

**Mindegyik feladat 1-1 pontot ér, a vizsgán a feltett 15 kérdésből legalább 9 jó választ kell elérni. Minden válaszhoz rövid, de áttekinthető indoklást is kérek! Mivel ezek „beugró” típusú kérdések, persze elég 1-2 mondat! Természetesen ezek csak példák, nem kizárólag ezek a kérdések szerepelhetnek!**

1. Két döntési fát készítünk különböző mintahalmazok felhasználásával ugyanazon probléma megoldására. Azt tapasztaljuk, hogy a két fának különböző számú levél csomópontja van. Lehet-e mindkét fa mindkét tanítóhalmazzal konzisztens?
2. Eszközünk tetszőleges 7 bemenetű (egy kimenetű) logikai függvényt képes megvalósítani. A konkrét problémához keresett függvényt tanítással igyekszünk előállítani. Mekkora a hipotézistér mérete?
3. Turbózással igyekszünk javítani osztályozónk teljesítményén, ADABOOST eljárást alkalmazunk. Az első osztályozó által elért pontosság 63%. Mekkora lesz a második osztályozó tanításánál az 5000 tanítómintánk súlyainak összege?
4. Milyen két fő csoportba osztjuk a klaszterezési eljárásokat?
5. Mit jelent az, hogy a gráf-alapú féligellenőrzött tanítási eljárások transzduktívak?
6. A transzduktív SVM eljárás adhatja-e ugyanazt a megoldást egy konkrét mintahalmaz esetén, mint a klasszikus SVM?
7. Egy osztályozási feladatnál a  $P(y^{(\text{tanító})} \neq d^{(\text{tanító})})$  valószínűség mit jelent? (Az eszköz pontossága, általánosító képessége, hibája stb.)
8. Döntési fát építünk kétosztályos feladatra, és a gyökér csomópontban 5000-5000 mintánk van a két osztályból. Kétféle attribútum-tesztet vizsgálunk meg a gyökér csomópontban, mindkettőnek kétféle kimenetele van. Az első teszt esetén az első gyermek csomópontba 4000 minta jut az első osztályból és 1000 a másodikból. A második attribútum-teszt esetén a második gyermek csomópontba 500 minta jut az első osztályból és 4500 a másodikból. Melyik attribútum-teszt irrelevanciájának lesz nagyobb a valószínűsége a  $\chi^2$  teszt alapján?
9. Kétosztályos féligellenőrzött osztályozási feladatnál gráf-alapú (mincut) eljárást alkalmazunk. 10 ismert besorolású tanítómintánk van. Ha a 100 ismeretlen besorolású mintát is figyelembe vesszük, a gráf fölött értelmezett veszteségfüggvény hány értéke közül fogjuk a minimálist kiválasztani?
10. Hogyan lehet a MINCUT eljárást úgy módosítani, hogy az ismeretlen besorolású pontokra adjon a javasolt címkére vonatkozó valószínűségi információt is?
11. Féligellenőrzött regressziónál az ismeretlen válaszü pontok
12. Egy kétosztályos problémára tanított MOE struktúrájánál vita alakult ki, mert a tesztelésnél észrevették, hogy a két megtanított szakértő közül az egyik kimenete mindig 0, másiké mindig 1. A szakértőket megvalósító fjelsztő azzal védekezett, hogy ez nem lehetséges, hiszen eredőben a nemlineáris problémán 98%-os pontosságot ért el a MOE, tehát nyilván a teszt rossz. Igaza van?
13. Generatív klaszterezésnél az egyes mintaosztályokat generáló modell Gauss eloszlású kell legyen?
14. Igaz-e, hogy féligellenőrzött regressziónál azért is érdemes lehet lokális vizsgálatokat végezni, hogy a számítási igény kisebb legyen?
15. Előfordulhat-e, hogy a mintatér mechanikus részekre bontása is segít a bonyolult problémák részfeladatok
16. 3 osztályos féligellenőrzött tanulást végzünk, minden osztályból 10-10 ismert besorolású mintánk van, ezen felül 10.000 nem ismert besorolású mintánk van. Iniciált (seeded) k-átlagképző eljárásnál egy-egy mintapontból indítjuk-e a középpontok kiinduló értékeit?

17. Ha a k-átlagképző eljárásnál a tér véletlen pontjaiból indítjuk a középpontok kiinduló értékeit, akkor előfordulhat-e, hogy valamelyik középpont a harmadik iterációban nem reprezentál egyetlen mintapontot sem?
18. Előfordulhat-e a k-átlagképző eljárásnál, hogy valamelyik középpont az iterációk végén egy egyelemű halmazt reprezentál?
19. Kétbemenetű lineáris eszközünk VC dimenziója 3. A 2-dimenziós tér összes (akár speciális helyzetű) ponthármását tetszőleges címkézés mellett képes – megfelelő paraméterezés esetén – szeparálni?
20. Egy N-bemenetű 1-kimenetű rendszerről mintákat gyűjtöttünk, a bemeneti értékek vektora  $\mathbf{x}$ , a kimenet  $y$ . Melyik eloszlás írja le a rendszer bemeneti→kimeneti modelljét? (Pl.  $p(y)$ ,  $p(y, \mathbf{x})$ ,  $p(\mathbf{x})$ ,  $p(y|\mathbf{x})$ ,  $p(x, y)$ ,  $p(\mathbf{x}|y)$  stb.)
21. k-legközelebbi szomszéd (kNN) osztályozónál hogyan adható meg egy ismeretlen címkéjű mintapont besorolásának valószínűsége? (Tehát a kNN osztályozó szerint az  $\mathbf{x}$  pont az  $j$ -dik osztályba tartozik, és ennek a besorolásnak a bizonyosságát akarjuk jellemezni.)
22. Az EM algoritmusnál a következő összefüggéssel számoltunk:

$$E\{z_j^{(l)}\} = P(j|\mathbf{x}^{(l)}) = \frac{P(\mathbf{x}^{(l)}|j)P(j)}{\sum_r P(\mathbf{x}^{(l)}|r)P(r)}$$

EM algoritmussal ismeretlen besorolású minták eloszlását próbáljuk megbecsülni. A mintákról semmi többet nem tudunk csak azt, hogy 3 forrásból származnak. Mekkora állítsuk az egyes csoportok a priori valószínűségeit?