

Adatbányászati algoritmusok

2006/2007. tanév I. félév

VIZSGAFELADATOK – 2007. JANUÁR 10.

- (1) Adja meg a tudáskinyerés lépéseit (**4 pont**). Ismertesse a legfontosabb adatbányászati feladatokat. (**4 pont**)?
- (2) Írja le az FP-growth algoritmust (**7 pont**). Hogyan támogatja az FP-growth algoritmust az FP-fa (**3 pont**)?
- (3) Ismertesse az asszociációs szabályok kinyerésének feladatát (**3 pont**). Vázzolja a χ^2 -alapú függetlenségvizsgálatot (**6 pont**).
- (4) Írja le a neurális hálózat alapú osztályozót (**8 pont**). Milyen előnyökkel/hátrányokkal rendelkezik a neurális hálózat alapú osztályozó (**2 pont**)?
- (5) Tegyük fel, hogy a bemeneti adatbázis minden i sora négy bináris értéket tartalmaz (x_i , y_i , w_i és z_i). Minden i -re igaz, hogy $z_i = x_i + y_i \bmod 2$, továbbá a sorok 80%-ában $w_i = z_i$, a maradék 20%-ban $w_i = 1 - z_i$. Mit fogunk kezdeni egy ilyen bemenettel a döntésifa-alapú osztályozók, amennyiben z a magyarázandó attribútum? Válaszát indokolja (**4 pont**). A döntési fák milyen hibájára mutat rá ez az egyszerű példa (**3 pont**)?
- (6) A sztochasztikus szörfölő modellje alapján felállított rangsoroló módszer mely hibáira ad megoldást a PageRank algoritmus (**4 pont**). Ismertesse ezt a módszert (**6 pont**).
- (7) Milyen elvárásaink vannak klaszterező algoritmusokkal szemben (**4 pont**)? Ismertesse a k-közép klaszterező eljárást (**6 pont**). Milyen előnyei és hátrányai vannak az eljárásnak (**4 pont**)?
- (8) Fogadás a lóversenyen. Mi a vagyonnövekedési ráta, mi az optimális portfólió és az információ értéke (definíciók, tételek, bizonyítások) (**20 pont**).