Subject code	Credits
INF5002	6

Course title in Lithuanian

MAŠININIS MOKYMAS

Course title in English

MACHINE LEARNING

Short course annotation in Lithuanian (up to 500 characters)

Kurso tikslas - pristatyti didelę mašininio mokymo srities algoritmų įvairovę ir jų taikymo galimybes. Kurso metu pristatomi dažniausiai naudojami mokymo su instruktoriumi (supervised learning), mokymo pastiprinimu (reinforcement learning) ir klasterizacijos (unsupervised learning) sąvokos ir metodai. Mašininio mokymo algoritmai analizuojami iš dviejų perspektyvų: koks jų efektyvumas ir tinkamumas spręsti realius industrinius atpažinimo uždavinius ir kokios šių algoritmų savybės leidžia juos laikyti įdomiais intelekto modeliais. Iliustracijai naudojami šnekos atpažinimo ir teksto atpažinimo sričių uždaviniai, kuriems spręsti reikia pritaikyti, adaptuoti ir integruoti skirtingus mašininio mokymo metodus. Laboratorinių darbų metu studentai realizuoja nedidelius programinės įrangos kūrimo projektus, kad įgytų praktinės mašininio mokymo metodų taikymo patirties.

Short course annotation in English (up to 500 characters)

The goal of this course is to introduce students to the broad landscape of existing machine learning techniques and their possible applications. This course presents the most popular unsupervised learning, reinforcement learning, and supervised learning concepts and techniques. The algorithms are discussed from two perspectives: as useful data mining tools for solving real-world recognition tasks as well as interesting computational models of intelligence. The real-world tasks from the fields of speech and natural language processing are used to illustrate the ways to fit and to combine many machine learning techniques together. Students will complete small software projects to gain practical experience with the techniques covered in this course.

Prerequisites for entering the course

Undergraduate courses: Introduction to programming, Calculus, Probability theory and mathematical statistics

Course aim

The aim of the course is to introduce students to the broad set of existing machine learning techniques and their possible applications.

Content

No	Content (topics)
1.	**Basic concepts of machine learning (ML)**. Induction, generalization, training data, instances, features, feature spaces, classes, classifiers and predictors, classifier assessment, n-fold cross validation, training, optimization and test data subsets, compact class distribution in feature space, feature selection, data over-fitting, minimum description length principle
2.	**Supervised deterministic ML techniques**. K-nearest neighbour technique, prediction by linear regression, artificial neural networks, multilayer perceptron, error backpropagation learning, support vector machines.
3.	**Supervised probabilistic ML techniques**. Optimum Bayes' classifier, probability density mixture models, Expectation Maximization technique, hidden Markov models, learning probabilistic belief networks.
4.	**Supervised symbolic ML techniques**. Learning decision trees, learning regression trees, learning general logic descriptions.
5.	**Additional ML techniques**. Generalization as search, version spaces formalism, Inductive logic programming, FOIL algorithm
6.	**Unsupervised ML techniques**. Hierarchical clustering, k-means clustering, EM clustering.
7.	**Combining decisions of multiple classifiers**. Bagging, Boosting, and Stacking; AdaBoost algorithm
8.	**Machine learning for signal processing applications**. Speech recognition, dynamic time warping, Viterbi algorithm, speaker identification, automatic music transcription.
9.	**Machine learning for natural language processing applications**. Statistical language modelling, automatic text categorization.
10.	**Semi-supervised ML techniques**. Reinforcement learning, Q-learning, TD learning, ADP learning.
11	**ML techniques in advanced AI systems**. Simulation of adaptive behaviour, grammatical inference, learning deterministic finite automata, L* algorithm, constructive induction, feature construction.

Distribution of workload for students (contact and independent work hours)

Lectures	**45 hours**
Laboratory work	**15 hours**
Individual students work	**100 hours**
Total:	**160 hours**

Structure of cumulative score and value of its constituent parts

Final written exam (50%), mid-term written exam (17%), and assessments of laboratory (practical) work (33%).

Recommended reference materials

No.	Publication year	Authors of publication and title	Publishing house	Number of copies in		
				University library	*Self-study rooms*	*Other libraries*
Basic materials						
1.	2011	Witten, I.H., E. Frank. Data Mining: Practical Machine Learning Tools and Techniques (3rd ed.).	Elsevier	unlimited online content, through ScienceDirect	unlimited, electronic book	
2.	2006	Theodoridis S., K. Koutroumbas. Pattern Recognition	Prentice-Hall	unlimited online content, through ScienceDirect	unlimited, electronic book	
3.	2007	Raškinis, G. Intelektika: Uždaviniai ir jų sprendimo būdai	Vytautas Magnus University	unlimited online content, through Moodle at VMU	unlimited, electronic book	
4.	Present	Wikipedia: The Free Encyclopedia	Wikimedia Foundation,	unlimited online content	unlimited online content	
Supplementary materials						
1.	2000	Jurafsky, D., J. H. Martin. Speech and Language Processing	Prentice-Hall	1	4	
2.	2010	De la Higuera, C. Grammatical Inference: Learning Automata and Grammars	Cambridge University Press	1		
3.	2002	Russel, S., P. Norvig Artificial Intelligence: A Modern Approach (2nd ed.)	Prentice-Hall	1	2	
4.	1998	Sutton, R. S., A. G. Barto. Reinforcement Learning: An Introduction	The MIT Press	e-book available online at http://webdocs.cs.ualberta.ca/~sutton/book/ebook/the-book.html		

Course programme designed by

Prof. Dr. Gailius Raškinis, Systems Analysis Department