

Adatbányászati algoritmusok
2011/12 őszi félév
Vizsgatematika

1. Adathalmazokkal kapcsolatos alapfogalmak, változók típusai. A tudásfeltárás folyamata. Adatok előfeldolgozása: normalizáció, transzformáció (beleértve változók diszkrétizálását, Wavelet- és Discrete Fourier transzformációkat).
2. A nagy dimenziószám átka, dimenziószám-csökkentő eljárások: PCA, MDS, ISOMAP. A csomósodás (hubness) jelensége, és ennek összefüggése az adatok dimenziószámával.
3. Döntési fák. Döntési fák tanítási algoritmusai.
4. Legközelebbi szomszéd alapú osztályozó modellek, idősorok osztályozása, a DTW algoritmus.
5. Bayes-tétel. Bayesi osztályozók: Naive Bayes és Bayes-hálózatok. Az „optimális” osztályozó Bayes-i hibája.
6. A perceptron algoritmus. Többrétegű neurális hálók, a Back-propagation algoritmus.
7. Szupport vektor gépek.
8. Ensemble modellek. Ensemble modellek helyességének Dietterich-féle indokai, Bagging és Boosting.
9. Osztályozó modellek kiértékelése, az overfitting jelensége.
10. Asszociációs szabályokkal kapcsolat alapfogalmak (gyakori elemhalmazok, támogatottság, konfidencia, stb.). A szupport függvény antimonotonitása. Asszociációs szabályok konfidenciájának antimonotonitása. Az apriori algoritmus, Hash-fás implementációja és Christian Borgelt implementációja.
11. Az ECLAT algoritmus. Asszociációs szabályok jóságának kiértékelésére használt mutatók és azok jellemzői. A szupport függvény ferde (skewed) eloszlása, Liu algoritmus. Simpson paradoxona.
12. A klaszterezés feladata. Kleinberg tétele. A k -közép klaszterező algoritmus és változatai.
13. Klaszterezések és klaszterek típusai. Hierarchikus klaszterező algoritmusok, a DB-Scan klaszterező algoritmus.
14. Anomália-felismerési eljárások.