Skoltech

Course Syllabus

Course Title	Machine Learning
Course Title (in Russian)	Машинное обучение
Lead Instructor	Burnaev, Evgeny
Co-Instructor	Zaytsev, Alexey

1. Annotation

Course Description

The course is a general introduction to machine learning (ML) and its applications. It covers fundamental topics in ML and describes the most important algorithmic basis and tools. It also provides important aspects of the algorithms' applications. The course starts with an overview of canonical ML applications and problems, learning scenarios, etc. Next, we discuss in-depth fundamental ML algorithms for classification, regression, clustering, etc., their properties, and practical applications. The last part of the course is devoted to advanced ML topics such as Gaussian processes, neural networks. Within practical sections, we show how to use the ML methods and tune their hyper-parameters. Home assignments include the application of existing algorithms to solve data analysis problems. The students are assumed to be familiar with basic concepts in linear algebra, probability, real analysis, optimization, and python programming.

On completion of the course students are expected to:
➢ Have a good understanding of the fundamental issues and challenges of ML: data, model selection, model complexity among others;
➢ Have an understanding of the strengths and weaknesses of many popular ML approaches;
➢ Appreciate the basic underlying mathematical relationships within and across ML algorithms and the paradigms of supervised and unsupervised learning.
➢ Be able to design and implement various machine learning algorithms in a range of real-world applications.

Course Description (in Russian)

Курс представляет собой общее введение в машинное обучение (МЛ) и его приложения. Курс охватывает фундаментальные темы в ML и описывает наиболее важные алгоритмические основы и подходы, аспекты применения алгоритмов. Курс начинается с обзора канонических приложений и задач ML, сценариев обучения и т.д. Далее, в курсе подробно обсуждаются фундаментальные алгоритмы ML для классификации, регрессии, кластеризации и т.д., их свойства и практическое применение. Последняя часть курса посвящена продвинутым темам ML, таким как гауссовские процессы, нейронные сети. В практических сессиях мы покажем, как использовать методы ML и настраивать их гиперпараметры. Домашние задания включают в себя применение существующих алгоритмов для решения задач анализа данных. Предполагается, что студенты, посещающие данный курс, уже знакомы с основными понятиями линейной алгебры, теории вероятностей, математического анализа, оптимизации и программирования на python.

По окончании курса студенты будут:

➢ Иметь хорошее понимание фундаментальных проблем и проблем МЛ: анализ свойств данных, выбор типа модели и сложности соответствующего функционального класса, и т.п.;
➢ Иметь представление о сильных и слабых сторонах многих популярных подходов ML;
➢ Понимать лежащие в основе ML методов математические концепции, позволяющие проводить обучение с учителем и обучение без учителя;
➢ Уметь разрабатывать и использовать алгоритмы машинного обучения для решения прикладных задач.

2. Basic Information

Number of ECTS credits 6

Course Prerequisites / Recommendations

Course prerequisites are numerical linear algebra, calculus, probability theory, optimization and programming (python).

Type of Assessment Graded

Mapping from grades to percentage:

A:	86
B:	76
C:	66
D:	56
E:	46
F:	0

Term Term 3

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

Masters Programs	PhD Programs
Data Science	Computational and Data Science and Engineering Engineering Systems

Maximum Number of Students

	Maximum Number of Students
Overall:	120
Per Group (for seminars and labs):	120

Course Stream Science, Technology and Engineering (STE)

3. Course Content

Topic	Summary of Topic	Lectures (# of hours)	Seminars (# of hours)	Labs (# of hours)
Introduction Lecture	- Introduction. Some canonical applications and problems. Definitions and terminology - Types of problems: Supervised, Semi-supervised, Empirical Risk Minimization, Cross-validation	1.5	1.5	
Regression, Kernel Trick	- Regression (problem statement) - Linear Regression. Closed form Solution Ridge Regression. Closed-form Solution. Direct Dual Solution - Non-linear case. Kernels. Kernel Ridge Regression - LASSO. L1 Sparsity. ElasticNet	1.5	1.5	
Classification	- Binary classification, loss curve, ROC/AUC, precision and recall - Learning a classifier. Surrogate LossEmpirical risk minimization (ERM), Overfitting. - Regularization, Log Loss - Two class and Multiclass Logistic Regression - k-Nearest Neighbour CLassifier. Compactness and continuity hypotheses. Distance functions - Classification and regression trees. Ensembles - Naive Bayes Classifier	1.5	1.5	
Tree-based methods. Random forest	- Decision Tree Classifiers. Divide-and-conquer algorithm and ID3 algorithm. Greedy tree for classification. - Design choices for decision tree learning: choice of root; purity, entropy, information gain and gini index; loss function. - Pruning and overfitting: pre- and post-pruning. Stopping rules. - Bagging. Bootstrapping. Random Forest.	1.5	1.5	
Adaboost	- Ensembles of classifiers. Bagging. - Stacked generalization. - Boosting. Motivation. Spam filtering example: majority vote boosting approach. Boosting for binary classification. Boosting heuristics: continuous and exponential upper bounds. - AdaBoost. Toy examples and noisy problems. Boosting stumps.	1.5	1.5	
Gradient boosting	- Gradient Boosting. The gradient boosting machine (GBM). GBM for regression. GBM regularization via shrinkage. Theoretical motivation. - Gradient Boosting Decision Trees. Tree Ensembles.	1.5	1.5	

Topic	Summary of Topic	Lectures (# of hours)	Seminars (# of hours)	Labs (# of hours)
Advanced classification. Imbalanced and Multilabel cases	- Imbalanced classification. Imbalance ratio. - Resampling methods: random oversampling (ROS); random undersampling (RUS); synthetic minority oversampling technique (SMOTE). - Multi-Class Classification. One-vs-All. One-vs-One. Approach based on Error-Correcting Codes. - Multi-class Algorithms: Logistic Regression, SVMs. Bayesian Approach. - Extras: Nonparametric Estimation. Mean Integrated Square Error (MISE). Oversmoothing, undersmoothing. - Extras: Kernel density estimation (KDE). Multidimensional KDE. - Extras: Nonparametric regression. Nadaraya-Watson estimate.	1.5	1.5	
Model and Feature Selection	- Overfitting. Error Decomposition. Bias-Variance Tradeoff. - Hyperparameters. Model Selection. Train, Test and Validation Error. Cross-Validation. Model Consistency. - Feature Selection. LASSO Feature Selection. - Regularization. Bayesian View on Regularization. - Mallows' Cp Statistic. AIC. BIC. - Extras: Sensitivity Analysis. Elementary Effects. Sobol Indices. EASI and CSTA.	1.5	1.5	
Shallow Artificial Neural Networks	- Motivation. Real-life Neurons. - Perceptron algorithm and SGD. Convergence. - Feed-forward Neural Networks. Activation function. - Neural networks for classification. Log loss. - Forward and Backward propagation. Backprop algorithm. Issues.	1.5	1.5	
Deep Artificial Neural Networks	- Old School approach to feature engineering. Feature construction magic. - Neural network as a computational graph. - Examples of practical applications. Causes of Deep NN breakthrough - Universal Approximation - Training computational graphs via backprop. - Image representation (RGB). Convolutional layers. Pooling Layers. - AlexNet. VGG. Inception V3. ResNet. - Transfer learning via fine-tuning. - Recurrent nets (Brief). - Batchnorm. Weight Norm. Dropout. Data Augmentation	1.5	1.5	

Topic	Summary of Topic	Lectures (# of hours)	Seminars (# of hours)	Labs (# of hours)
Neural networks for sequential data	- Recurrent neural networks - LSTM and GRU - Basic attention idea - Self-attention - Transformers for sequential data - Visual transformers - Sensitivity vs attention	1.5	1.5	
Gaussian Processes	- Bayesian Modeling. Stochastic Process. Random function. - Gaussian Process definition. Joint Gaussian distribution. Mean value and covariance function. - Gaussian Kernel as Covariance function. Interpretability of Kernel Parameters. - Kernel arithmetics. Construction of New Kernels. - GP Regression with noise: Model. Prediction. Interpolation. Smoothing. - GP optimization. - GP classification. Sigmoidal likelihood. - Non-stationary GP.	1.5	1.5	
Dimensionality Reduction	- Problem Statement. Examples: Faces, Airfoils, MNIST. - PCA - Multidimensional Scaling (MDS) - Replicative Neural Networks (Autoencoders) - Graph based on nearest neighbours: ISOMAP, LLE - t-Stochastic Neighbour Embedding	1.5	1.5	
Anomaly Detection	- Anomaly detection. - Nearest neighbours based methods - One-class SVM. Kernel choice - Other approaches to anomaly detection	1.5	1.5	
Clustering	- Hierarchical clustering (agglomerative/divisive models). - K-means - Cluster validity. External Measures: Entropy, Mutual Information, Jaccard Index, Rand Index, Silhouette Coefficient Mixture models, etc. - Hard and soft assignment with K-means, Gaussian Mixture Models (GMM) - Expectation-Maximisation algorithm. EM convergence. K-means in comparison to Learning GMM	1.5	1.5	
Project consultations	Consultations about the final group projects.		18	
Contact hours	Additional contact hours in case of extra questions from students.			9

4. Learning Outcomes

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

5. Assignments and Grading

Assignment Type	Assignment Summary	% of Final Course Grade
Homework Assignments	In this homework, you need to implement some tasks with classic machine learning Python packages, such as Numpy, Matplotlib, etc., and implement and compare basic supervised machine learning algorithms.	20
Homework Assignments	In this homework, you need to solve some tasks connected with neural networks.	10
Homework Assignments	In this homework, you need to solve some tasks connected with Gaussian processes, and unsupervised machine learning algorithms	10
Final Project	Early draft of the project report. At the beginning of the term, the instructor presents a set of projects to students. Students will form groups. They will then work on the selected project. Projects can consist of reproduction studies, novel work, or analytical tasks. Students must propose an earlier draft of the final project with the aim of the project, main tasks, and literature review.	5
Final Exam	The exam consists of several theoretical questions, mostly multiple-choice.	25
Final Project	Concise project report with 4-6 pages and a github repo with fully reproducible code. After the project submission deadline, students will have to present the project in class. The final version of the project, code, and presentation will be graded.	30

6. Assessment Criteria

Assignment 1 Type Homework Assignments

Or Upload Sample of Assignment 1

Sample of Assignment 1

Assessment Criteria for Assignment 1

Homeworks cost 40% of the final grade. Each Homework consists of several coding problems. Each problem (or its bullit-point) is graded as 1 if the problem is completely solved (fully reproducible code is provided, evaluation output is preserved and is correct, necessary comments are provided by student), 0 - otherwise.

Assignment 2 Type

Final Project

Sample of Assignment 2

Concise project report with 4-6 pages, github repo with fully reproducible code. After the project submission deadline, students will have to present the project in the class.

5/35 - (group) grade for the quality of early report
20/35 - (group) grade for the quality of the final project
10/35 - (group) grade for the quality of presentation

Assignment 3 Type

Final Exam

Assessment Criteria for Assignment 3

Exam cost 25% of the final grade. It consists of several theoretical questions, mostly multiple choice.

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)
Bishop, C.M. Pattern Recognition and Machine Learning. Springer, 2007	9780387310732
Barber, D. Bayesian Reasoning and Machine Learning. Cambridge University Press, 2012	9780521518147

Recommended Textbooks	ISBN-13 (or ISBN-10)
Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, 2009.	9780387848587
Shai Shalev-Shwartz, Shai Ben-David. Understanding Machine Learning: From Theory to Algorithms. Cambridge, 2014.	9781107057135

Recommended Textbooks	ISBN-13 (or ISBN-10)
R.E. Schapire, Y. Friend. Boosting. MIT, 2012	9780262526036
M. Mohri, A. Rostamizadeh, A. Talwalkar. Foundations of Machine Learning. MIT, 2012.	9780262018258
B. Clarke, E. Fokoue, H.H. Zhang. Principles and Theory for Data Mining and Machine Learning. Springer, 2009	9780387981352
Kevin P. Murphy. Machine Learning: A Probabilistic Perspective. MIT Press, 2012.	9780262018029
Sutton and Barto. Reinforcement Learning: An Introduction. MIT Press, 1998.	9780262193986

Web-resources (links)	Description
http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.HomePage	online version of Barber's "Bayesian Reasoning and Machine Learning"
http://gaussianprocess.org/gpml/	Carl Rasmussen and Christopher Williams. Gaussian Processes for Machine Learning. The MIT Press, 2006.
http://wol.ra.phy.cam.ac.uk/mackay/itila/book.html	D. J. C. MacKay (2003) Information Theory, Inference, and Learning Algorithms.

8. Facilities

Software
Google Colab

9. Additional Notes