

KARTA OPISU MODUŁU KSZTAŁCENIA		
Nazwa modułu/przedmiotu Skalowanie i wizualizacja danych wielowymiarowych		Kod 1010512331010510208
Kierunek studiów Informatyka	Profil kształcenia (ogólnoakademicki, praktyczny) ogólnoakademicki	Rok / Semestr 2 / 3
Ścieżka obieralności/specjalność Inteligentne systemy wspomagania decyzji	Przedmiot oferowany w języku: polski	Kurs (obligatoryjny/obieralny) obligatoryjny
Stopień studiów: II stopień	Forma studiów (stacjonarna/niestacjonarna) stacjonarna	
Godziny Wykłady: 30 Ćwiczenia: - Laboratoria: 15 Projekty/seminaria: -		Liczba punktów 4
Status przedmiotu w programie studiów (podstawowy, kierunkowy, inny) inny		(ogólnouczelniany, z innego kierunku) ogólnouczelniany
Obszar(y) kształcenia i dziedzina(y) nauki i sztuki		Podział ECTS (liczba i %)
Odpowiedzialny za przedmiot / wykładowca:		
<p>Robert Susmaga email: Robert.Susmaga@cs.put.poznan.pl tel. 61 6652934 Instytut Informatyki ul. Piotrowo 2, 60-965 Poznań</p>		
Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:		
1	Wiedza:	Podstawowa wiedza z algebry liniowej (proste operacje na wektorach i macierzach) oraz geometrii analitycznej (tworzenie wykresów prostych funkcji).
2	Umiejętności:	Projektowanie, implementowanie i testowanie prostych programów komputerowych realizujących podstawowe operacje wektorowo-macierzowe i generujące wykresy podstawowych funkcji.
3	Kompetencje społeczne	(Pożądane) Ciekawość poznawcza, wytrwałość w dążeniu do poszerzania swojej wiedzy, spora doza uczciwości i kultury osobistej.
Cel przedmiotu:		
<p>1. Przekazanie studentom szczegółowej wiedzy dotyczącej:</p> <p>a) wybranych aspektów algebry liniowej, w szczególności: operacji wektorowo-macierzowych w przestrzeniach wielowymiarowych oraz rozkładu macierzy kwadratowych względem wartości własnych (ang. 'eigenvalue decomposition', EVD) (opcjonalnie: rozkładu macierzy względem wartości osobliwych (ang. 'singular value decomposition', SVD)), wraz z ich zastosowaniami w poniższych metodach,</p> <p>b) wybranych metod skalowania i wizualizacji danych, w tym metody składowych głównych (ang. 'principal component analysis', PCA), metody analizy czynnikowej (ang. 'factor analysis', FA), metody skalowania wielowymiarowego (ang. 'multidimensional scaling', MDS) oraz metody analizy korespondencji (ang. 'correspondence analysis', CA).</p> <p>2. Rozwijanie u studentów umiejętności</p> <p>a) identyfikowania, formułowania i rozwiązywania problemów badawczych związanych ze skalowaniem i wizualizacją danych wielowymiarowych,</p> <p>b) projektowania, tworzenia i testowania programów implementujących omawiane metody.</p>		
Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia		
Wiedza:		

<p>1. ma uporządkowaną i podbudowaną teoretycznie wiedzę ogólną związaną z kluczowymi zagadnieniami z zakresu analizy danych, w szczególności dotyczącymi analizy danych wielowymiarowych, wraz z ich zaletami (interpretacje geometryczne) i wadami ("kłątwa wymiarowości") - [K2st_W2]</p> <p>2. ma 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 anych wielowymiarowych (przede wszystkim: metodę MDS, systemy współrzędnych barycentrycznych) - [K2st_W3]</p> <p>3. ma 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]</p> <p>4. zna zaawansowane metody, techniki i narzędzia stosowane przy rozwiązywaniu złożonych zadań inżynierskich i prowadzeniu prac badawczych w obszarze analizy danych wielowymiarowych, przede wszystkim elementy analizy wariancji, przestrzeni liniowych i rozkładów macierzy (głównie: EVD i SVD) - [K2st_W6]</p>
<p>Umiejętności:</p> <p>1. potrafi 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]</p> <p>2. potrafi ? 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 - [K2st_U5]</p> <p>3. potrafi 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]</p> <p>4. potrafi dokonać krytycznej analizy istniejących rozwiązań technicznych (w szczególności, np. dziedzinach: uczenia maszynowego i eksploracji danych -- rozwiązań wymagających skutecznej redukcji wymiarowości danych) oraz zaproponować ich ulepszenia (usprawnienia) - [K2st_U8]</p> <p>5. potrafi - zgodnie z zadaną specyfikacją, uwzględniając 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]</p>
<p>Kompetencje społeczne:</p> <p>1. rozumie, że w informatyce wiedza i umiejętności bardzo szybko stają się przestarzałe - [K2st_K1]</p> <p>2. potrafi odpowiednio określić priorytety służące realizacji określonego przez siebie lub innych zadania - [K2st_K2]</p>

<p>Sposoby sprawdzenia efektów kształcenia</p>
<p>Efekty kształcenia przedstawione wyżej weryfikowane są w następujący sposób:</p> <p>Ocena formująca: (w zakresie laboratoriów):</p> <ul style="list-style-type: none"> - na podstawie oceny bieżącego postępu realizacji zadań. <p>Ocena podsumowująca (zarówno w zakresie wykładów jak i laboratoriów):</p> <ul style="list-style-type: none"> - 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).
<p>Treści programowe</p>
<p>Program wykładu obejmuje następujące zagadnienia: Wstęp: Idea mierzenia i skalowania, typy skal pomiarowych, podstawowe transformacje zmiennych, przykłady pomiarów (fizyka), przykłady skalowań (psychologia); idea zmiennych ukrytych (przypadek dwuwymiarowy); dane i pomiary wielowymiarowe, wielowymiarowe zmienne ukryte. Idea wizualizacji, paradygmaty wizualizacji, wady i zalety, wizualizacja danych "niskowymiarowych" i jej różne aspekty; wizualizacja danych wielowymiarowych i jej różne aspekty. Twierdzenie Vivianiego i 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, wyrażenia i funkcje skalarne w notacji macierzowo-wektorowej. Wielowymiarowe przestrzenie wektorowe, iloczyn skalarny wektorów, rzut wektora, kąt między wektorami, wektory ortogonalne; (euklidesowa) norma wektora. Miary zależności zmiennych, kowariancja, korelacja, miary podobieństwa, miary odległości (euklidesowa, Minkowskiego, Mahalanobisa, "miara" kosinusowa). Macierze i podstawowe operacje na macierzach, interpretacja macierzy jako nośników danych (tabele zmiennych/obiektów, tabele kontyngencji, macierze kowariancji/korelacji) i jako operatorów przekształcających (macierze skalujące, macierze rzutu, macierze rotacji/translacji). Podstawowe charakterystyki skalarnie macierzy: wyznacznik, norma; 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). Opcjonalnie: wartości i wektory osobliwe macierzy, rozkład względem wartości osobliwych (ang. "singular value decomposition", SVD. Interpretacje i zastosowania rozkładów w algebrze macierzy (funkcje macierzowe, odwrotności i pseudoodwrotności w problemie regresji liniowej) i analizie danych (redukcja wymiarowości/kompresja i wygładzanie danych w uogólnionym problemie regresji liniowej). Metody skalowania i wizualizacji. Idea metody składowych głównych (ang. "principal component analysis", PCA), zależności 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),</p>

macierze odległości, mapy obiektów, procedura metody MDS; wykorzystanie rozkładów macierzy w MDS, warianty metody (skalowanie niesymetrycznych danych, skalowanie zmiennych), przykładowe zastosowania MDS. Opcjonalnie: idea analizy czynnikowej (ang. "factor analysis", FA), założenia i ograniczenia modelu, procedura metody FA; wykorzystanie rozkładów macierzy w FA, składowe główne jako czynniki, rotacja czynników, przykładowe zastosowania FA. Opcjonalnie: idea analizy korespondencji (ang. "correspondence analysis", CA), tablice kontyngencji, inercje i profile, odległość c2, procedura metody CA; wykorzystanie rozkładów macierzy w CA, uogólnienia metody, przykładowe zastosowania CA. Opcjonalnie: idea wybranej metody wizualizacji nieliniowej (np. ang. "t-distributed Stochastic Neighbour Embedding", t-SNE), dywergencja Kullbacka-Leiblera, problem optymalizacyjny, przykładowe zastosowania t-SNE. Program laboratorium obejmuje następujące zagadnienia: 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, wektory ortogonalne, macierze ortogonalne. Rozkład EVD macierzy, przykładowe zastosowania w funkcjach macierzowych. Rozkład SVD macierzy, przykładowe zastosowania w kompresji danych. Metoda PCA, przykładowe zastosowania w redukcji wymiarowości. Metoda MDS, przykładowe zastosowania w wizualizacji danych ciągłych.

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
3. B. Kaczmarek: Elementy algebry i analizy macierzy, Wydawnictwo PP, 1689, Poznań, 1992

Literatura uzupełniająca:

1. A. Biela: Skalowanie wielowymiarowe jako metoda badań naukowych, Towarzystwo Naukowe KUL, Lublin 1992
2. I.T. Jolliffe: Principal Component Analysis, Springer-Verlag, Nowy Jork, USA, 2002
3. H.H. Harman: Modern Factor Analysis, The University of Chicago Press, Chicago, 1967
4. T.F. Cox, M.A.A. Cox: Multidimensional Scaling, Chapman & Hall/CRC Press, Boca Raton, USA, 2001
5. I. Borg, P.J.F. Groenen: Modern Multidimensional Scaling, Springer Science+Business Media, Nowy Jork, USA, 2005
6. M. Greenacre: Correspondence Analysis in Practice, Chapman & Hall/CRC Press, Nowy Jork, USA, 2007
7. H. Dudycz: Wizualizacja danych, Wydawnictwo AE, Wrocław, 1998

Bilans nakładu pracy przeciętnego studenta

Czynność	Czas (godz.)
1. udział w wykładach	30
2. udział w zajęciach laboratoryjnych	15
3. teoretyczne przygotowanie do ćwiczeń laboratoryjnych (literatura przedmiotu, języki i systemy programowania)	5
4. praktyczna praca nad zebraniem danych, tworzeniem i testowaniem oprogramowania	10
5. przeprowadzenie badań / eksperymentów obliczeniowych	5
6. tworzenie sprawozdań	10
7. przygotowanie do sprawdzianu (laboratorium)	5
8. przygotowanie do sprawdzianu (wykład)	15

Obciążenie pracą studenta

forma aktywności	godzin	ECTS
Łączny nakład pracy	95	4
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	45	2
Zajęcia o charakterze praktycznym	40	1