

STATYSTYCZNE SYSTEMY UCZĄCE SIĘ WPROWADZENIE

Prof. dr hab. inż. Jacek Koronacki, dr inż. Paweł Teisseyre

Celem wykładu oraz towarzyszących mu zajęć laboratoryjnych jest przedstawienie szerokiego wachlarza statystycznych metod maszynowego uczenia się. Wykład zacznie się nietypowo, a mianowicie od zarysowania nieklasycznych metod statystycznych stosowanych do rozwiązywania problemów, w których obserwacje są wektorami o wielkim wymiarze. Omówione zostaną testy wielokrotne oraz metody wyboru modelu regresji liniowej. Do problematyki tej wrócimy także na ostatniej godzinie wykładu w kontekście zadania klasyfikacji.

Zasadnicza część wykładu dotyczyć będzie trzech obszarów: uczenia pod nadzorem, bez nadzoru i z częściowym nadzorem. W ramach pierwszego obszaru omówione zostaną zagadnienia klasyfikacji (analizy dyskryminacyjnej) oraz regresji, w tym: (uogólnione) podejście Fishera do analizy dyskryminacyjnej; liniowa maszyna wektorów podpierających (liniowa SVM); dyskryminacja logistyczna; klasyfikator bayesowski i metody nieparametrycznej estymacji gęstości w klasach oraz metody najbliższych sąsiadów; drzewa klasyfikacyjne; komitety klasyfikatorów; nieliniowe maszyny wektorów podpierających i problem regularyzacji; procedura Grama-Schmidta i dekompozycja QR w regresji liniowej; regresja grzbietowa; metoda LASSO i metoda LAR; regresja nieliniowa – m.in. drzewa regresyjne i metoda MARS;

W ramach uczenia bez nadzoru omówione zostaną zagadnienia: analizy struktury danych i redukcji wymiarowości, m.in. analiza składowych głównych; analizy skupień, w tym metoda k-średnich, hierarchiczna analiza skupień, analiza skupień w przestrzeni cech i metody spektralne (m.in. algorytm DaSpec), metody oparte na teorii grafów i teorii informacji. Spośród metod uczenia pod częściowym nadzorem omówione zostaną m.in. metody TSVM (transductive SVM) i LapSVM (Laplacian SVM).

Wykład będzie miał charakter wprowadzenia w ww. tematykę – nacisk położony będzie na aspekty metodologiczne i podanie intuicyjnych podstaw omawianych podejść. Nieco dokładniejsze, bardziej formalne ujęcie znajdzie uczestnik kursu w notatkach do wykładu. Bardzo ważnym uzupełnieniem wykładu będą komputerowe zajęcia laboratoryjne. Wykład zajęcia laboratoryjne będą prowadzone w języku angielskim.