



KARTA OPISU PRZEDMIOTU - SYLABUS

Nazwa przedmiotu

Wizualizacja danych wielowymiarowych [S2Inf1-SzInt>WDW]

Przedmiot

Kierunek studiów
Informatyka

Rok/Semestr
2/3

Studia w zakresie (specjalność)
Sztuczna inteligencja

Profil studiów
ogólnoakademicki

Poziom studiów
drugiego stopnia

Język oferowanego przedmiotu
polski

Forma studiów
stacjonarne

Wymagalność
obligatoryjny

Liczba godzin

Wykład
16

Laboratorium
16

Inne
0

Ćwiczenia
0

Projekty/seminaria
0

Liczba punktów ECTS

2,00

Koordynatorzy

dr hab. inż. Robert Susmaga
robert.susmaga@put.poznan.pl

Wykładowcy

Wymagania wstępne

Wiedza: Podstawowa wiedza z algebry liniowej (proste operacje na wektorach i macierzach) oraz geometrii analitycznej (tworzenie wykresów prostych funkcji). Umiejętności: Umiejętność projektowania i implementowania prostych programów realizujących podstawowe operacje wektorowo-macierzowe i generujących wykresy podstawowych funkcji. (Pożądane) Ciekawość poznawcza, wytrwałość w dążeniu do poszerzania swojej wiedzy, spora doza uczciwości i kultury osobistej.

Cel przedmiotu

1. Przekazanie studentom szczegółowej wiedzy dotyczącej: a) leżących u podstaw analiz wielowymiarowych wybranych operacji wektorowo-macierzowych w przestrzeniach wielowymiarowych oraz rozkładu macierzy kwadratowych względem wartości własnych (ang. "eigenvalue decomposition", EVD), wraz z ich zastosowaniami w przedstawionych poniżej metodach, b) wybranych metod przekształcania i wizualizowania danych wielowymiarowych, w tym metody składowych głównych (ang. "principal component analysis", PCA), metody skalowania wielowymiarowego (ang. "multidimensional scaling", MDS), a także sposobów reprezentacji danych w barycentrycznych układach współrzędnych. 2. Rozwijanie u studentów umiejętności a) identyfikowania, formułowania i rozwiązywania problemów badawczych związanych z analizowaniem i wizualizowaniem danych wielowymiarowych, b) projektowania, tworzenia i testowania programów implementujących omawiane metody.

Przedmiotowe efekty uczenia się

Wiedza:

w ramach przedmiotu studenci:

zdołają zdobyć zaawansowaną wiedzę szczegółową dotyczącą wybranych zagadnień z zakresu analizy danych wielowymiarowych, w szczególności dotyczących redukcji wymiarowości (przede wszystkim: z zakresu metody pca), wraz z ich zastosowaniami w selekcji, wygładzaniu i wizualizacji danych wielowymiarowych (przede wszystkim: metodę mds, systemy współrzędnych barycentrycznych)

[k2st_w3]

zdołają zdobyć wiedzę o trendach rozwojowych i najistotniejszych nowych osiągnięciach informatyki, w szczególności na polu uczenia maszynowego i eksploracji danych, w których najnowsze osiągnięcia najczęściej wykorzystują skuteczne algorytmy optymalizacji w przestrzeniach wielowymiarowych

[k2st_w4]

Umiejętności:

dzięki przedmiotowi studenci:

potrafią wykorzystać do formułowania i rozwiązywania zadań inżynierskich i prostych problemów badawczych metody analityczne, symulacyjne oraz eksperymentalne, w szczególności dotyczące przekształceń i analiz danych wielowymiarowych [k2st_u4]

potrafią — przy formułowaniu i rozwiązywaniu zadań inżynierskich, w szczególności dotyczących uczenia maszynowego i eksploracji danych — integrować wiedzę z różnych obszarów matematyki (algebra liniowa, geometria wielowymiarowa, itp.), uwzględniając także aspekty pozatechniczne, np. aspekty komunikacji człowiek-komputer w wizualizacji [k2st_u5]

potrafią ocenić przydatność i możliwość wykorzystania nowych osiągnięć (metod i narzędzi) oraz nowych produktów informatycznych, przede wszystkim z dziedzin dotyczących analizy i przetwarzania danych wielowymiarowych (np. metod redukcji/selekcji cech) [k2st_u6]

potrafią dokonać krytycznej analizy istniejących rozwiązań technicznych — w szczególności (w dziedzinach analizy danych) rozwiązań wymagających redukcji wymiarowości — oraz zaproponować ich ulepszenia (usprawnienia) [k2st_u8]

potrafią — zgodnie z zadaną specyfikacją, uwzględniającą aspekty pozatechniczne — zaprojektować złożony system informatyczny oraz zrealizować ten projekt (co najmniej w części) używając właściwych metod, technik i narzędzi, w tym przystosowując do tego celu istniejące lub opracowując nowe narzędzia analizy danych wielowymiarowych, w szczególności: redukcji wymiarowości [k2st_u11]

Kompetencje społeczne:

studenci tego przedmiotu:

rozumieją, że w informatyce wiedza i umiejętności bardzo szybko stają się przestarzałe [k2st_k1]

rozumieją znaczenie wykorzystywania najnowszej wiedzy z zakresu informatyki w rozwiązywaniu problemów badawczych i praktycznych [k2st_k2]

Metody weryfikacji efektów uczenia się i kryteria oceny

Efekty uczenia się przedstawione wyżej weryfikowane są w następujący sposób:

Ocena formująca: (w zakresie laboratoriów):

- na podstawie oceny bieżącego postępu realizacji przydzielanych zadań.

Ocena podsumowująca (zarówno w zakresie wykładów jak i laboratoriów):

- ocena wiedzy i umiejętności wykazanych na pisemnym sprawdzianie wiedzy zawierającym w kilk (4-6) zadań (analogicznych do zadań prezentowanych na zajęciach); czas przewidziany na zaliczenie to 60-90

(wykłady) / 30-45 (laboratoria) minut; aby uzyskać ocenę pozytywną trzeba zdobyć przynajmniej $1 + \lfloor m/2 \rfloor$ (zaokrąglenie w dół) punktów, gdzie m jest punktacją maksymalną (np. aby uzyskać ocenę pozytywną przy $m = 30$ należy zdobyć przynajmniej 16 punktów).

Treści programowe

Wybrane metody przekształcania i wizualizowania danych wielowymiarowych, w tym reprezentacje danych w barycentrycznych układach współrzędnych oraz specjalizowane metody wizualizacyjne; w szczególności metoda składowych głównych (ang. "principal component analysis", PCA) oraz metoda skalowania wielowymiarowego (ang. "multidimensional scaling", MDS).

Tematyka zajęć

(Wykłady) Wstęp: Przestrzenie wielowymiarowe. Idea wizualizacji, wady i zalety, wizualizacja danych "niskowymiarowych" i jej różne aspekty; wizualizacja danych wielowymiarowych i jej różne aspekty. Barycentryczne układy współrzędnych, trójwymiarowe i czterowymiarowe. Zastosowanie czterowymiarowych układów barycentrycznych: wizualizacja miar konfirmacji i miar trafności klasyfikowania.

Wektory i macierze, podstawowe operacje wektorowe i macierzowe. Wielowymiarowe przestrzenie wektorowe, iloczyn skalarny wektorów, rzut wektora, kąt między wektorami, wektory ortogonalne, norma wektora. Macierze i podstawowe operacje na macierzach, interpretacja macierzy jako nośników danych i jako operatorów przekształcających. Podstawowe charakterystyki skalarne macierzy, macierze odwrotne i macierze ortogonalne oraz ich interpretacja graficzna. Analiza spektralna macierzy: wartości własne i ich właściwości, wektory własne i ich właściwości. Idea rozkładu macierzy, rozkład względem wartości własnych (ang. "eigenvalue decomposition", EVD): konstrukcja i podstawowe właściwości. Interpretacje i zastosowania rozkładów w algebrze macierzy i w analizie danych (w szczególności: redukcja wymiarowości, wygładzanie i kompresja). Idea metody składowych głównych (ang. "Principal Component Analysis", PCA), zależność zmiennych, macierze kowariancji/korelacji, procedura metody PCA, wykorzystanie rozkładów macierzy w PCA, dobór liczby redukowanych składowych, operacja odtwarzania danych, przykładowe zastosowania PCA. Idea metody skalowania wielowymiarowego (ang. "Multidimensional Scaling", MDS), macierze odległości, mapy obiektów, procedura metody MDS; wykorzystanie rozkładów macierzy w MDS, przykładowe zastosowania MDS. Opcjonalnie: idea analizy korespondencji (ang. "Correspondence Analysis", CA), idea metody wizualizacji nieliniowej t-SNE (ang. "t-distributed Stochastic Neighbour Embedding").

(Laboratoria) Wprowadzenie do języka Python i wybranych bibliotek tego języka: NumPy i Matplotlib. Tworzenie prostych programów działających na danych skalarnych, wektorowych i macierzowych. Wizualizacja danych skalarnych, wektorowych i macierzowych, wykresy rozrzutu, barycentryczne układy współrzędnych. Iloczyn skalarny wektorów, ortogonalność wektorów, normy wektorów. Rozkład EVD macierzy, przykładowe zastosowania w funkcjach macierzowych. Metoda PCA, przykładowe zastosowania w redukcji wymiarowości i wizualizacji. Metoda MDS, przykładowe zastosowania w wizualizacji. Opcjonalnie: metoda CA, metoda t-SNE.

Metody dydaktyczne

Wykłady: prezentacja multimedialna uzupełniona przykładami obliczeniowymi, demonstracja wybranych wizualizacji.

Laboratoria: modelowanie przykładowych problemów dotyczących przetwarzania i wizualizacji danych wielowymiarowych i rozwiązywanie tych problemów, wykonywanie eksperymentów symulacyjnych, dyskusja, praca w zespole, demonstracja i pokaz multimedialny.

Literatura

Podstawowa

1. G. Banaszak, W. Gajda: Elementy algebry liniowej część I i II, WNT, Warszawa, 2002

2. J. Koronacki, J. Ćwik: Statystyczne systemy uczące się, WNT, Warszawa, 2005

Uzupełniająca

1. I.T. Jolliffe: Principal Component Analysis, Springer-Verlag, Nowy Jork, USA, 2002

2. I. Borg, P.J.F. Groenen: Modern Multidimensional Scaling, Springer Science+Business Media, Nowy Jork, USA, 2005

3. M. Greenacre: Correspondence Analysis in Practice, Chapman & Hall/CRC Press, Nowy Jork, USA, 2007

Bilans nakładu pracy przeciętnego studenta

	Godzin	ECTS
Łączny nakład pracy	50	2,00
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	32	1,00
Praca własna studenta (studia literaturowe, przygotowanie do zajęć laboratoryjnych/ćwiczeń, przygotowanie do kolokwium/egzaminu, wykonanie projektu)	18	1,00