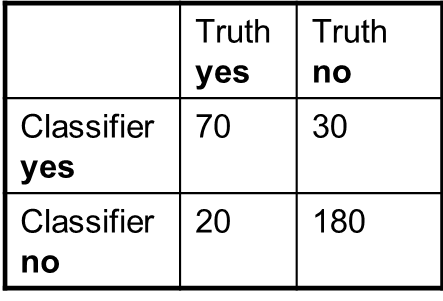
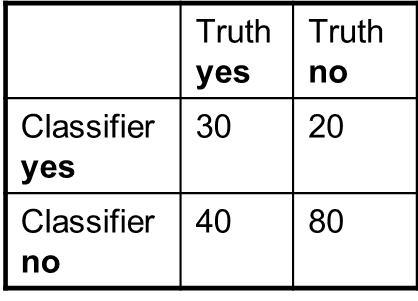
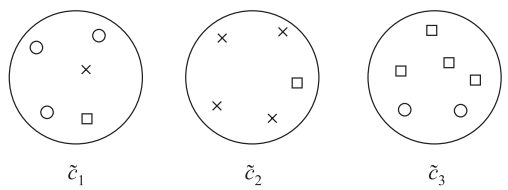
Lehetséges típusfeladatok ZH-ra

1. Mi az adatbányászat definíciója?
2. Ismertesse a CRISP-DM módszertan lényegét!
3. Ismertesse a 4 legfontosabb adatbányászati feladattípust!
4. Mi a különbség és hasonlóság az osztályozás és a regresszió között? Hibamérték, célfüggvény is.
5. Mi a különbség és hasonlóság az osztályozás és a csoportosítás között?
6. Ismertesse a 4 alapvető adattípust, és milyen módon lehet őket összehasonlítani?
7. Hogyan hasonlítana össze olyan objektumokat, amelyek attribútumai többféle adattípusból kerülnek ki?
8. Definiálja a távolság metrikát!
9. Mit jelent az adatsimítás, mikor alkalmazzuk, és adjon egy példa módszert!
10. Mi a diszkretizálás ill. a normalizálás, és mikor alkalmazzuk?
11. Mire utal a felügyelt és a felügyelet nélküli tanítás terminológia? Adjon példát mindkét esetre!
12. Ismertesse az információ-visszakeresés (IR) osztályzásnál használt alapvető mértékeit!
13. Milyen kiértékelő mértéket használna osztályozásnál akkor, ha egy objektum több kategóriába is tartozhat (de akár 0-ba is)? Indokolja!
14. Számolja ki a mikro és makroátlagolt pontosságot az alábbi esetekben. Mi okozza a különbséget?



1. Ismertesse a lineáris osztályozó fogalmát és adjon rá egy példát!
2. Mi a különbség a többcímkés és a kiválasztó osztályozás között (példával)?
3. Ismertesse a Rocchio-osztályozót!
4. Ismertesse a legközelebbi szomszédok módszert! Hogyan használjuk a KD-fákat a gyorsításra?
5. Ismertesse a naív Bayes osztályozó alapötletét; miért a naív jelző?
6. Hogyan lehet az alulcsordulást megelőzni a naív Bayes-nél?
7. Mi a naív Bayes bonyolultsága?
8. Mi a különbség a binomiális és a multinomiális modell paraméterbecslése között (NB)?
9. Mikor érdemes használni a naív Bayes-t és mikor a döntési fákat?
10. Ismertesse a döntési fákkal történő osztályozást kitérve a vágási függvényekre és nyesési technikákra!
11. Ismertesse az ID3 döntési fa algoritmust!
12. Jellemezze a jó vágási függvényeket (Taylor-Silverman)!
13. Ismertesse az SVM vázlatát!
14. Mi a különbség a lineárisan szeparálható, a lineárisan nem szeparálható (gyengítő változós eset) és a magtértranszformációt használó SVM között?
15. Milyen csoportok közti távolságfüggvényeket ismer?
16. Ismertesse a hierarchikus csoportosítók működését és hasonlítsa össze a 4 csoporthasonlósági függvényt!
17. Ismertesse az inverzió fogalmát csoportosításnál!
18. Mire használjuk a prioritási tömböket csoportosításnál?
19. Ismertesse a k-átlag módszert!
20. Mikor érdemes k-medoid módszert használni k-átlag helyett?
21. Határozza meg a tisztaság és az FP, FN, TP, TN értékeket az alábbi csoportok esetén:



1. Hogyan határozzuk meg a klaszterek címkéit?
2. Számolja ki az egyszerű kapcsolódással az alábbi elemek csoportosítását! Ábrázolja az eredményt dendogramon!

:

1. Ismertesse az asszociációs szabályok kinyerésének alapjait!
2. Ismertesse az apriori algoritmus működését!
3. Hogyan használjuk az apriori-elvet (antimonotonság) az apriori algoritmusokban?
4. Miért van szükség az elemek lexikografikus rendezésére az apriori algoritmusnál?
5. Mutassa be az apriori algoritmus működését az alábbi példán! Legyen a támogatottsági küszöb 2!

|  |  |
| --- | --- |
| tranzakció | elemek |
| 10 | a b c |
| 20 | b c d e |
| 30 | b c e |
| 40 | a b |
| 50 | d |
| 60 | d e c a |

1. Hasonlítsa össze az apriori és aprioriTID algoritmusokat!
2. Ismertesse a Direkt hash és pruning módszert!
3. Ismertesse az FP-growth algoritmust!
4. Ismertesse az ECLAT algoritmust!
5. Milyen alternatívák vannak az elemhalmazok reprezentálására?
6. Miért vezetünk be érdekességi feltételeket az asszociációs szabályoknál?
7. Ismertesse a TD-IDF súlyozást és indokolja a bevezetését!
8. Mi a különbség a szótövezés és a lemmatizálás között?