

Adatbányászati algoritmusok

2007/2008. tanév I. félév

ZÁRTHELYI VIZSGA – 2007. NOVEMBER 13.

- (1) Milyen feltételei vannak egy sikeres adatbányászati projektnek az adatokra és az alkalmazási területre vonatkozóan (**5 pont**)?
- (2) Hogyan mérjük a mintavételezés hibáját, ha mintavételezés során a cél a relatív gyakoriságok megőrzése (**3 pont**)? Mit tudunk mondani a hiba, az elem előfordulásának valószínűsége és a minta mérete közötti kapcsolatra (**5 pont**)?
- (3) Hogyan lehet alkalmazni az equisupport nyesést az Eclat algoritmusban (**7 pont**)?
- (4) Definiálja az empirikus kovarianciát és az empirikus korrelációt (**2 pont**). Fejezze ki az $X \rightarrow Y$ asszociációs szabály ezen értékeit az X és Y relatív gyakoriságaival (az eredményt bizonyítsa) (**3 pont**). Legyen a nullhipotézisünk az, hogy az X és Y függetlenek egymástól. Meg tudunk-e adni értelmes elfogási tartományt, amennyiben csak az empirikus korrelációt és a tranzakciók számát ismerjük (**4 pont**)?
- (5) Adja meg a klaszterezés klasszikus jósági mutatóit (**4 pont**). Milyen hátrányai vannak azoknak az algoritmusoknak, amelyek ezen jósági mutatókat használják (**3 pont**)?