental scientific knowledge in the field of physical and mathematical sciences	Gen.Pro.C-1.1 Apply fundamental scientific knowledge in the field of physical and mathematical sciences
	Gen.Pro.C-1.2 Consolidate and critically assess professional experience and research findings
	Gen.Pro.C-1.3 Understand interdisciplinary relations in applied mathematics and computer science and apply them in professional settings

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

As a result of studying the course the student should:

know:

theory of machine and deep learning (concepts, axioms, methods);
modern problems of the relevant sections of deep learning;
basic libraries and software for the implementation of their own or typical projects using deep learning.

be able to:

understand the task and evaluate its correctness;
use your knowledge to solve applied problems;
find algorithms for solving problems and analyze them;
analyze scientific literature on deep learning.

master:

skills of mastering a large amount of information and solving problems (including complex ones);
skills of independent work and mastering new disciplines;
culture of setting, analyzing and solving mathematical and applied problems;
the subject language of deep learning and the skills of competently describing problem solving and presenting the results.

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	Introduction to deep learning	2	2		3
2	Fighting overfitting in neural networks	4	4		6
3	Convolutional neural networks	4	4		6
4	Recurrent neural networks	4	4		6
5	Seq2seq models	4	4		6
6	Attention mechanism	4	4		6
7	Learning without a teacher	4	4		6
8	Generative models	4	4		6
AH in total		30	30		45
Exam preparation		30 AH.			
Total complexity		135 AH., credits in total 3			

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

Semester: 2 (Spring)

1. Introduction to deep learning

Neural networks. Backpropagation of the error. Stochastic gradient descent. Weighing initialization. Implementation using the Pytorch library.

2. Fighting overfitting in neural networks

Advanced optimization techniques. Optimization of hyperparameters. Batch normalization. Regularization. Dropout. Fading gradient problem.

3. Convolutional neural networks

Image tasks: classification, localization, detection, segmentation, semantic segmentation. Basic Convolutional Network Architectures. Transfer learning.

4. Recurrent neural networks

GRU and LSTM. Application of recurrent networks for natural language processing problems. Vector representations of words. word2vec: CBOW and skip-gram. Language models. GPT models - (1,2,3).

5. Seq2seq models

Seq2seq models. Translation tasks, text2speech, speech recognition.

6. Attention mechanism

Attention mechanism. Transformers (Transformer). BERT, RoBERTa models.

7. Learning without a teacher

Autoencoders. Variation autoencoder.

8. Generative models

Generative Adversarial Networks (GAN). Generation of images.

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

Equipment needed for lectures and seminars: computer and multimedia eqipment (projector, sound system).

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

Main literature

Provided at the department:

1. I. Goodfellow et al. Deep Learning. http://deeplearningbook.org
2. С. Николенко и др. Введение в глубокое обучение. - СПб. - Питер, 2017.

Additional literature

Provided at the department:

1. http://cs231n.stanford.edu/
2. http://deeplearning.cs.cmu.edu/F20/index.html

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

1. http://web.stanford.edu/class/cs224n/
2. https://pytorch.org/

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

For some of the lessons, you will need Zoom. Google Drive to access course materials. The presence of smartphones / laptops during classes is encouraged to participate in interactive exercises.

9. Guidelines for students to master the course

A student who studies discipline must, on the one hand, master a general conceptual apparatus, and on the other hand, must learn to apply theoretical knowledge in practice.

As a result of studying the discipline, the student should know the basic definitions of the discipline, be able to apply this knowledge to solve various problems.

Successful learning requires:

- visits to all classes provided by the curriculum for the discipline;
- conducting the abstract of occupations;
- intense independent work of the student.

Independent work includes:

- reading recommended literature;

- study of educational material, preparation of answers to questions intended for self-study;
- solving problems offered to students in the classroom;
- preparation for performance of tasks of the current and intermediate certification.

An indicator of possession of the material is the ability to answer questions on discipline topics without an outline.

It is important to achieve an understanding of the material being studied, and not its mechanical memorization. If it is difficult to study individual
topics, questions, you should seek advice from the teacher.

Intermediate control of students' knowledge in the form of problem solving in accordance with the subject of classes is possible

SUPPLEMENT

Assessment funds for course (training module)

major:	Applied Mathematics and Physics
specialization:	Applied Bioinformatics/Прикладная биоинформатика Phystech School of Biological and Medical Physics Chair of Bioinformatics and Systems Biology
term:	1
qualification:	Master

Semester, form of interim assessment: 2 (spring) - Exam

Author: A.A. Zverkov

1. Competencies formed during the process of studying the course

Code and the name of the competence	Competency indicators
UC-1 Use a systematic approach to critically analyze a problem,and develop an action plan	UC-1.1 Systematically analyze the problem situation, identify its components and the relations between them
	UC-1.2 Search for solutions by using available sources
	UC-1.3 Develop a step-by-step strategy for achieving a goal, foresee the result of each step, evaluate the overall impact on the planned activity and its participants
Gen.Pro.C-1 Gain fundamental scientific knowledge in the field of physical and mathematical sciences	Gen.Pro.C-1.1 Apply fundamental scientific knowledge in the field of physical and mathematical sciences
	Gen.Pro.C-1.2 Consolidate and critically assess professional experience and research findings
	Gen.Pro.C-1.3 Understand interdisciplinary relations in applied mathematics and computer science and apply them in professional settings

2. Competency assessment indicators

As a result of studying the course the student should:

know:

theory of machine and deep learning (concepts, axioms, methods);
modern problems of the relevant sections of deep learning;
basic libraries and software for the implementation of their own or typical projects using deep learning.

be able to:

understand the task and evaluate its correctness;
use your knowledge to solve applied problems;
find algorithms for solving problems and analyze them;
analyze scientific literature on deep learning.

master:

skills of mastering a large amount of information and solving problems (including complex ones);
skills of independent work and mastering new disciplines;
culture of setting, analyzing and solving mathematical and applied problems;
the subject language of deep learning and the skills of competently describing problem solving and presenting the results.

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

During the current control, the student should be able to answer the following questions:

1. Backpropagation method for training neural networks.
2. Implementation of neural networks using Pytorch.
3 Implementation of the convolution operation.
4. The task of image classification / segmentation.
5. The problem of text classification.

During the class, interactive discussions can take place in the course chats, which will be homework. It is possible to perform patent search as an independent task. Successful completion of all tasks in the course and the implementation of control slices of knowledge gives an advantage on exam.

4. Evaluation criteria

1. Backpropagation method for training neural networks.
2. Implementation of neural networks using Pytorch.
3 Implementation of the convolution operation.

4. The task of image classification / segmentation.
5. The problem of text classification.
6. Batch normalization.
7. The problem of fading gradient.
8. Tasks with images: classification, localization, detection, segmentation, semantic segmentation.
9 Application of recurrent networks for natural language processing tasks.
11. Models BERT, RoBERTa.
12. Generative models.

The mark is excellent (10 points) - it is given to a student who has shown comprehensive, systematic, deep knowledge of the curriculum of the discipline, who has an interest in this subject area, has demonstrated the ability to confidently and creatively put them into practice in solving specific problems, and a free and proper substantiation of decisions.

The mark is excellent (9 points) - it is given to a student who has shown comprehensive, systematic, in-depth knowledge of the curriculum of the discipline and the ability to confidently put them into practice in solving specific problems, free and proper substantiation of the decisions made.

The mark is excellent (8 points) - given to a student who has shown comprehensive, systematic, in-depth knowledge of the curriculum of the discipline and the ability to confidently apply them in practice in solving specific problems, correct justification of decisions made, with some shortcomings.

A mark is good (7 points) - it is put up for a student, if he knows the material firmly, sets it up competently and in essence, knows how to apply the knowledge gained in practice, but does not competently substantiate the results obtained.

Evaluation is good (6 points) - it is put up to a student, if he knows the material firmly, sets it up correctly and in essence, knows how to apply this knowledge in practice, but admits some inaccuracies in the answer or in solving problems.

A mark is good (5 points) - it is given to a student, if he basically knows the material, correctly and essentially sets it out, knows how to apply this knowledge in practice, but allows a sufficiently large number of inaccuracies to answer or solve problems.

Grade satisfactorily (4 points) is given to a student who has shown the fragmented, fragmented nature of knowledge, insufficiently correct formulations of basic concepts, violations of the logical sequence in the presentation of program material, but at the same time he has mastered the main sections of the curriculum necessary for further education and can apply knowledge is modeled in a standard situation.

Grade satisfactorily (3 points) - given to a student who showed the fragmented, scattered nature of knowledge, making mistakes in formulating basic concepts, disrupting the logical sequence in presenting program material, poorly masters the main sections of the curriculum required for further education and even applies the knowledge gained in a standard situation.

The rating is unsatisfactory (2 points) - is given to a student who does not know most of the main content of the curriculum of the discipline, makes gross mistakes in the wording of the basic principles and does not know how to use this knowledge when solving typical tasks.

Unsatisfactory mark (1 point) - is given to a student who does not know the main content of the discipline's curriculum, makes gross errors in the wording of the basic concepts of the discipline and does not have any skills to solve typical practical problems.

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

During the oral exam, the student is given 30 minutes to prepare. The interview for a student in an oral exam must not exceed one astronomical hour.