

STATYSTYCZNE METODY ANALIZY DANYCH

Prof. dr hab. inż. Jacek Koronacki

Plan kursu:

1. Preliminaria: czym są statystyczne systemy uczące się. Liniowa analiza regresji: estymatory MNK, wstępna analiza rezyduów (obserwacje odstające, wpływowe, nieadekwatność modelu), problem zmiennych ukrytych, korelacja a regresja (współczynnik determinacji), wnioskowanie statystyczne o parametrach modelu, problemy niejednorodności wariancji oraz autokorelacji rezyduów, wnioskowanie na podstawie studentyzowanych rezyduów, odległość Cooka, problem współliniowości, rozkład QR; problem wyboru i oceny modelu; analiza wariancji i kowariancji; wzmianka o modelach mieszanych i uogólnionych modelach liniowych.
2. Liniowa analiza dyskryminacyjna: (uogólnione) podejście Fishera; dyskryminacja oparta na regresji liniowej oraz dyskryminacja logistyczna; liniowe maszyny wektorów podpierających. Klasyfikator bayesowski i metoda największej wiarygodności w analizie dyskryminacyjnej.
3. Zasady nieparametrycznej estymacji gęstości w klasach, metoda najbliższych sąsiadów i metody pokrewne; (nieliniowe) maszyny wektorów podpierających; metody oparte na prototypach; problem wyboru i oceny klasyfikatora.
4. Drzewa klasyfikacyjne; rodziny klasyfikatorów (bagging, boosting, lasy losowe Breimana i ich modyfikacje).
5. Analiza skupień: metoda k-średnich średnich i metody pokrewne (np. SOM); hierarchiczna analiza skupień. Redukcja wymiarowości: analiza składowych głównych (liniowa i nieliniowa), analiza czynnikowa i skalowanie wielowymiarowe; wzmianka o analizie składowych niezależnych.
6. Nieparametryczna analiza regresji. Problemy nieklasyczne: m.in. ścieżki rozwiązań dla metod regresji i klasyfikacji pod nadzorem; analiza skupień na podzbiorach atrybutów; problem uwzględnienia własności geometrycznych skupień; klasyfikacja pod częściowym nadzorem.

Literatura:

J. Koronacki i J. Ćwik, *Statystyczne systemy uczące się*, wyd. II, Exit 2008 (wyd. I, WNT 2005).

T. Hastie, R. Tibshirani i J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, wyd. II, Springer 2009 (wyd. I, Springer 2001).

B. Ripley, <http://www.stats.ox.ac.uk/~teo/statmethod/formulae.pdf>

J. Koronacki i J. Mielniczuk, *Statystyka dla studentów kierunków technicznych i przyrodniczych*, WNT 2001, 2004, 2006, 2009.