



Foundations of Machine Teaching

Course description sheet

Basic information

Field of study Computer Science Major - Organisational unit Faculty of Computer Science Study level First-cycle (engineer) programme Form of study Full-time studies Profile General academic		Didactic cycle 2022/2023 Course code WIINF5.II20.08770.22 Lecture languages Polish Mandatoriness Elective Block General Modules Course related to scientific research Yes
Course coordinator	Witold Dzwinel	
Lecturer	Witold Dzwinel, Radosław Łazarz, Marcin Orchel	
Period Semester 6	Method of verification of the learning outcomes Completing the classes Activities and hours Lectures: 14 Laboratory classes: 14	Number of ECTS credits 3

Goals

C1	Celem jest zapoznanie studentów z głównymi algorytmami uczenia maszynowego i inżynierii cech.
----	---

Course's learning outcomes

Code	Outcomes in terms of	Learning outcomes prescribed to a field of study	Methods of verification
Knowledge - Student knows and understands:			
W1	Student zna i rozumie zasady nadzorowanej i nienadzorowanej analizy danych.	INF1A_W02, INF1A_W03, INF1A_W04	Activity during classes, Test
W2	Student ma wiedzę o metodach generowania, selekcji i ekstrakcji cech.	INF1A_W02, INF1A_W03	Activity during classes, Test
W3	Zna i rozumie podstawowy formalizm matematyczny związany z metodami uczenia maszynowego	INF1A_W01, INF1A_W02	Activity during classes, Test
Skills - Student can:			
U1	Student potrafi opracować system samouczący się, wykorzystujący dostarczone dane.	INF1A_U01, INF1A_U03, INF1A_U05, INF1A_U09	Execution of laboratory classes, Completion of laboratory classes
U2	Student potrafi opracować system, który można wykorzystać do klasyfikacji obiektów oraz zastosowania metod regresji do danych wielowymiarowych.	INF1A_U01, INF1A_U05, INF1A_U09	Execution of laboratory classes, Completion of laboratory classes
Social competences - Student is ready to:			
K1	Posiada kompetencje pracy w zespole i umie zorganizować sobie pracę nad złożonym zagadnieniem informatycznym	INF1A_K02, INF1A_K03, INF1A_K04	Execution of a project

Program content ensuring the achievement of the learning outcomes prescribed to the module

Introducing the student to the problems of image recognition and machine learning. The aim is to develop skills of representation generation, pattern recognition and building data analysis models.

Student workload

Activity form	Average amount of hours* needed to complete each activity form
Lectures	14
Laboratory classes	14
Preparation for classes	10
Realization of independently performed tasks	15
Examination or final test/colloquium	2
Contact hours	5
Preparation of project, presentation, essay, report	25
Student workload	Hours 85

Workload involving teacher	Hours 28
-----------------------------------	--------------------

* hour means 45 minutes

Program content

No.	Program content	Course's learning outcomes	Activities
1.	Ćwiczenie 1: Wizualizacja danych z repozytorium UCI przy pomocy metody PCA oraz LDA	U1, U2, K1	Laboratory classes
2.	Podstawowe pojęcia i problemy teorii uczenia maszynowego: Metody uczenia maszynowego, a metody inteligencji obliczeniowej. Podstawowe elementy statystyki wykorzystywane w uczeniu maszynowym. Entropia, cross-entropia. Podstawowe terminy: przestrzeń cech, wektor cech, obraz, klasyfikator, regresja. Podstawy teorii PAC. Fundamentalne twierdzenie PAC. Teoria Vapnika-Chervonienkis. Wymiar V-C. Podstawowe ograniczenia w teorii V-C. Moc klasyfikatora a wymiarowość przestrzeni cech.	W1, W2, W3	Lectures
3.	Ćwiczenie 2: Zastosowanie klasyfikatora najbliższego sąsiada do danych testowych. Problem przekleństwa wymiaru. Metody weryfikacji wyników: cross-validation, ROC, Precision-Recall. F-index etc.	U1, U2, K1	Laboratory classes
4.	Miary podobieństwa i niepodobieństwa: Pojęcie miary podobieństwa i niepodobieństwa. Przestrzenie Minkowskiego. Inne miary Chudhuri, Mahalanobisa. Różne aspekty wykorzystania miary Tanimoto. Miary zbior-zbiór. Miara Hausdorfa. Miary oparte na macierzach odległości. Miary podobieństwa wykorzystujące najbliższych sąsiadów. Dywergencja. Miara Kullbacka-Leiblera i inne. Miara "koparkowa" Wassersteina.	W1, W2, W3	Lectures
5.	Ćwiczenie 4: Zastosowanie klasyfikatora boostującego i porównanie do klasyfikatora SVM.	U1, U2, K1	Laboratory classes
6.	Cechy w metodach uczenia maszynowego, Metody ekstrakcji cech: Przekleństwo wymiaru. Dychotomie a wymiar przestrzeni cech. Filtry i wrapery. Proste algorytmy redukcji cech: FFS, FBS i pochodne. Metody selekcji cech oparte o heurystyki. Transformata Karhunen-Loeve. Metody PCA, LDA oraz SVD. Przykłady zastosowań. Nieliniowe metody ekstrakcji cech skalowanie wielowymiarowe i pochodne. Sieci złożone i deskryptory w metodach rozpoznawania obrazów. Reprezentacje i cechy w sieciach neuronowych. Embedding.	W1, W2, W3	Lectures
7.	Klasyfikatory proste: Klasyfikator 1-NN. Klasyfikator K-NN. Algorytmy poszukiwania najbliższego obrazu. Metoda Friedmana. Kd-drzewa. Przybliżone metody znajdowania grafu k-NN. Interesujące zastosowania klasyfikatora NN, kernel k-NN.. Klasyfikatory liniowe. Perceptron. Metody uczenia. Klasyfikator Fischera. Liniowy klasyfikator SVM. Klasyfikatory kawałkami liniowe.	W1, W2, W3	Lectures

No.	Program content	Course's learning outcomes	Activities
8.	Ćwiczenie 6: Zastosowanie algorytmów klasteryzacji k-means oraz algorytmów hierarchicznych do analizy dużych zbiorów danych.	U1, U2, K1	Laboratory classes
9.	Ćwiczenie 7 (projekt): Stworzenie własnego systemu klasyfikacji danych wielowymiarowych obejmującego wszystkie etapy uczenia maszynowego: generacja cech, selekcja cech, ekstrakcja i wizualizacja, klasteryzacja i klasyfikacja. Jako klasyfikator zastosować (w zależności od danych) głęboka sieć neuronową MLP lub CNN. Zadania są indywidualne i bazują na artykułach naukowych z dziedziny Uczenia Maszynowego.	U1, U2, K1	Laboratory classes
10.	Klasyfikatory Bayesowskie: Wielowymiarowe rozkłady gęstości prawdopodobieństwa. Zasada Bayesa. Klasyfikatory Bayesa. Drzewa Bayesa. Estymacja rozkładów losowych.	W1, W2, W3	Lectures
11.	Klasyfikatory nieliniowe SVM: Aproksymacja, interpolacja a problemy uczenia maszynowego. Nieliniowa interpretacja klasyfikatora SVM. Trik „kernelowy”. Przykłady zastosowań.	W1, W2, W3	Lectures
12.	Klasyfikatory zespołowe: Klasyczne klasyfikatory zespołowe. Klasyfikatory boostujące. Ada boost. Metody tworzenia klasyfikatorów zespołowych opartych o współdziałanie klasyfikatorów prostych. Dane niebalansowane. Klasyfikatory one-class.	W1, W2, W3	Lectures
13.	Metody uczenia nienadzorowanego: Klasteryzacja, Aglomeratywne i całościowe algorytmy klasteryzacji. Metoda k-means, Współczesne algorytmy klasteryzacji (DBSCAN, CHAMELEON, SNN, Afinity propagation).	W1, W2, W3	Lectures
14.	Nowoczesne kierunki rozwoju metod uczenia maszynowego: Sieci neuronowe wielowarstwowe. MLP oraz CNNs. Sieci neuronowe głębokie, sieci rekurencyjne LSTM, GANs. Metody osadzania w generacji cech dla niestrukturalnych danych. Podstawowe pojęcia problemu optymalnego transportu.	W1, W2, W3	Lectures
15.	Problemy uczenia maszynowego w analizie danych niestrukturalnych (tekst, grafy): Metody generacji reprezentacji tekstu (BOW, tf-idf, metody osadzania Word2Vec, Doc2Vec). Inne metody typu Paragraph Vector. Metody proste i generatywne analizy tekstu przy pomocy sieci głębokich. Metody osadzania grafów. Metody DeepWalk, Graph2Vec.	W1, W2, W3	Lectures

Extended information/Additional elements

Teaching methods and techniques :

Lectures, Discussion

Activities	Methods of verification	Credit conditions
Lectures	Activity during classes, Test	Prezentacja wyników projektów na kolokwium ustnym

Activities	Methods of verification	Credit conditions
Lab. classes	Execution of a project, Execution of laboratory classes, Completion of laboratory classes	Wykonanie wszystkich projektów w zadanym czasie. Ocena z prezentacji i tutoriala.

Method of determining the final grade

1. In order to obtain a positive final grade, it is necessary to obtain a positive assessment from the laboratory and the oral test (colloquium) during the submission of an individual project. The assessment is obtained for the quality of the project, which is assessed by the teacher. 2. We calculate the weighted average of laboratory assessments (75%) and lectures (25%) obtained on all dates. 3. Set a final grade based on the relationship: if $sr > 4.75$ then OK: = 5.0 else if $sr > 4.25$ then OK: = 4.5 else if $sr > 3.75$ then OK: = 4.0 else if $sr > 3.25$ then OK: = 3.5 else OK: = 3. 4. If a positive grade from the laboratory and passing the lecture was obtained on the first date and the final grade is lower than 5.0, the final grade is raised by 0.5. Absence in laboratory classes requires the exercise at home with a set (short) deadline for implementation. A larger number of absences than 3 requires also passing a verbal colloquium covering the exercise material before the teacher.

Prerequisites and additional requirements

1. Knowledge of the basic problems of mathematical analysis and algebra.
2. Knowledge of algorithms of numerical methods.
3. Good knowledge of algorithms.

Rules of participation in given classes, indicating whether student presence at the lecture is obligatory

Lectures: Studenci uczestniczą w zajęciach poznając kolejne treści nauczania zgodnie z sylabusem przedmiotu. Studenci winni na bieżąco zadawać pytania i wyjaśniać wątpliwości. Rejestracja audiowizualna wykładu wymaga zgody prowadzącego. Laboratory classes: Studenci wykonują ćwiczenia laboratoryjne zgodnie z materiałami udostępnionymi przez prowadzącego. Student jest zobowiązany do przygotowania się w przedmiocie wykonywanego ćwiczenia, co może zostać zweryfikowane kolokwium w formie ustnej lub pisemnej. Zaliczenie zajęć odbywa się na podstawie zaprezentowania rozwiązania postawionego problemu.

Literature

Obligatory

1. Literatura
2. 1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.
3. 2. Bishop Ch.K. Pattern Recognition and Machine Learning, Springer, New York, 2007.
4. 3. Theodoris S and Koutroumbas K, Pattern Recognition, Academic Press, San Diego, London, Boston, 1998
5. Literatura uzupełniająca:

Optional

1. Artykuły naukowe z czasopism: Pattern Recognition, Pattern Recognition Letters, Elsevier.

Scientific research and publications

Publications

1. 1. Czech, W., Mielczarek, W., Dzwiniel, W., Distance-based graph invariants for analysis and visualization of complex networks, Concurrency and Computation: Practice and Experience, 29(9):e4054, 2017 IF =1.133
2. 2. Dzwiniel, W., Wcisło, R., Czech, W., ivga: A fast force-directed method for interactive visualization of complex networks, Journal of Computational Science, 21C (2017) pp. 448-459 IF=1.748
3. 3. Wójcik P., Quellet T., Balcerzak M., Dzwiniel, W., Identification of biomarker genes for resistance to a pathogen by a novel method for meta-analysis of single-channel microarray datasets. Journal of Bioinformatics and Computational Biology, 13(4), 1550013 (19 pages) 2015. IF=0.785
4. 4. Pawliczek P, Dzwiniel W, Yuen DA, Visual exploration of data by using multidimensional scaling on multi-core CPU,

- GPU and MPI cluster, *Concurrency and Computation: Practice and Experience*, 26(3): 662-682, 2014. IF =0.997
5. 5. Kurdziel M, Boryczko K, Dzwinel W, Procrustes analysis of truncated least squares multidimensional scaling, *Computing and Informatics*, 31 (6), 1417-1440, 2012, IF =0.254
 6. 6. Dzwinel, W., Wcisło, R., ivhd: A robust linear-time and memory efficient method for visual exploratory data analysis, 13thConference on Machine Learning and Data Mining, MLDM, New York, July 15-20, 2017, *Lecture Notes of Artificial Intelligence, LNAI*, 10358, 345-360, 2017
 7. 7. Dzwinel, W., Wcisło, R., Very fast interactive visualization of large sets of high-dimensional data, 2015 International Conference of Computational Science, ICCS 2015, Reykjavik, 1-3.06.2015. *Procedia of Computer Science*, 51, 572-581, 2015.
 8. 8. Pawliczek, P., Dzwinel, W., Yuen, DA, Visual Exploration of Data with Multi-thread MIC Computer Architectures, in Rutkowski et al. (Eds.): ICAISC 2015, Part II, *Lecture Notes of Artificial Intelligence, LNAI* 9120, 25-35, 2015.

Learning outcomes prescribed to a field of study

Code	Content
INF1A_K02	Ma świadomość ważności i rozumie pozatechniczne aspekty i skutki działalności inżyniera-informatyka
INF1A_K03	Potrafi współdziałać i pracować w grupie, przyjmując w niej różne role, ma świadomość odpowiedzialności za pracę własną i za wspólnie realizowane zadania.
INF1A_K04	Potrafi odpowiednio określić priorytety służące realizacji określonego przez siebie lub innych zadania oraz adekwatnie zaplanować pracę.
INF1A_U01	Potrafi pozyskiwać informacje z literatury, baz danych i innych źródeł; potrafi integrować uzyskane informacje, dokonywać ich interpretacji, a także wyciągać wnioski oraz formułować i uzasadniać opinie. Ma umiejętność samokształcenia się, m.in. w celu podnoszenia kompetencji zawodowych.
INF1A_U03	Potrafi wykorzystywać różne techniki komunikacyjne dla realizacji zadań związanych z pracą inżyniera-informatyka.
INF1A_U05	Potrafi wykorzystać poznane metody i modele do tworzenia programów o charakterze użytkowym, a także potrafi adekwatnie wykorzystać znane algorytmy i struktury danych w budowie systemu komputerowego.
INF1A_U09	Potrafi zaprojektować, wykonać i oprogramować urządzenie z wykorzystaniem mikrokontrolerów lub mikroprocesorów
INF1A_W01	Ma wiedzę w zakresie matematyki i fizyki
INF1A_W02	Ma szczegółową wiedzę w zakresie podstaw algorytmiki, struktur danych oraz złożoności obliczeniowej, a także w zakresie podstaw teoretycznych budowy wybranych narzędzi i systemów informatycznych.
INF1A_W03	Ma szczegółową wiedzę w zakresie wybranych języków i technik programowania oraz uporządkowaną wiedzę w zakresie inżynierii oprogramowania, w tym projektowania i testowania systemów.
INF1A_W04	Ma szczegółową wiedzę w zakresie technik i narzędzi implementacyjnych uwzględniających wybrane aspekty budowy oprogramowania, w szczególności systemów baz danych, aplikacji działających w środowiskach sieciowych i budowy interfejsu użytkownika.