Skoltech

Course Syllabus

Course Title	Theoretical Methods of Deep Learning
Course Title (in Russian)	Теоретические методы глубокого обучения
Lead Instructor	Yarotsky, Dmitry

1. Annotation

Course Description

Deep Learning (DL) is an extremely important applied science that, at present, is poorly understood theoretically. We know that neural networks work well, but cannot fully explain why. Nevertheless, in recent years, there has been a rapid growth of publications that shed light on the new mathematics underlying DL, and we see now many interesting connections between DL and various other fields, e.g. approximation theory, differential equations, information theory, random matrix theory, statistical physics. This course aims to introduce students to these cutting-edge developments.

Course Description (in Russian)

Глубокое обучение (Deep Learning - DL) является чрезвычайно важной прикладной дисциплиной, в настоящее время недостаточно хорошо понимаемой на теоретическом уровне. Мы знаем, что нейронные сети хорошо работают, но не можем этого полностью объяснить. Тем не менее, в последние годы наблюдается быстрый рост публикаций, проливающих свет на новую математику, лежащую в основе DL, и мы видим много интересных связей между DL и другими областями, такими как теория аппроксимаций, дифференциальные уравнения, теория информации, теория случайных матриц и статистическая физика. Данный курс ставит целью познакомить студентов с этими передовыми достижениями.

2. Basic Information

Course Academic Level

MSc PhD

Number of ECTS credits

3

Course Prerequisites / Recommendations

This is a theoretical course that involves relatively complex and diverse mathematics. Also, familiarity with machine learning and deep learning is desirable (but not critical).

Type of Assessment

Graded

Mapping from grades to percentage:

A: 80

B: 70

C: 60

D: 50

E: 40

F: 0

Term

Term 2

Students of Which Programs do You Recommend to Consider this Course as an Elective?

BSc Programs	Masters Programs	PhD Programs
	Advanced Computational Science Data Science Mathematical and Theoretical Physics	Computational and Data Science and Engineering Mathematics and Mechanics Physics

Maximum Number of Students

	Maximum Number of Students
Overall:	30
Per Group (for seminars and labs):	30

Course Stream

Science, Technology and Engineering (STE)

3. Course Content

Topic	Summary of Topic	Contact Hours: Lectures	Contact Hours: Seminars	Contact Hours: Labs	Non-contact Hours: Student's Independent Study
Introduction and overview	Neural networks. Deep learning and current status of its theoretical understanding. Main research directions. Overview of connections to statistical learning theory, approximation theory, random matrix theory, information theory, statistical physics, and other fields.	2	0	0	0
Expressiveness	Measures of network expressiveness. Approximation with ANNs. The universal approximation theorem. Sobolev spaces. Approximation rates. Canonical rates of continuous parametric approximation. Fast approximation of smooth functions by ReLU networks. Vapnik-Chervonenkis dimension. VC-dimensions of perceptrons. Upper and lower bounds for VC-dim of deep networks. Application to generalization bounds. The bit extraction technique. Deep narrow networks. Expressiveness in terms of the number of linear pieces and Betti numbers of level sets. Entropy and complexity of Sobolev balls.	5	5	1	14
Hierarchical generative models	Information flow on trees. The binary symmetric channel. Connection to the Ising model on Cayley tree. The problem of reconstruction solvability. The phase transition between solvable and non-solvable regimes.	2	2	1	7
Optimization	Gradient descent and backpropagation. Local minima and saddle points on the loss surface. The rate of convergence to local minima. Evasion of non-degenerate saddle points. Deep linear networks. Absence of non-global minima in linear networks.	3	3	1	10
Wide networks	Mean field theory and Gaussian approximation. Order-to-chaos transition in wide networks. Depth of information propagation. Random matrix theory. Wishart and Wigner ensembles. The semicircular and Marchenko-Pastur laws. Free probability and noncommutative central limit theorem. Applications of random matrix theory to wide networks. Neural tangent kernel.	5	5	1	14

4. Learning Outcomes

Skoltech Learning Outcomes are indicated as per Skoltech Learning Outcomes Framework.

1. FUNDAMENTAL KNOWLEDGE

1.1. KNOWLEDGE OF MATHEMATICS AND NATURAL SCIENCES

1.2. KNOWLEDGE OF APPLIED SCIENCE AND ENGINEERING SCIENCE INCLUDING CONTEMPORARY METHODS AND TOOLS

1.4. INTERDISCIPLINARY THINKING, KNOWLEDGE STRUCTURE AND INTEGRATION

2.1. COGNITION AND MODES OF REASONING

2.1.1. Analytical reasoning and problem solving

2.1.3. Creative thinking

2.2. ATTITUDES AND LEARNING PROCESS

2.2.3. Responsibility, intensity, perseverance, urgency and will to deliver

2.2.4. Resourcefulness, flexibility and an ability to adapt

2.2.5. Self-awareness and a commitment to self-improvement, lifelong learning and educating

2.3. ETHICS, EQUITY AND OTHER RESPONSIBILITIES

2.3.1. Ethical action, integrity and courage

3.1. COMMUNICATIONS IN INTERNATIONAL ENVIRONMENTS

3.1.2. Written, electronic and graphical communication

3.1.3. Oral presentation and discussion

3.1.5. Communications in English in scientific, business and social settings

3.2. TEAMWORK AND LEADERSHIP

3.2.1. Forming effective teams

4.1. MAKING SENSE OF GLOBAL SOCIETAL ENVIRONMENTAL AND BUSINESS CONTEXT

4.1.3. Understanding the technical products, systems and infrastructure of the sector

4.2. VISIONING — INVENTING NEW TECHNOLOGIES THROUGH RESEARCH

4.2.1. The research process — hypothesis, evidence and defense

4.2.2. Basic research leading to new scientific discovery

5. Assignments and Grading

Physical Attendance Requirement (% of classes)	70

Assignment Type	Assignment Summary	% of Final Course Grade
Homework Assignments	Weekly assignments involving theoretical and programming exercises	30
Final Project	A project performed individually or in a small team. The students can work out and present details of some relevant publication (preferably along with computer experiments), or present their own research if it fits the topic of the course.	25
Final Exam	A written theoretical 3-hour exam containing 10-15 questions and problems.	40
Class participation	Active participation in class discussions and exercises.	5

6. Assessment Criteria

Assignment 1 Type

Homework Assignments

Sample of Assignment 1

Prove that the family of arbitrary-depth neural networks with width 2 and the ReLU activation function has an infinite VC-dimension. Describe a size-N set shattered by this family. Write a program that, given an assignment of N outcomes 0-1, finds the network weights such that the network outputs these outcomes on the shattered set.

Assessment Criteria for Assignment 1

A correct proof and a working program.

Assignment 2 Type

Final Project

Sample of Assignment 2

Study numerically the low-lying part of the loss surface of a deep network, following recent publications on this topic. Perform multiple optimization runs with random initial conditions and analyze the resulting distribution of loss values. Discuss how the results can be explained theoretically (e.g., using the spin-glass model of Choromanska et al. or some other publications).

Assessment Criteria for Assignment 2

A valid numerical experiment and its sound theoretical analysis.

Assignment 3 Type

Final Exam

Sample of Assignment 3

1) Formulate the universal approximation theorem.
2) For the given neural network model, determine if the gradient descent optimization converges to the global optimum.

Assessment Criteria for Assignment 3

Correct and full answers and solutions.

Assignment 4 Type

Class participation

Sample of Assignment 4

Compute the VC-dimension of the given family of classifiers.

Assessment Criteria for Assignment 4

Correct and full answers and solutions.

7. Textbooks and Internet Resources

You can request at most two required textbooks. Additionaly, you can suggest up to nine recommended textbooks.

Required Textbooks	ISBN-13 (or ISBN-10)
-	

Recommended Textbooks	ISBN-13 (or ISBN-10)
M. Anthony, P. Bartlett, Neural Network Learning: Theoretical Foundations, Cambridge University Press, 1999	978-0521118620

Papers	DOI or URL
J.Berner, Ph.Grohs, G.Kutyniok, Ph.Petersen, The Modern Mathematics of Deep Learning, arXiv:2105.04026	https://arxiv.org/abs/2105.04026

Web-resources (links)	Description
https://neurips.cc/	The conference NeurIPS
https://icml.cc/	The conference ICML
https://iclr.cc/	The conference ICLR
http://unsupervised.cs.princeton.edu/deeplearningtutorial.html	S. Arora, Toward theoretical understanding of deep learning // (an ICML tutorial giving an overview of the subject)
https://www.cs.princeton.edu/courses/archive/fall19/cos597B/lecnotes/bookdraft.pdf	A theory of deep learning book (draft)
https://www.deeplearningbook.org/	A standard introduction to practical deep learning
https://random-matrix-learning.github.io/	A tutorial on application of Random Matrix Theory to machine learning
https://github.com/google/neural-tangents	Neural Tangent Kernel: an implementation and experiments

Web-resources (links)	Description
https://terrytao.files.wordpress.com/2011/02/matrix-book.pdf	Terence Tao, Topics in random matrix theory

8. Facilities

Equipment
A large (black- or white-) board and markers for writing; a projector.

Labs for Education

Laboratory of Mathematical Foundations of Artificial Intelligence

9. Additional Notes

The proposed course 1) has explicit academic content and requirements for receiving credits, 2) is in alignment with the program's learning outcomes, 3) adheres to policies and Skoltech regulations.

Lead Instructor confirms