



Podstawy uczenia maszynowego

Karta opisu przedmiotu

Informacje podstawowe

Kierunek studiów Informatyka (kierunek wspólny - WI) Specjalność - Jednostka organizacyjna Wydział Informatyki Poziom kształcenia Studia inżynierskie I stopnia Forma studiów Stacjonarne Profil studiów Ogólnoakademicki		Cykl dydaktyczny 2022/2023 Kod przedmiotu WIINF5.II20.08770.22 Języki wykładowe polski Obligatoryjność Do wyboru Blok zajęciowy Przedmioty ogólne Przedmiot powiązany z badaniami naukowymi Tak
Koordynator przedmiotu	Witold Dzwiniel	
Prowadzący zajęcia	Witold Dzwiniel, Radosław Łazarz, Marcin Orchel	
Okres Semestr 6	Forma zaliczenia Zaliczenie Forma prowadzenia i godziny zajęć Wykład: 14 Ćwiczenia laboratoryjne: 14	Liczba punktów ECTS 3

Cele kształcenia dla przedmiotu

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

Efekty uczenia się dla przedmiotu

Kod	Efekty w zakresie	Kierunkowe efekty uczenia się	Metody weryfikacji
Wiedzy - Student zna i rozumie:			
W1	Student zna i rozumie zasady nadzorowanej i nienadzorowanej analizy danych.	INF1A_W02, INF1A_W03, INF1A_W04	Aktywność na zajęciach, Kolokwium
W2	Student ma wiedzę o metodach generowania, selekcji i ekstrakcji cech.	INF1A_W02, INF1A_W03	Aktywność na zajęciach, Kolokwium
W3	Zna i rozumie podstawowy formalizm matematyczny związany z metodami uczenia maszynowego	INF1A_W01, INF1A_W02	Aktywność na zajęciach, Kolokwium
Umiejętności - Student potrafi:			
U1	Student potrafi opracować system samouczący się, wykorzystujący dostarczone dane.	INF1A_U01, INF1A_U03, INF1A_U05, INF1A_U09	Wykonanie ćwiczeń laboratoryjnych, Zaliczenie laboratorium
U2	Student potrafi opracować system, który można wykorzystać do klasyfikacji obiektów oraz zastosowania metod regresji do danych wielowymiarowych.	INF1A_U01, INF1A_U05, INF1A_U09	Wykonanie ćwiczeń laboratoryjnych, Zaliczenie laboratorium
Kompetencje społecznych - Student jest gotów do:			
K1	Posiada kompetencje pracy w zespole i umie zorganizować sobie pracę nad złożonym zagadnieniem informatycznym	INF1A_K02, INF1A_K03, INF1A_K04	Wykonanie projektu

Treści programowe zapewniające uzyskanie efektów uczenia się dla modułu zajęć

Wprowadzenie studenta w problematykę rozpoznawania obrazów i uczenia maszynowego. Celem jest wykształcenie umiejętności generacji reprezentacji danych, rozpoznawania wzorców i budowy modeli analizy danych. Kurs obejmuje podstawową terminologię oraz elementy teorii klasycznego uczenia maszynowego, a także wprowadzenie do zrozumienia współczesnych metod uczenia głębokiego w kontekście aktualnych wyzwań związanych z gwałtownym rozwojem metod sztucznej inteligencji.

Nakład pracy studenta

Rodzaje zajęć studenta	Średnia liczba godzin* przeznaczonych na zrealizowane aktywności
Wykład	14
Ćwiczenia laboratoryjne	14
Przygotowanie do zajęć	10
Samodzielne studiowanie tematyki zajęć	15
Egzamin lub kolokwium zaliczeniowe	2
Dodatkowe godziny kontaktowe	5
Przygotowanie projektu, prezentacji, pracy pisemnej, sprawozdania	25

Łączny nakład pracy studenta	Liczba godzin 85
Liczba godzin kontaktowych	Liczba godzin 28

* godzina (lekcyjna) oznacza 45 minut

Treści programowe

Lp.	Treści programowe	Efekty uczenia się dla przedmiotu	Formy prowadzenia zajęć
1.	Ćwiczenie 1: Wizualizacja danych z repozytorium UCI przy pomocy metody PCA oraz LDA	U1, U2, K1	Ćwiczenia laboratoryjne
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	Wykład
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	Ćwiczenia laboratoryjne
4.	Miary podobieństwa i niepodobieństwa: Pojęcie miary podobieństwa i niepodobieństwa. Przestrzenie Minkowskiego. Inne miary Chudhuri, Mahalanobisa. Różne aspekty wykorzystania miary Tanimoto. Miary zbiór-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	Wykład
5.	Ćwiczenie 4: Zastosowanie klasyfikatora boostującego i prorównanie do klasyfikatora SVM.	U1, U2, K1	Ćwiczenia laboratoryjne
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	Wykład

Lp.	Treści programowe	Efekty uczenia się dla przedmiotu	Formy prowadzenia zajęć
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	Wykład
8.	Ćwiczenie 6: Zastosowanie algorytmów klasteryzacji k-means oraz algorytmów hierarchicznych do analizy dużych zbiorów danych.	U1, U2, K1	Ćwiczenia laboratoryjne
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	Ćwiczenia laboratoryjne
10.	Klasyfikatory Bayesowskie: Wielowymiarowe rozkładów gęstości prawdopodobieństwa. Zasada Bayesa. Klasyfikatory Bayesa. Drzewa Bayesa. Estymacja rozkładów losowych.	W1, W2, W3	Wykład
11.	Klasyfikatory nieliniowe SVM: Aproksymacja, interpolacja a problemy uczenia maszynowego. Nieliniowa interpretacja klasyfikatora SVM. Trik „kernelowy”. Przykłady zastosowań.	W1, W2, W3	Wykład
12.	Klasyfikatory zespołowe: Klasyczne klasyfikatory zespołowe. Klasyfikatory boostujące. Ada boost. Metody tworzenia klasyfikatorów zespołowych opartych o współdziałanie klasyfikatorów prostych. Dane niezbalansowane. Klasyfikatory one-class.	W1, W2, W3	Wykład
13.	Metody uczenia nienadzorowanego: Klasteryzacja, Aglomeratywne i całościowe algorytmy klasteryzacji. Metoda k-means, Współczesne algorytmy klasteryzacji (DBSCAN, CHAMELEON, SNN, Afinity propagation).	W1, W2, W3	Wykład
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	Wykład
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	Wykład

Informacje rozszerzone

Metody i techniki kształcenia :

Mini wykład, Dyskusja

Rodzaj zajęć	Metody zaliczenia	Warunki zaliczenia przedmiotu
Wykład	Aktywność na zajęciach, Kolokwium	Prezentacja wyników projektów na kolokwium ustnym
Ćwiczenia laboratoryjne	Wykonanie projektu, Wykonanie ćwiczeń laboratoryjnych, Zaliczenie laboratorium	Wykonanie wszystkich projektów w zadanym czasie. Ocena z prezentacji i tutoriala.

Dodatkowy opis

Studenci na ćwiczeniach prowadzą tutoriale na tematy związane z wykładem
Wykłady z przedmiotu będą prowadzone w sposób zdalny z wykorzystaniem platformy Webex/Teams.
Pozostałe zajęcia będą odbywać się w salach. Dotyczy to także zaliczeń i egzaminów odbywających się w sesjach egzaminacyjnych.

Warunki i sposób zaliczenia poszczególnych form zajęć, w tym zasady zaliczeń poprawkowych, a także warunki dopuszczenia do egzaminu

Warunkiem zaliczenia ćwiczeń laboratoryjnych jest wykonanie zadania oraz napisanie raportu z jego wykonania. Raport powinien zawierać: założenia, dyskusję rezultatów i wnioski. Warunkiem zaliczenia projektu jest zrozumienie oraz implementacja algorytmu uczenia maszynowego z (w miarę) nowego artykułu naukowego.

Sposób obliczania oceny końcowej

1. Aby uzyskać pozytywną ocenę końcową niezbędne jest uzyskanie pozytywnej oceny z laboratorium oraz kolokwium (ustnego) podczas oddawania projektu indywidualnego. Ocenę uzyskuje się za jakość wykonania projektu, którą ocenia prowadzący. 2. Obliczamy średnią ważoną z ocen z laboratorium (75%) i prezentacji (tutorialu) (25%) uzyskanych we wszystkich terminach. 3. Wyznaczymy ocenę końcową na podstawie zależności: $\text{if } sr > 4.75 \text{ then OK} = 5.0 \text{ else if } sr > 4.25 \text{ then OK} = 4.5 \text{ else if } sr > 3.75 \text{ then OK} = 4.0 \text{ else if } sr > 3.25 \text{ then OK} = 3.5 \text{ else OK} = 3$ 4. Jeżeli pozytywną ocenę z laboratorium i zaliczenia wykładu uzyskano w pierwszym terminie oraz ocena końcowa jest mniejsza niż 5.0 to ocena końcowa jest podnoszona o 0.5 Nieobecność na zajęciach laboratoryjnych wymaga wykonania ćwiczenia w domu z zadanym (krótkim) terminem realizacji. Większa liczba nieobecności niż 3 wymaga ponadto zdania przed prowadzącym ustnego kolokwium obejmującego materiał ćwiczeń.

Sposób i tryb wyrównywania zaległości powstałych wskutek nieobecności studenta na zajęciach

Do uzgodnienia z prowadzącym zajęcia. Najczęściej kolokwium (ustne) w godzinach konsultacji.

Wymagania wstępne i dodatkowe

1. Znajomość podstawowych zagadnień analizy matematycznej, algebry oraz statystyki matematycznej
2. Znajomość algorytmów metod numerycznych.
3. Dobra znajomość algorytmiki.

Zasady udziału w poszczególnych zajęciach, ze wskazaniem, czy obecność studenta na zajęciach jest obowiązkowa

Wykład: 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. Ćwiczenia laboratoryjne: 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.

Literatura

Obowiązkowa

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:

Dodatkowa

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

Badania i publikacje

Publikacje

1. 1. Czech, W., Mielczarek, W., Dzwinel, 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. Dzwinel, 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., Dzwinel, 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, Dzwinel 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.

Kierunkowe efekty uczenia się

Kod	Treść
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.