

Adatbányászati algoritmusok

2005/2006. tanév I. félév

VIZSGA FELADATOK – 2005. DECEMBER 20.

- (1) Milyen mérőszámok állnak rendelkezésünkre a hasonlóság definiálására
 - bináris (**2 pont**), illetve valós (**3 pont**) attribútumokkal leírt objektumok,
 - elemsorozatok (**1 pont**)
 - és dokumentumok (**1 pont**)esetén.
- (2) Adja meg a gyakori elemhalmazok kinyerésének feladatát (**2 pont**). Hogyan definiáljuk a zárt mintákat (**2 pont**)? Írja le a gyakori elemhalmazokat kinyerő Eclat algoritmust (**6 pont**).
- (3) Vázzon a szinguláris felbontás (SVD) alapján történő dimenziócsökkentési eljárást (**7 pont**). Hogyan szokás definiálni a dimenziócsökkentés hibáját (**1 pont**). Mit tudunk erről mondani az SVD esetében (**2 pont**)?
- (4) Hogyan definiáljuk egy asszociációs szabály függetlenségei mutatóját (lift érték) (**2 pont**)? Miért nem elég, ha csak támogatottság, bizonyosság és érdekességi mutató alapján szűrjük az asszociációs szabályokat (**3 pont**)?
- (5) Mutassa be a döntési fák előállításánál használt, divergencia-függvény alapján definiált vágási függvények osztályát (**4 pont**). Mely ismert vágási függvények tartoznak ebbe a családba (**2 pont**). Milyen tulajdonságok jellemzik ezt a családot (**4 pont**)?
- (6) Milyen elvárásaink vannak egy jó klaszterező algoritmussal szemben (**5 pont**)? Írja le a single-link működését. Milyen előnyei, hibái vannak az algoritmusnak (**6 pont**)?
- (7) A sztochasztikus szörfölő modellje alapján felállított rangsoroló módszer mely hibáira ad megoldást a PageRank algoritmus (**4 pont**). Ismertesse ezt a módszert (**5 pont**).