Government of Russian Federation

Federal State Autonomous Educational Institution of High Education

«National Research University Higher School of Economics»

National Research University
High School of Economics
Faculty of Computer Science

Syllabus for the course
«Machine Learning and Data Mining»

(Машинное обучение и майнинг данных)

01.04.02 «Applied Mathematics and Informatics»,
«**Data Science**» Master program

Authors:
Dmitry I. Ignatov, Candidate of sciences (Ph.D), associate professor, dignatov@hse.ru
Maxim A. Borisyak, Master of sciences (MSc), senior lecturer mborisyak@hse.ru

Approved by: Head of Data Analysis and Artificial Intelligence Department, Sergey O. Kuznetsov

Recommended by:

Moscow, 2018

1. Teachers

Author, lecturer: Borisyak Maxim, National Research University Higher School of Economics, Department of Data Analysis and Artificial Intelligence, senior lecturer

2. Scope of Use

The present program establishes minimum demands of students' knowledge and skills, and determines content of the course.

The present syllabus is aimed at department teaching the course, their teaching assistants, and students of the Master of Science program 01.04.02 «Data Sciences»,

This syllabus meets the standards required by:

- Educational standards of National Research University Higher School of Economics;
- Educational program «Data Sciences» of Federal Master's Degree Program 01.04.02 «Applied Mathematics and Informatics», 2018;
- University curriculum of the Master's program in «Data Sciences» for 2018.

3. Summary

The course "**Machine Learning and Data Mining**" introduces students to new and actively evolving interdisciplinary field of modern data analysis. Started as a branch of Artificial Intelligence, it attracted attention of physicists, computer scientists, economists, computational biologists, linguists and others and become a truly interdisciplinary field of study. In spite of the variety of data sources that could be analyzed, objects and attributes that from a particular dataset poses common statistical and structural properties. The interplay between known data and unknown ones give rise to complex pattern structures and machine learning methods that are the focus of the study. In the course we will consider methods of Machine Learning and Data Mining. Special attention will be given to the hands-on practical analysis of the real world datasets using available software tools and modern programming languages and libraries.

4. Learning Objectives

Learning objectives of the course "**Machine Learning and Data Mining**" (MLDM) are to familiarize students with a new rapidly evolving filed of machine learning and mining, and provide practical knowledge experience in analysis of real world data.

5. Learning outcomes

After completing the study of the discipline "Machine Learning and Data Mining", the student should:

- Know basic notions and terminology used in MLDM
- Understand fundamental principles of modern data analysis
- Learn to develop mathematical models of MLDM
- Be capable of analyzing real world data

After completing the study of the discipline "Machine Learning and Data Mining" the student should have the following competences:

Competence	**Code**	**Code (UC)**	**Descriptors (indicators of achievement of the result)**	**Educative forms and methods aimed at generation and development of the competence**
The ability to reflect developed methods of activity.	SC-1	SC-M1	The student is able to reflect developed mathematical methods for machine learning and data mining (data sciences)	Lectures and tutorials, group discussions, presentations, paper reviews.
The ability to propose a model to invent and test methods and tools of professional activity	SC-2	SC-M2	The student is able to improve and develop research methods as applicable to machine learning and data mining (data sciences)	Classes, home works.
Capability of development of new research methods, change of scientific and industrial profile of self-activities	SC-3	SC-M3	The student obtains necessary knowledge in machine learning and data mining, which is sufficient to develop new methods	Home tasks, paper reviews
The ability to describe problems and situations of professional activity in terms of humanitarian, economic and social sciences to solve problems which occur across sciences, in allied professional fields.	PC-5	IC-M5.3_5.4_5.6_2.4.1	The student is able to describe data analysis problems in terms of computational mathematics.	Lectures and tutorials, group discussions, presentations, paper reviews.
The ability to detect, transmit common goals in the professional and social activities	PC-8	SPC-M3	The student is able to identify mathematical aspects in machine learning and data mining tasks, evaluate correctness of the used methods, and their applicability in each current situation	Discussion of paper reviews; cross discipline lectures

6. Place of the discipline in the Master's program structure

The course "Machine Learning and Data Mining" is a course taught in the first year of the Master's program 01.04.02 "Data Science" and is a base course for specialization "Intelligent Systems and Structural Analysis"

Prerequisites

The course is based on knowledge and understanding of

- Discrete mathematics
- Algorithms and data structures
- Linear algebra
- Theory of probability and statistical analysis

It also requires some programming experience in the Python programming language.

7. Schedule

One pair consists of 1 academic hour for lecture and 1 academic hour for classes after lecture.

№	Topic	Total hours	Contact hours		Self-study
			Lectures	Seminars	
1	Introduction to Machine Learning and Data Mining, No-Free-Lunch theorems	26	3	3	20
2	Bias-variance decomposition, regularization techniques.	26	3	3	20
3	Introduction to meta-algorithms, bootstrap, boosting.	26	3	3	20
4	Optimization techniques: black-box methods, first order methods.	26	3	3	20
5	Miscellaneous topics: imbalanced datasets, importance sampling, one-class classification methods.	26	3	3	20
6	Introduction and overview of deep learning methods.	26	3	3	20
7	Deep generative models: energy-based models, Boltzmann machines and deep belief networks.	28	4	4	20
8	Deep generative models: Variational AutoEncoders.	26	3	3	20
9	Deep generative models: Generative Adversarial Networks.	30	4	4	22

10	Meta-learning: concept learning, learning how to learn.	26	3	3	20
	Total	266	32	32	202

8. Requirements and Grading

	Type of work	Characteristics		
		1	2	
Type of grading	Homework	3		Solving homework tasks and examples.
	Special homework – research projects		1	Independent modelling and verification of research papers results
	Exam		1	Oral exam
Final				

9. Assessment

The assessment consists of classwork and homework, assigned after each lecture. Students have to demonstrate their knowledge in each lecture topic concerning both theoretical facts, and practical tasks' solving. All tasks are connected through the discipline and have increasing complexity.

Final assessment is the final exam. Students have to demonstrate knowledge of theory facts, but the most of tasks would evaluate their ability to solve practical examples, present straight operation, and recognition skills to solve them.

The grade formula:

The exam will consist of 1 paper review, total 10 points for the exam

Final course mark *is obtained from the following formula*:

Ofinal = 0,5*Ocumulative + 0,5*Oexam.

The grades are rounded in favour of examinee.

Table of Grade Accordance

Ten-point Grading Scale	**Five-point Grading Scale**	
1 - very bad 2 – bad 3 – no pass	Unsatisfactory - 2	**FAIL**

4 – pass 5 – highly pass	Satisfactory – 3	**PASS**
6 – good 7 – very good	Good – 4	
8 – almost excellent 9 – excellent 10 – perfect	Excellent – 5	

10. Course Description

The following list describes main topics covered by the course with lecture order.

Topic 1. Introduction to Machine Learning and Data Mining, No-Free-Lunch theorems

Content: Introduction to No-Free-Lunch theorems, discussion about the role of prior knowledge in Machine Learning. Discussion of the general Machine Learning workflow. Assumptions behind the most popular Machine Learning methods.

Topic 2. Bias-variance decomposition, regularization techniques.

Content: Model complexity through bias-variance decomposition, methods to control complexity of the models, the most common regularization techniques.

Topic 3. Introduction to meta-algorithms, bootstrap, boosting.
Content: Meta-algorithms as a tool for regulating bias/variance of the model. Introduction to bootstrap: Random Forest. Stacking. Introduction to boosting: AdaBoost, Gradient Boosting Machine, XGBoost.

Topic 4. Optimization techniques: black-box methods, first order methods.
Content: Brief overview of first order optimization methods: stochastic gradient descent, momentum, adam/adamax. Detailed discussion of black-box optimization methods: Bayesian optimization, Variational Optimization. Examples of black-box optimization: hyper-parameter tuning.

Topic 5. Miscellaneous topics: imbalanced datasets, importance sampling, one-class classification methods.
Content: Discussion of the problems caused by imbalanced datasets, particularly, for gradients based methods: change of priors, importance sampling. One-class classification: one-class SVM, density based methods, popular heuristics: dimensionality reduction (e.g. through AutoEncoders), Radial Basis Networks.

Topic 6. Introduction and overview of deep learning methods.
Content: Introduction to Deep Learning through No-Free-Lunch theorem lens. Popularity of Deep Learning methods. Correspondence between the most common Deep Learning methods and prior assumptions.

Topic 7. Deep generative models: energy-based models, Boltzmann machines and deep belief networks.

Content: Definition of a generative problem, types of generative problems. Energy-based models and contrastive divergence: Boltzmann machines, Deep Belief Networks, Restricted Boltzmann Machines and their connection to the AutoEncoders.

Topic 8. Deep generative models: Variational AutoEncoders.

Content: Variational bounds on likelihood, Variational AutoEncoder, Conditional Variational AutoEncoder.

Topic 9. Deep generative models: Generative Adversarial Networks (GANs).

Content: Jensen-Shannon divergence and Wasserstein distance as minimization problems. Adversarial Neural Networks: classical GAN, WGAN, energy-based GAN. Difficulties in adversarial training, gradient penalty for WGAN. Practical applications beyond generative problems. Adversarial AutoEncoder, BiGAN, CycleGAN, Adversarial Variational Bayes.

Topic 10. Meta-learning: concept learning, learning how to learn.

Content: Concept learning: Neural Statistician, Generative Matching Networks. Learning how to learn: optimization procedure as a learning problem, gradient-based optimization algorithms.

11. Term Educational Technology

The following educational technologies are used in the study process:

- discussion and analysis of the results of the home task in the group;
- individual education methods, which depend on the progress of each student;
- group projects on analysis of real data.

12. Recommendations for course lecturer

Course lecturer is advised to use interactive learning methods, which allow participation of the majority of students, such as slide presentations, combined with writing materials on board, and MLDM software tools for demonstration and practicing purposes. The course is intended to be introductory, that is rather broad in nature, but it is normal to differentiate tasks and projects in a group if necessary, and direct fast or more dedicated learners to solve more complicated tasks. The final group project and computational homework tasks are inevitable constituents of the cource.

13. Recommendations for students

Lectures are combined with classes. Students are invited to ask questions and actively participate in-group discussions and projects. There will be special office hours for students, which would like to get more precise understanding of each topic. Teaching assistant will also help you. All tutors are ready to answer your questions online by official e-mails that you can find in the “contacts” section. Note that the final mark is a cumulative value of your term activity and final results.

14. Sample final exam questions

1. Choose a machine learning algorithm for the following problems and explain your choice:
    a. two-class classification with classes distributed normally with equal variance;
    b. two-class classification with a weakly correlated features.
    c. two-class classification with individual features not having any discriminative power, i.e. marginal distributions of each feature conditioned on class are equal.
    d. object recognition in natural images.

2. Contrastive divergence training, major computational difficulties for contrastive divergence, motivation behind Restricted Boltzmann machines.
3. What types of base learners are best suited for boosting methods? For bootstrap methods?

15. Reading and Materials

15.1. Required Reading

1. Mohammed J. Zaki and Wagner Meira, Jr., Data Mining and Analysis: Fundamental Concepts and Algorithms, Cambridge University Press, 2014
2. Peter Flach Machine Learning: The Art and Science of Algorithms that Make Sense of Data, Cambridge University Press, 2012

15.2. Recommended Reading

1. C.M. Bishop. Pattern Recognition and Machine Learning Springer (2006) http://research.microsoft.com/en-us/um/people/cmbishop/PRML/index.htm
2. D. Barber Bayesian Reasoning and Machine Learning, Cambridge University Press, 2012
3. http://www-stat.stanford.edu/~tibs/ElemStatLearn/
4. J. Han, M. Kamber, J. Pei. Data Mining: Concepts and Techniques, Third Edition. — Morgan Kaufmann Publishers, 2012. — 703 p.
5. T. Hastie, R.Tibshirani, J. Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition, Springer 2009
6. T. Mitchel. Machine Learning, McGraw Hill, 1997. http://www.cs.cmu.edu/~tom/mlbook.html
7. Witten, E. Frank, M. Hall. Data Mining: Practical Machine Learning Tools and Techniques, 2011, Morgan Kaufmann Publishers http://www.cs.waikato.ac.nz/ml/weka/book.html

Supplementary reading:

8. B. G. Mirkin. Core Concepts in Data Analysis: Summarization, Correlation, Visualization, Springer, 2011, 388 p.
9. B.G. Mirkin Clustering: A Data Recovery Approach. CRC Press, 2012.

15.3. List of review papers

1 Wolpert, D.H. and Macready, W.G., 1997. No free lunch theorems for optimization. IEEE transactions on evolutionary computation, 1(1), pp.67-82.
2 Schaffer, C., 1994. A conservation law for generalization performance. In Proceedings of the 11th international conference on machine learning (pp. 259-265).
3 Hastie, T., Rosset, S., Zhu, J. and Zou, H., 2009. Multi-class adaboost. Statistics and its Interface, 2(3), pp.349-360.
4 Bengio, Y., Roux, N.L., Vincent, P., Delalleau, O. and Marcotte, P., 2006. Convex neural networks. In Advances in neural information processing systems (pp. 123-130).
5 Kingma, D. and Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
6 Zeiler, M.D., 2012. ADADELTA: an adaptive learning rate method. arXiv preprint arXiv:1212.5701.
7 Wierstra, D., Schaul, T., Glasmachers, T., Sun, Y., Peters, J. and Schmidhuber, J., 2014. Natural evolution strategies. Journal of Machine Learning Research, 15(1), pp.949-980.
8 Glorot, X. and Bengio, Y., 2010, March. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (pp. 249-256).
9 Ciregan, D., Meier, U. and Schmidhuber, J., 2012, June. Multi-column deep neural networks for image classification. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on (pp. 3642-3649). IEEE.
10 Graham, B., 2014. Fractional max-pooling. arXiv preprint arXiv:1412.6071.
11 Courbariaux, M., Bengio, Y. and David, J.P., 2015. Binaryconnect: Training deep neural networks with binary weights during propagations. In Advances in Neural Information Processing Systems (pp. 3123-3131).
12 Springenberg, J.T. and Riedmiller, M., 2013. Improving deep neural networks with probabilistic maxout units. arXiv preprint arXiv:1312.6116.
13 Liao, Z. and Carneiro, G., 2015. Competitive multi-scale convolution. arXiv preprint arXiv:1511.05635.
14 Ren, S., He, K., Girshick, R. and Sun, J., 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).
15 Wu, J., Leng, C., Wang, Y., Hu, Q. and Cheng, J., 2016. Quantized convolutional neural networks for mobile devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4820-4828).
16 Kim, Y.D., Park, E., Yoo, S., Choi, T., Yang, L. and Shin, D., 2015. Compression of deep convolutional neural networks for fast and low power mobile applications. arXiv preprint arXiv:1511.06530.
17 Lebedev, V., Ganin, Y., Rakhuba, M., Oseledets, I. and Lempitsky, V., 2014. Speeding-up convolutional neural networks using fine-tuned cp-decomposition. arXiv preprint arXiv:1412.6553.
18 Tieleman, T. and Hinton, G., 2009, June. Using fast weights to improve persistent contrastive divergence. In Proceedings of the 26th Annual International Conference on Machine Learning (pp. 1033-1040). ACM.

19 Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V. and Courville, A., 2017. Improved training of wasserstein gans. arXiv preprint arXiv:1704.00028.
20 Salakhutdinov, R. and Hinton, G., 2009, April. Deep boltzmann machines. In Artificial Intelligence and Statistics (pp. 448-455).
21 Louppe, G. and Cranmer, K., 2017. Adversarial Variational Optimization of Non-Differentiable Simulators. arXiv preprint arXiv:1707.07113.
22 Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M. and Lempitsky, V., 2016. Domain-adversarial training of neural networks. Journal of Machine Learning Research, 17(59), pp.1-35.
23 Pu, Y., Wang, W., Henao, R., Chen, L., Gan, Z., Li, C. and Carin, L., 2017. Adversarial symmetric variational autoencoder. In Advances in Neural Information Processing Systems (pp. 4331-4340).
24 Zhu, J.Y., Park, T., Isola, P. and Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv preprint arXiv:1703.10593.
25 Gatys, L.A., Ecker, A.S. and Bethge, M., 2016. Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2414-2423).
26 Kim, T. and Bengio, Y., 2016. Deep directed generative models with energy-based probability estimation. arXiv preprint arXiv:1606.03439.
27 Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I. and Abbeel, P., 2016. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In Advances in Neural Information Processing Systems(pp. 2172-2180).
28 Arjovsky, M. and Bottou, L., 2017. Towards principled methods for training generative adversarial networks. arXiv preprint arXiv:1701.04862.
29 Dziugaite, G.K. and Roy, D.M., 2015. Neural network matrix factorization. arXiv preprint arXiv:1511.06443.
30 Mnih, A. and Salakhutdinov, R.R., 2008. Probabilistic matrix factorization. In Advances in neural information processing systems (pp. 1257-1264).
31 Salakhutdinov, R., Mnih, A. and Hinton, G., 2007, June. Restricted Boltzmann machines for collaborative filtering. In Proceedings of the 24th international conference on Machine learning (pp. 791-798). ACM.
32 Rendle, S., Gantner, Z., Freudenthaler, C. and Schmidt-Thieme, L., 2011, July. Fast context-aware recommendations with factorization machines. In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval (pp. 635-644). ACM.
33 Hu, Y., Koren, Y. and Volinsky, C., 2008, December. Collaborative filtering for implicit feedback datasets. In Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on (pp. 263-272). Ieee.
34 Andrychowicz, M., Denil, M., Gomez, S., Hoffman, M.W., Pfau, D., Schaul, T. and de Freitas, N., 2016. Learning to learn by gradient descent by gradient descent. In Advances in Neural Information Processing Systems (pp. 3981-3989).

15.4. Course telemaintenance

All material of the discipline are posted in informational educational site at NRU HSE portal www.ami.hse.ru. Students are provided with links to research papers, electronic books, data and software.

16. Equipment

The course requires a laptop, projector, and acoustic systems.
It also requires opportunity to install programming software, such as:

- Python
- sklearn
- tensorflow
- keras

on student personal computers.

Lecture materials, course structure and syllabus are prepared by Maxim Borisyak.