

Intelligens orvosi műszerek (VIMIA023)

Második zh

2018. november 27.

(60 perc)

Minden válaszhoz rövid, tömör indoklást is kérek, kivéve az igaz/hamis feleletválasztós feladatokat! Fontos javaslat az elmúlt évek tapasztalatai alapján: nem érdemes kapkodni, alaposan olvassa el a feladatot, gondolkodjon el rajta, értelmezze, mielőtt elkezdi megoldani! (Pl. egy „nem” vagy „mindig” szó a feladat szövegében sokat változtathat a helyes válaszon.)

NÉV (nyomtatott betűvel): NEPTUN-KÓD:

ALÁÍRÁS:

1. A következő állítások közül melyik hamis, melyik igaz?
- a. A szenzitivitás és specificitás összege mindig 1. a. **Igaz Hamis**
 - b. Az FN és a TP esetek összege a betegek összes számát adja meg. b. **Igaz Hamis**
 - c. A döntés várható összköltségét nem befolyásolja a valós pozitív (TP) esetek költsége, hiszen őket mindenképpen kezelniük kell. c. **Igaz Hamis**
 - d. Ha mindenkit betegnek (pozitívnak) veszünk, akkor a ROC görbe (0,0) pontját kapjuk. d. **Igaz Hamis**
 - e. A döntési eljárás költségkomponensei közül C_{00} minden értelmes valós esetben nagyobb, mint C_{10} . e. **Igaz Hamis**
 - f. Ha egy csoportban a hamis pozitív diagnózisok (TN helyett FP) nagyobb költségnövekedéssel járnak, mint a hamis negatív diagnózisok (TP helyett FN), akkor ezt a csoportot betegnek érdemes tekinteni a diagnosztikai eljárás során. f. **Igaz Hamis**
 - g. A ROC görbe analízisnél a véletlen találgatás kb. 0,5 görbe alatti területet eredményez. g. **Igaz Hamis**
 - h. Ha a mintahalmazban ugyanannyi pozitív, mint amennyi negatív eset van, akkor a jó döntéshez szükséges információszükséglet 0. h. **Igaz Hamis**
 - i. A tanulásnál korai leállásnak nevezzük, ha akkor állunk le, amikor a tesztminta-halmazon nőni kezd a hiba, miközben a tanítóminta-halmazon még csökken. i. **Igaz Hamis**
 - j. Ha a döntési küszöböt változtatjuk, akkor a TP esetek száma és a FP esetek száma általában ugyanabban az irányban változik (mindkettő nő vagy mindkettő csökken). j. **Igaz Hamis**
 - k. Az egyszerű eszközök hajlamosabbak a túltanulásra. k. **Igaz Hamis**
 - l. A hibaarány-komplexitás alapú metszésnél általában korai leállást alkalmazunk. l. **Igaz Hamis**

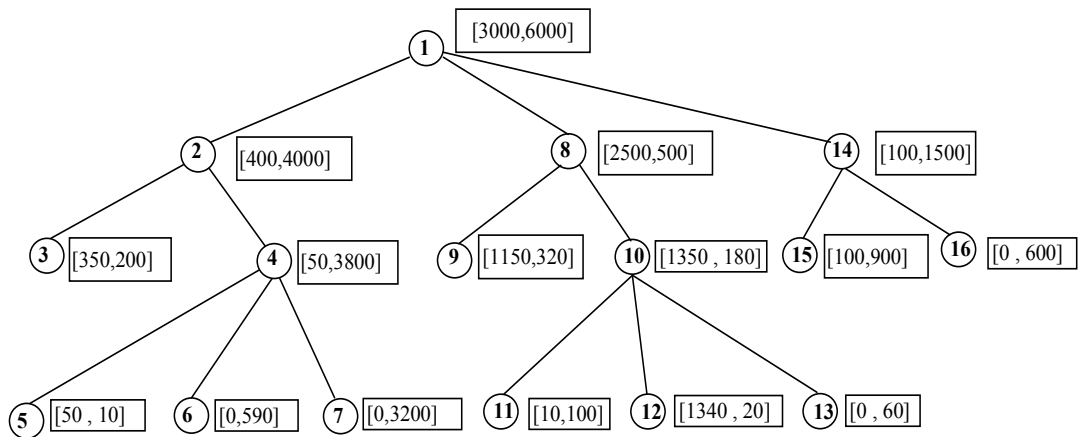
(≤ 6 jó válasz **0 pont**,

$6 <$ jó válasz: **(jó válaszok száma-6) pont**,

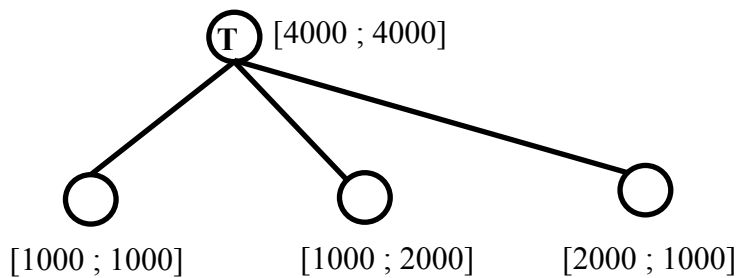
12 jó válasz **6 pont**)

A túloldalon is vannak még feladatok!

2. Szűrési eljárást fejlesztünk egy 8.700.000 népességű társadalomra. Az derült ki, hogy négyszer akkora egy főre eső költségnövekedést jelent, ha tévesen egészségesnek veszünk egy embert (FN), mint ha tévesen betegnek vennénk (FP). Szűrési eljárásunk egy mérések alapján számított z paraméteren alapul, és a $z=130$ értéket produkáló csoportba a teljes népességből várhatóan 810.000 egészséges és 17.000 beteg kerül. Mi az olcsóbb, ha a $z=130$ -at produkáló embereket betegnek vagy egészségesnek diagnosztizáljuk? **(4 pont)**
3. Az alábbi ábrán egy 9.000 minta alapján tanítással előállított döntési fa látható. A csomópontok mellett, mindig tőlük jobbra bekeretezve szögletes zárójelben látható két szám, ez a C1, illetve C2 osztályba jutó tanítóminták száma az adott csomópontban. Mind az osztályozás hibaaránya, mind az eszköz komplexitása költséget jelent nekünk: az egységnyi komplexitásnövekedés 250 bitcoin költségnövekedést okoz, míg az egységnyi hibaarány-növekedés 4000 bitcoin költségnövekedést. Olcsóbb-e a kifejtett döntési fát használnunk, mint ha eldobjuk az egész döntési fát, és a gyökércsomópontot (1-es) használjuk levélként? (Tehát az eredeti tanító halmaz alapján hozni a döntést, egyetlen csomópontkifejtést se végezve.) **(5 pont)**



4. Az alábbi teszt során mekkora lesz az információnyereség? A minták az A, illetve a B osztályba tartoznak, eredetileg 4000 - 4000 tanítóminta volt a két osztályból. A gyökércsomópont mellett (jobbra), illetve a teszt után létrejövő gyermekcsomópontok alatt található két-két szám azt mutatja, hogy abba a csomópontba hány A, illetve B osztálybeli tanítóminta jutott el [#A,#B] sorrendben. Természetesen a választát számítással indokolja! **(5 pont)**



$$K(Tn) = R(Tn) + \alpha \cdot |T(Tn)|$$

$$I(P(v_1), P(v_2), \dots, P(v_k)) = \sum_{j=1}^k -P(v_j) \cdot \log_2(P(v_j))$$

$$\log_2(x) = \frac{\log_{10}(x)}{\log_{10}(2)}$$

$$TPR = \frac{TP}{TP + FN}$$

$$TNR = \frac{TN}{TN + FP}$$

Jó munkát!