



KARTA OPISU PRZEDMIOTU - SYLABUS

Nazwa przedmiotu

Zaawansowane aspekty redukcji wymiarowości [S2SI1E>ZARW]

Przedmiot

Kierunek studiów

Sztuczna inteligencja/Artificial Intelligence

Rok/Semestr

1/2

Studia w zakresie (specjalność)

–

Profil studiów

ogólnoakademicki

Poziom studiów

drugiego stopnia

Język oferowanego przedmiotu

angielski

Forma studiów

stacjonarne

Wymagalność

obieralny

Liczba godzin

Wykład

15

Laboratorium

15

Inne

0

Ćwiczenia

0

Projekty/seminaria

0

Liczba punktów ECTS

3,00

Koordynatorzy

dr hab. inż. Robert Susmaga

robert.susmaga@put.poznan.pl

Wykładowcy

Wymagania wstępne

Podstawowa wiedza z algebry liniowej (proste operacje na wektorach i macierzach) oraz geometrii analitycznej (tworzenie wykresów prostych funkcji). Projektowanie, implementowanie i testowanie prostych programów komputerowych 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) wybranych aspektów algebry liniowej, w szczególności: operacji wektorowo-macierzowych w przestrzeniach wielowymiarowych oraz rozkładu macierzy kwadratowych względem wartości własnych (ang. 'eigenvalue decomposition', EVD) i rozkładu macierzy względem wartości osobliwych (ang. 'singular value decomposition', SVD), wraz z ich zastosowaniami w poniższych metodach, b) wybranych metod redukcji wymiarowości, w tym metody składowych głównych (ang. 'principal component analysis', PCA) i metody skalowania wielowymiarowego (ang. 'multidimensional scaling', MDS), a także metod "zanurzeń" SNE (ang. 'stochastic neighbor embedding') i t-SNE (ang. 't-distributed stochastic neighbor embedding'). c) zastosowań metod redukcji wymiarowości danych do wizualizacji danych wielowymiarowych, w szczególności: prostych wizualizacji opartych na barycentrycznych układach współrzędnych i wizualizacji opartych na zaawansowanych metodach redukcji wymiarowości. 2. Rozwijanie u studentów umiejętności a) identyfikowania, formułowania i rozwiązywania problemów badawczych związanych z redukcją wymiarowości i jej zastosowaniami w wizualizacji danych wielowymiarowych, b) projektowania, tworzenia i testowania programów implementujących omawiane metody.

Przedmiotowe efekty uczenia się

Wiedza

Student:

- ma zaawansowaną i pogłębioną wiedzę z zakresu metod, narzędzi i środowisk dotyczących redukcji (K2st_W1)
- 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 (K2st_W2)
- 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 danych wielowymiarowych (przede wszystkim: metodę MDS, systemy współrzędnych barycentrycznych) (K2st_W3)
- 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)
- zna zaawansowane metody, techniki i narzędzia stosowane przy rozwiązywaniu złożonych zadań inżynierskich i prowadzeniu prac badawczych w obszarze analizy danych wielowymiarowych, przede wszystkim redukcji wymiarowości i jej zastosowań w wizualizacji (K2st_W6).

Umiejętności

Student

- potrafi pozyskiwać informacje z literatury, baz danych oraz innych źródeł (w języku polskim i angielskim), integrować je, dokonywać ich interpretacji i krytycznej oceny, wyciągać wnioski oraz formułować i wyczerpująco uzasadniać opinie (K2st_U1)
- potrafi planować i przeprowadzać eksperymenty, interpretować uzyskane wyniki i wyciągać wnioski oraz formułować i weryfikować hipotezy związane problemami redukcji wymiarowości (K2st_U3)
- 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)
- 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)
- 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)
- 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)
- potrafi ocenić przydatność metod i narzędzi służących do rozwiązania zadania redukcji wymiarowości (K2st_U9)
- potrafi - stosując m.in. koncepcyjnie nowe metody - rozwiązywać złożone zadania informatyczne dotyczące redukcji wymiarowości (K2st_U10)

-- 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).
Kompetencje społeczne

Student:

-- rozumie, że w informatyce wiedza i umiejętności bardzo szybko stają się przestarzałe (K2st_K1),
-- potrafi odpowiednio określić priorytety służące realizacji określonego przez siebie lub innych zadania (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 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 $1 + \lceil m/2 \rceil$ (zaokrąglenie w dół) punktów, gdzie m jest punktacją maksymalną (np. aby uzyskać ocenę pozytywną przy $m = 30$ należy zdobyć przynajmniej 16 punktów).

Treści programowe

Program wykładu obejmuje następujące zagadnienia: Wstęp: Idea mierzenia i skalowania, podstawowe transformacje zmiennych; dane i pomiary wielowymiarowe, zmienne ukryte, redukcja zmiennych (wymiarowości). Idea wizualizacji danych. Twierdzenie Vivianiego i barycentryczne układy współrzędnych, trójwymiarowe i czterowymiarowe. Zastosowanie czterowymiarowych układów barycentrycznych: wizualizacja miar konfirmacji i miar trafności klasyfikowania. 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). 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. Wartości i wektory osobliwe macierzy, rozkład względem wartości osobliwych (ang. "singular value decomposition", SVD). Idea, procedury i zastosowania metod: składowych głównych (ang. "principal component analysis", PCA) oraz skalowania wielowymiarowego (ang. "multidimensional scaling", MDS).

Miary i problem optymalizacyjne (minimalizacja wybranych miar) w redukcji wymiarowości.

Zagadnienie najbliższych sąsiadów, dywergencja Kullbacka-Leiblera. Idea, procedury i zastosowania metod: SNE i t-SNE (ang. /t-distributed/ stochastic neighbour embedding).

Tematyka zajęć

Program wykładu obejmuje następujące zagadnienia: Wstęp: Idea mierzenia i skalowania, podstawowe transformacje zmiennych; dane i pomiary wielowymiarowe, zmienne ukryte, redukcja zmiennych (wymiarowości). Idea wizualizacji danych. Twierdzenie Vivianiego i barycentryczne układy współrzędnych, trójwymiarowe i czterowymiarowe. Zastosowanie czterowymiarowych układów barycentrycznych: wizualizacja miar konfirmacji i miar trafności klasyfikowania. 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). 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. Wartości i wektory osobliwe macierzy, rozkład względem wartości osobliwych (ang. "singular value decomposition", SVD). Idea, procedury i zastosowania metod: składowych głównych (ang. "principal component analysis", PCA) oraz skalowania wielowymiarowego (ang. "multidimensional scaling", MDS).

Miary i problem optymalizacyjne (minimalizacja wybranych miar) w redukcji wymiarowości.

Zagadnienie najbliższych sąsiadów, dywergencja Kullbacka-Leiblera. Idea, procedury i zastosowania

metod: SNE i t-SNE (ang. /t-distributed/ stochastic neighbour embedding).

Metody dydaktyczne

Wykłady: prezentacja multimedialna uzupełniona przykładami podawanymi na tablicy, demonstracja wybranych systemów wizualizacji danych.

Laboratoria: modelowanie przykładowych problemów dotyczących skalowania i wizualizacji i rozwiązywanie ich metodami dostępnymi w laboratorium, wykonywanie eksperymentów symulacyjnych, dyskusja, praca w zespole, demonstracja i pokaz multimedialny.

Literatura

Podstawowa

1. J. Koronacki, J. Ćwik: Statystyczne systemy uczące się , WNT, Warszawa, 2005
2. G. Banaszak, W. Gajda: Elementy algebry liniowej część I i II, WNT, Warszawa, 2002

Uzupełniająca

1. Materiały wykładowe
2. I.T. Jolliffe: Principal Component Analysis , Springer-Verlag, Nowy Jork, USA, 2002
3. I. Borg, P.J.F. Groenen: Modern Multidimensional Scaling , Springer Science+Business Media, Nowy Jork, USA, 2005
4. M. Greenacre: Correspondence Analysis in Practice , Chapman & Hall/CRC Press, Nowy Jork, USA, 2007.

Bilans nakładu pracy przeciętnego studenta

	Godzin	ECTS
Łączny nakład pracy	75	3,00
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	30	1,50
Praca własna studenta (studia literaturowe, przygotowanie do zajęć laboratoryjnych/ćwiczeń, przygotowanie do kolokwii/egzaminu, wykonanie projektu)	45	1,50