

Doktori kurzus tematikája, de MSc-sek is felvehetik

A többváltozós adatelemzés modern módszerei 1

A tantárgyfelelős személy és a tantárgy előadója:

Héberger Károly, Magyar Tudományos Akadémia Természettudományi Kutatóközpont

Ajánlott előtanulmányok:

Valószínűségszámítás / Statisztika; Analitikai Kémia; Műszeres Analitika

A tantárgy célkitűzése:

Azoknak a sokváltozós adatfeldolgozási és értékelési módszereknek megismerése és alkalmazása, amelyeket a kemometria gyakorlatban az utóbbi időben fejlesztettek ki és kezdtek széles körben alkalmazni.

Részletes tematika:

1. A kemometria tárgya, felosztása (alapfogalmak, szakkifejezések). A matematikai statisztika elemei (eloszlásfüggvények, várható érték, variancia, hipotézisvizsgálat, normalitás-vizsgálat stb.)
2. Robusztus és nem-paraméteres módszerek, alkalmazhatóságuk (letörési pont, lineáris eset, aszimmetrikus konfidencia intervallum, fuzzy halmazok és regresszió).
3. A fekete doboz modell és alkalmazhatósága. {empirikus, illetve statisztikai modellezés, a változók (faktorok) megválasztása. A korrelációs koefficiens (index) és a vele való visszaélés}.
4. Görbeillesztés (lineáris és nem-lineáris eset, modell megkülönböztetés, linearitás ellenőrzése, stb.) Az illesztések során elkövetett leggyakoribb hibák és kiküszöbölésük.
5. Reziduumvizsgálat, példák, az illeszkedés hiánya, a modell megválasztása, trend észrevése. A hibalefutás (szerkezet) szerepe, elemzése. Súlyozás, heteroszkedaszticitás, változó transzformáció, új változó modellbe építése.
6. Többváltozós (multivariate) módszerek 1. A többszörös lineáris regresszió (F próba t próba). LFER, QSAR, példák.
7. Változók (faktorok) kiválasztása a mérési pontok leírása szempontjából egyenértékű változók közül. A Páronkénti korrelációs módszer és általánosítása. A kereszt-ellenőrzés.
8. Többváltozós (multivariate) módszerek 2. Felügyelt tanítású mintázat felismerés, osztályozás: Lineáris diszkriminancia analízis, kanonikus korreláció analízis.
9. Többváltozós (multivariate) módszerek 3. Felügyeletlen tanítású mintázat felismerés. Csoportelemzés, főkomponens-elemzés, faktoranalízis. Mikor használjuk?
10. Többváltozós (multivariate) módszerek 4. Főkomponens-regresszió (PCR). A parciális legkisebb négyzetek módszere (PLS), előrejelzés, a kemometria módszerek validálása.
11. Globális minimumkereső eljárások: szimulált megeresztés, genetikus algoritmus.
12. Módszerek és modellek összehasonlítása, rangsorolásuk és csoportosításuk, a keresztellenőrzés változatai.
13. A kísérlettervezés alapelemei. Hallgatói problémák elemzése. Programbemutató (Statistica, Unscrambler, stb.), konzultáció.
14. Összefoglalás. számonkérés.

A tantárgy oktatásának módja:

Előadás, programbemutatókkal

Az aláírás és a jegy megszerzésének feltétele: kiselőadás tartása, írásbeli vizsga

Jegyzet, tankönyv, felhasználható irodalom:

T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Data Mining Inference and Prediction, Springer, New York, 2nd edition, 2009 February. Interneten elérhető:

<https://web.stanford.edu/~hastie/Papers/ESLII.pdf>

Szerzői kollektíva: Sokváltozós adatelemzés (kemometria). Nemzeti Tankönyvkiadó, Bp, 2001.