

Krakowska Interdyscyplinarna Szkoła Doktorska

Opis przedmiotu/ course description

Przedmiot/ Course :	Metody uczenia maszynowego w eksploracji danych
Moduł kształcenia/ Training module:	Moduł specjalistyczny
Okres realizacji/ Implementation period :	II rok, semestr zimowy
Język wykładowy/ Language:	język polski/ Polish
Prowadzący/ Lecturer:	dr hab. Małgorzata Jakubowska, prof. AGH
Wymiar godzin przedmiotu/duration :	30 (10 godzin - wykład, 20 godzin – ćwiczenia obliczeniowe)
Forma prowadzenia zajęć/ Form of teaching :	wykład 10 godzin, ćwiczenia 20 godzin
Opis przedmiotu/ course content:	<p>Przedmiot obejmuje teoretyczne podstawy dziedziny i przedstawia wstępne informacje o działaniu algorytmów uczenia maszynowego oraz sposobach ich wykorzystania w rozwiązywaniu problemów, które umożliwiają eksplorację danych pozyskanych w eksperymentach w różnych dyscyplinach nauki. Obliczenia realizowane będą z wykorzystaniem przykładów o charakterze dydaktycznym, a także zbiorów danych pozyskanych z repozytoriów internetowych oraz od kursantów.</p> <ol style="list-style-type: none">1. Wprowadzenie do algorytmów uczenia maszynowego Problem wielowymiarowości. Ogólna charakterystyka algorytmów uczenia maszynowego – metody bez nadzoru i z nadzorem. Zastosowania uczenia maszynowego w nauce i technice. Oprogramowanie wspierające uczenie maszynowe.2. Gromadzenie, kontrola i wstępne przetwarzanie danych Macierz danych wejściowych. Kompletowanie danych. Techniki pracy z danymi brakującymi. Problem danych odbiegających. Właściwy dobór i skalowanie zmiennych. Transformacja zmiennych. Podstawy przetwarzania sygnałów. Problem optymalnego próbkowania sygnałów. Algorytmy filtracji cyfrowej. Korekta linii bazowej.3. Metody wizualizacji danych wielowymiarowych. Wykres radarowy. Wykres macierzowy.4. Analiza skupień (<i>Cluster Analysis, CA</i>) – modelowanie danych nieoznakowanych. Hierarchiczna analiza skupień. Optymalizacja hiperparametrów algorytmu. Dendrogram, tworzenie i interpretacja. Zasady optymalnego podziału na skupienia. Metoda k-średnich.5. Analiza głównych składowych (<i>Principal Components Analysis, PCA</i>) – nienadzorowana redukcja wymiarowości przestrzeni danych

	<p>Geneza problemu. Ogólna zasada metody PCA. Strategie wyznaczania liczby istotnych głównych składowych. Interpretacja głównych składowych i ładunków składowych.</p> <p>6. Techniki uczenia z nadzorem - algorytmy klasyfikacyjne Etapy procesu klasyfikacji. Kryteria oceny klasyfikatora – dokładność, czułość, precyzja modelu. ROC, AUC. Macierz pomyłek. Strategie walidacji modeli klasyfikacyjnych. Selekcja zmiennych. Klasyfikator liniowy. Drzewa decyzyjne (<i>Classification and Regression Trees, CART</i>). Lasy losowe (<i>Random Forests, RF</i>). Maszyna wektorów wspierających (<i>Support Vector Machine, SVM</i>).</p> <p>7. Wprowadzenie do regresyjnych algorytmów uczenia maszynowego – modelowanie zmiennych ciągłych Selekcja zmiennych. Definiowanie i walidacja modelu regresyjnego. Metody oceny modelu - błąd średniokwadratowy kalibracji, walidacji krzyżowej i predykcji. Charakterystyka i działanie algorytmów: regresji głównych składowych (<i>Principal Components Regression, PCR</i>) i częściowych najmniejszych kwadratów (<i>Partial Least Squares Regression, PLSR</i>). Optymalizacja kompleksowości modelu.</p> <p>8. Statystyczne planowanie eksperymentów Geneza i cel planowania doświadczeń (<i>Design of Experiments, DOE</i>). Ogólna zasad DOE. Definiowanie zmiennych planu. Przegląd typowych planów. Modele liniowe – plan czynnikowy kompletny i frakcyjny, dwu- i wielopoziomowy. Modele liniowo-kwadratowe – plany kompozycyjne. Plan frakcyjny Placketta – Burmana. Ogólne podejście optymalizacyjne.</p>
<p>Efekty uczenia się wg 8PRK zgodnie z Programem kształcenia KISD/ <i>learning outcomes at level 8 of the PRK according to the KISD Training Program:</i></p>	<p>EU1,EU2,EU3,EU8,EU13</p>
<p>Forma weryfikacji efektów uczenia się/ <i>Method of verification of learning outcomes:</i></p>	<p>Zaliczenie</p>
<p>Wymagania wobec uczestników/ <i>Requirements for participants:</i></p>	