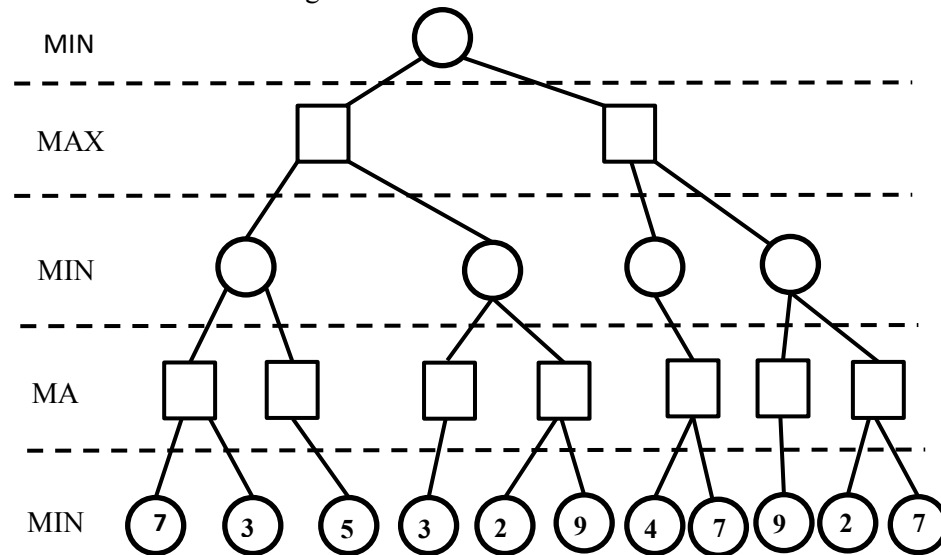


MI negyedik előadásához kapcsolódó kérdések, feladatok (Keresés ellenséges (kompetitív) környezetben)

1. Alkalmazza a MINIMAX algoritmust az ábrán látható játékfára, számolja ki a gyökér minimax értékét, és jelölje be a MINIMAX lépéseket! A gyökér a MIN játékos. Röviden mutassa be az alfa-béta nyelési eljárást! Feltéve, hogy a csomópontokat balról-jobbra értékeli ki, milyen alfa-béta vágásra van a fában lehetőség?



2. IGAZ/HAMIS?

- A csomópontok kiértékelésének sorrendje nem hat az alfa-béta nyelés hatékonyságára.
- Ha a lehető legjobb sorrendben értékeljük ki a csomópontokat, akkor adott erőforrások mellett kb. kétszeresre nő az elérhető mélység.
- A sakkjátékot teljes információszerzést lehetővé tevő determinisztikus környezetet ad.
- A bridzset nem teljes információt biztosító véletlen környezetben játszunk.
- Szimulált lehűtés esetén a hőmérséklet csökkenésével egyre kisebb a nem optimális irányba történő lépés valószínűsége.
- Hegymászó keresésnél a lokális szélsőértékek okozta problémák csökkentése érdekében véletlen újraindítású keresést végezhetünk.
- Hegymászó keresés esetén a számon tartott keresési gráf mérete a megoldás mélységének négyzetével nő.

MI ötödik előadásához kapcsolódó kérdések, feladatok

1. IGAZ/HAMIS?
 - a. Az irreleváns attribútum tesztje esetén a ROC görbe alatti terület 0,5.
 - b. Amikor a szenzitivitás (TPR) nő, a specificitás (TNR) együtt nő vele.
 - c. Ha minden esetet pozitívnak feltételezünk, akkor a hamis negatív esetek nem okoznak költségnövekedést.
 - d. Ha a hamis negatív esetek nagyobb költségnövekedést okoznak egy mintahalmazra, mint a hamis pozitív esetek, akkor ezt a csoportot pozitívnak célszerű tulajdonítani.
 - e. A C_{01} költség általában nagyobb, mint a C_{11} költség.

2. Előzetes vizsgálatot végeztünk egy diagnosztikus eljárás érzékenységére (TPR) és specificitására (TNR), a megvizsgált 71.218 esetből 33.105 volt hibátlan, a többi hibás. Eljárásunkkal összesen 37.321 esetet észleltünk hibásnak, de közülük csak 33.980 volt ténylegesen az. Ezen előzetes vizsgálat eredménye alapján számolva várhatóan mi lesz a szűrési eljárásunk érzékenysége és specificitása?

3. **A.** Eljárásunkat 48.384 fős mintán teszteltük, akik közül 24.192 egészséges, és ugyanannyi a beteg. Egy z paramétert mértünk, és különböző L küszöbértékeknél azt vizsgáltuk, hogy hány egészségesnek és hány betegnek esik L alá a mért z értéke. Ezek után azokat egészségesnek minősítettük, akiknek L alatti a z paramétere, és azokat betegnek, akiknél $z \geq L$. Rajzolja fel az eljárást jellemző ROC görbét, ha különböző L értékeknél az alábbi táblázatban található valós pozitív (TP) és valós negatív (TN) esetszámokat tapasztaltuk! Jelölje meg a görbén (legalább jellegre helyesen, tehát a 0,3 pont ne a 0,9-nél legyen!) az egyes kiszámolt pontokat, és azt, hogy melyik L értéknél kaptuk azokat!

L – küszöbszint	-100	-75	-50	-25	0	25	50	75
Az adott L küszöbszinthez tartozó TN esetszám	0	3.457	6.914	10.369	13.826	17.282	20.737	24.192
Az adott L küszöbszinthez tartozó TP esetszám	24.192	20.735	17.278	13.823	10.366	6.910	3.455	0

B. Tudjuk, hogy a teljes népességben a betegek aránya 0,2%. (Nem úgy, mint a tesztmintánkban!) A költségeink $C_{00}=3000$, $C_{10}= 8500$, $C_{11}=53000$, $C_{01}=179000$. Ennek alapján az $L_1=-25$ vagy az $L_2=0$ küszöbértékhez tartozik kisebb költség? (Ötlet: -25 és 0 közötti z értékkel 3457 egészséges és 3457 beteg rendelkezik.)

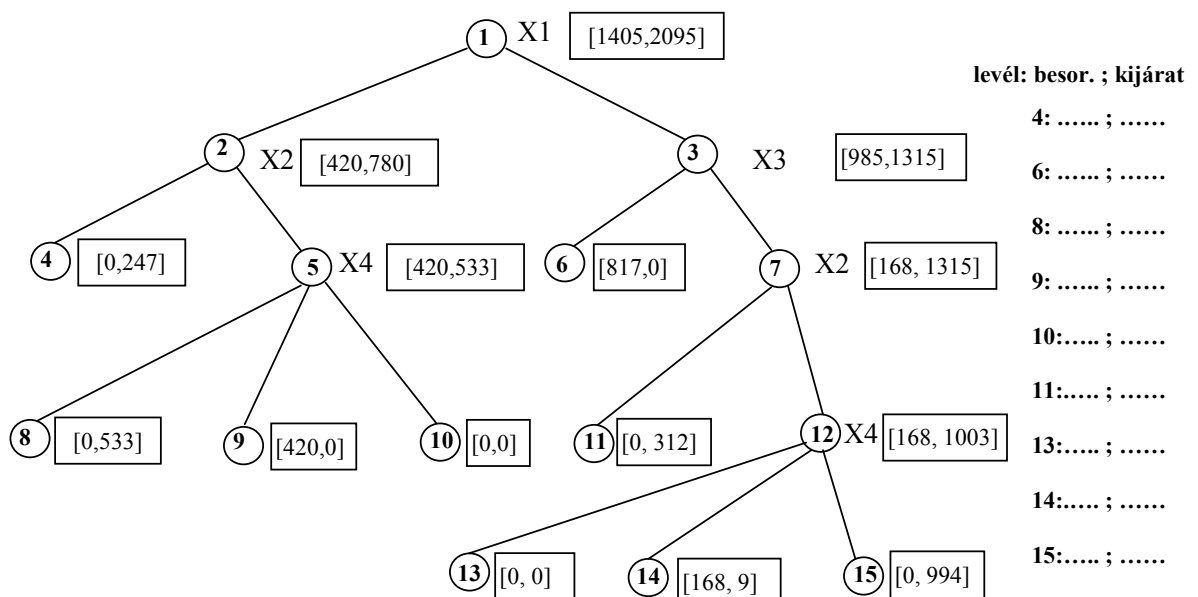
MI hatodik előadásához kapcsolódó kérdések, feladatok (döntési fák stb.)

1. A döntési fa tanulás pszeudókódja:

```

function DÖNTÉSI-FA-TANULÁS(példák,attribútumok,alapérték) returns egy döntési fa
inputs példák, a példák halmaza
           attribútumok, az attribútumok halmaza
           alapérték, a célpredikátum alapértéke
if példák üres
then
    return alapérték                                     %=> A
elseif példák minden elemének azonos a besorolása
    return közös besorolás                             %=> B
elseif attribútumok üres halmaz
    return TÖBBSÉGI-ÉRTÉK(példák)                    %=> C
else
    legjobb ← Attribútum-Választás(attribútumok, példák)
    fa ← egy új döntési fa, a gyökér a legjobb attribútum tesztje
    m ← TÖBBSÉGI-ÉRTÉK(példák)
    for each legjobb tesztjének minden  $v_i$  lehetséges értékére do
        példáki ← {a példák azon elemei, amelyre a legjobb attribútum értéke =  $v_i$ }
        részfa ← DÖNTÉSI-FA-TANULÁS(példáki,attribútumok-legjobb,m)
        a fa döntési fához adjunk egy  $v_i$  címkejű ágat és a részfa részfat
    endfor
    return fa                                         %=> D
endif
  
```

A következő ábrán egy 3500 minta alapján tanítással előállított döntési fa látható. A mintákat 4 attribútum jellemzi X1, X2, X3 és X4; az első három bináris, a negyedik 3 értékű diszkrét. A csomópontok mellett jobbra található annak az attribútumnak a neve, amelyet az adott csomópontban tesztelünk, és bekeretezve, szögletes zárójelben látható két szám, ez az adott csomóponthoz érkező, C1, illetve C2 osztályba tartozó tanítóminták száma.

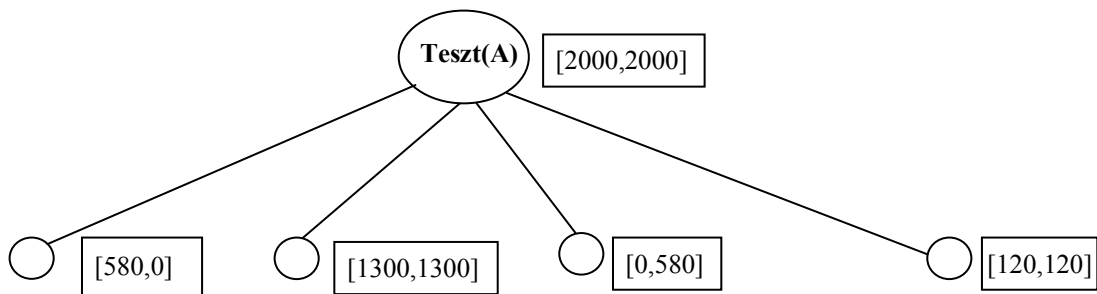


Adja meg, hogy melyik levéltől való visszatéréskor a pszeudókód melyik „kijáratánál” (A, B, C vagy

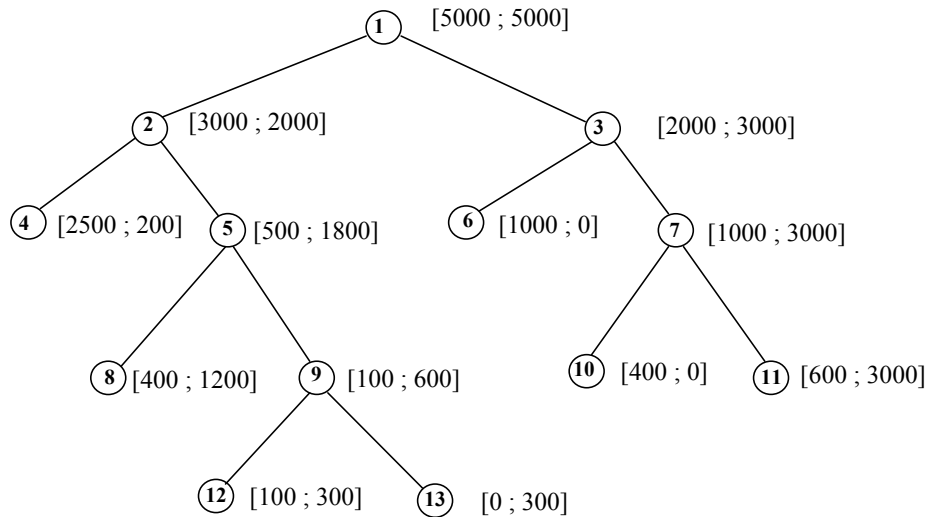
D) léptünk ki, és melyik levélhez milyen besorolás tartozik! (9 levél van: az ábra mellett jobbra felsorolva.)

2. Egy ismert besorolású elemekből álló mintahalmaz alapján alakítunk ki kétosztályos (bináris) döntést megvalósító döntési fát. Milyen okai lehetnek annak, ha a bemeneti attribútumainkat már mind teszteltük az adott részfában, de az egyik levélcsomóponthoz érkező tanítóminták mégse mind ugyanabba az osztályba tartoznak? Milyen választ rendelünk ilyenkor ehhez a levélhez?
3. Az alábbi ábrán egy 4000 minta alapján tanítással előállított döntési fa egyik csomópontja, és a gyermekcsomópontok 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.

A vizsgált csomópontban az A attribútumot teszteltük, a tesztnek 4 lehetséges értéke van. Az ábrán látható módon oszlottak meg a minták a 4 lehetséges érték mentén. Mekkora lesz a bemutatott hatású teszt információnyeresége?



4. IGAZ/HAMIS?
 - a. A komplexitás nem csak az erőforrásigény miatt rossz nekünk, de a nagyobb komplexitás a túltanulásra is hajlamosabbá teheti eszközünket.
 - b. Ha egy szabályos teljes franciakártya pakliból kihúzott lapnál meg akarjuk mondani, hogy milyen színű lesz, akkor az információszükségletünk 2 (bit).
 - c. Ha egy sakk-készletből, amelyikből hiányzik a sötét király, kiemelünk egy bábut, akkor a bábu színének megjósolásához szükséges információ kevesebb mint 1 (bit).
 - d. Egy tökéletes attribútum tesztje után az összes minta ugyanarra az ágra kerül a döntési fában.
 - e. Ha egy attribútum tesztje után az összes gyermekcsomópontnál ugyanaz a két osztály mintáinak aránya, mint a szülőcsomópontnál volt, akkor az a teszt nem hozott információnyereséget.
 - f. Ha a döntési fa egyik csomópontjához csak az egyik osztály mintái jutnak, akkor ott már nem érdemes tovább tesztelni az attribútumokat.
 - g. Ha már az összes rendelkezésre álló attribútumot leteszteltük a fában, akkor a csomópontok homogének kell legyenek.
4. A következő – bináris döntést végző – döntési fát tanítottuk egy mintahalmaz alapján. A minták az A, illetve a B osztályba tartoznak, eredetileg 5000-5000 tanítóminta volt mindkét osztályból. Az egyes csomópontok mellett (jobbra) található két szám vagy azt mutatja, hogy abba a csomópontba hány A, illetve B osztálybeli tanítóminta jutott el, vagy az adott csomópontban mért hibaarányt. A javaslat az, hogy vagy a 2-es vagy a 3-as csomópontot zárjuk be (az innen kiinduló részfát vágjuk le). Elvégezve a hibaarány-komplexitás metszési eljárást vizsgáljuk a kettő közül melyik csomópontból kiinduló részfa metszése a jobb? (Természetesen választát számítással támassza alá!)



5. Az alábbi ábrán egy 12.000 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 adott feladatban az egységnyi hibaarány költsége 0,1 bitcoin. Mekkora az egységnyi komplexitás költsége, ha a 8-as csomópont esetén mindegy, hogy bezárjuk (és levélként kezeljük), vagy megtartjuk a belőle kialakított részfat?

