AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY

Module name: Introduction to modern machine learning

Academic year: 2019/2020 Code: ZSDA-3-0164-s ECTS credits: 3

Faculty of: Szkoła Doktorska AGH

Field of study: Szkoła Doktorska AGH Specialty: —

Study level: Third-cycle studies Form and type of study: Full-time studies

Lecture language: Polski i Angielski Profile of education: Academic (A) Semester: 0

Course homepage: https://scikit-learn.org.edu.pl

Responsible teacher: Pawlak Mirosław (Miroslaw.Pawlak@agh.edu.pl)

Module summary

This graduate course introduces fundamental concepts and methods in machine learning. It describes several important supervised and unsupervised algorithms, provides the theoretical understanding of these methods, and illustrates key aspects for their applications. The course will examine methods ranging from kernel machines, model aggregation to regression, manifold learning and spectral clustering.

Description of learning outcomes for module

MLO code	Student after module completion has the knowledge/ knows how to/is able to	Connections with FLO	Method of learning outcomes verification (form of completion)
Skills: he can			
M_U001	Student is able to analyze machine learning methods accuracy and stability.	SDA3A_W02, SDA3A_W01	Execution of a project
M_U002	The students should be able to chose the proper learning method for the given practical problem, implement and assign the level of accuracy and interpretability.	SDA3A_U01	
Knowledge: he knows and understands			
M_W001	Making students to understand the basic properties of machine learning algorithms in terms of their accuracy, stability, and interpretability.	SDA3A_W02, SDA3A_W01	

M_W002	The students should grasp the fundamental knowledge of machine learning methodology that includes such concepts as supervised versus unsupervised learning, model selection and aggregation, regularization, variance-bias tradeoff, and parametric versus nonparametric methods.	SDA3A_W01	

Number of hours for each form of classes

	Form of classes										
Suma	Lectures	Auditorium classes	Laboratory classes	Project classes	Conversation seminar	Seminar classes	Practical classes	Fieldwork classes	Workshops	Prace kontrolne i przejściowe	Lektorat
20	20	0	0	0	0	0	0	0	0	0	0

FLO matrix in relation to forms of classes

MLO code	Student after module completion has the knowledge/ knows how to/is able to	Form of classes										
		Lectures	Auditorium classes	Laboratory classes	Project classes	Conversation seminar	Seminar classes	Practical classes	Fieldwork classes	Workshops	Prace kontrolne i przejściowe	Lektorat
Skills: he can												
M_U001	Student is able to analyze machine learning methods accuracy and stability.	+	-	-	-	-	-	-	-	-	-	-
M_U002	The students should be able to chose the proper learning method for the given practical problem, implement and assign the level of accuracy and interpretability.	-	-	-	-	-	-	-	-	-	-	-
Knowledge: he knows and understands												
M_W001	Making students to understand the basic properties of machine learning algorithms in terms of their accuracy, stability, and interpretability.	+	-	-	-	-	-	-	-	-	-	-

M_W002	The students should grasp the fundamental knowledge of machine learning methodology that includes such concepts as supervised versus unsupervised learning, model selection and aggregation, regularization, variance-bias tradeoff, and parametric versus nonparametric methods.	-	-	-	-	-	-	-	-	-	-	-

Student workload (ECTS credits balance)

Student activity form	Student workload
Udział w zajęciach dydaktycznych/praktyka	20 h
przygotowanie projektu, prezentacji, pracy pisemnej, sprawozdania	20 h
Summary student workload	40 h
Module ECTS credits	3 ECTS

Additional information

Module content

Lectures

What is Machine Learning ?
types of learning: supervised, unsupervised, reinforcement learning

Data Preprocessing
data visualization, data transformation, dealing with missing data

Fundamentals of Machine Learning
Bayes risk and rule, generative and discriminative ML,
Bayes risk consistency, bias-variance tradeoff, no free lunch theorem

Empirical Risk and Vapnik-Chervonenkis Theory
oracle bounds, concentration inequalities, VC dimension

Fundamental Algorithms for Supervised Learning: Parametric Algorithms
linear regression, logistic regression, support vector machines

Fundamental Algorithms for Supervised Learning: Nonparametric Algorithms
decision trees, k-nearest neighbors, kernel methods

Tuning and Learning Algorithm Selection
regularization, stochastic gradient descent, cross-validation

Learning Algorithm Aggregation
bagging and boosting, random forest

Model Performance Assessment
confusion matrix, ROC curves, accuracy and interpretability

Unsupervised Learning

confusion matrix, ROC curves, accuracy and interpretability

Unsupervised Learning
k-means and vector quantization

Principal Component Analysis
PCA methods, spectral clustering, robust PCA

Learning for sequences and time series data
sequence-to-sequence learning, functional data classification

Graphical Models for Machine Learning
Markov models, hidden MM

Learning in high-dimensional problems
data sparsity, Lasso algorithms

Other forms of learning
active learning, semi-supervised learning, metric learning

Teaching methods and techniques:
Lectures: Nie określono

Warunki i sposób zaliczenia poszczególnych form zajęć, w tym zasady zaliczeń poprawkowych, a także warunki dopuszczenia do egzaminu:
The course will be evaluated based on the activity during classes and the grade from the assigned projects. The project will be delivered both in the oral and written form.

Zasady udziału w poszczególnych zajęciach, ze wskazaniem, czy obecność studenta na zajęciach jest obowiązkowa:
Lectures:
– Attendance is mandatory: Yes
– Participation rules in classes: Presence in classes is obligatory
Two unjustified absences are allowed

Method of calculating the final grade
The final grade is calculated as a weighted average of the class activity and project marks.

Sposób i tryb wyrównywania zaległości powstałych wskutek nieobecności studenta na zajęciach:
In the case of absence the students are expected to present their projects in other suitable time.

Prerequisites and additional requirements
Basic knowledge of linear algebra, probability theory and statistics. Programming skill (MATLAB, Python, R)
is also expected.

Recommended literature and teaching resources
T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical
Learning, Springer, 2008.
M. Wainwright, High-Dimensional Statistics, Cambridge Press, 2019
M. Mohri, A. Rostamizadeh, and A. Talwakar, Foundation of Machine Learning, MIT Press 2012.
S. Shalev-Shwartz and S. Ben-David, Understanding Machine Learning: From Theory to Algorithms,

Cambridge Press, 2014.

Scientific publications of module course instructors related to the topic of the module

W. Greblicki and M. Pawlak. Nonparametric System Identification, Cambridge University Press, Cambridge, 2010.
W. Greblicki and M. Pawlak (2019). The weighted nearest neighbor estimate for a class of nonlinear time series systems, IEEE Trans. on Automatic Control, 64, pp.1550-1565.
W. Greblicki and M. Pawlak (2017). Hammerstein system identification with the nearest neighbor algorithm, IEEE Trans. on Information Theory, 63, 4746-4757.
D. Rzepka, M. Pawlak, D. Koscielnik and M. Miskowicz (2017). Bandwidth estimation from multiple level-crossings of stochastic signals, IEEE Trans. on Signal Processing, 65, 2488-2502.
M. Pawlak and U. Stadtmuller (2020). Nonparametric specification testing for signal models, IEEE Trans. Information Theory, to appear.
J. Lv and M. Pawlak (2019). Additive modeling and prediction of transient stability boundary in large-scale power systems using the group Lasso algorithm, Electrical Power and Energy Systems, 113, 963-970.
J. Lv, M. Pawlak and U. Annakkage (2017). Prediction of the transient stability boundary based on nonparametric additive modeling, IEEE Trans. Power Systems, 32, 4362-4369.

Additional information

None