

10:00 - 11:00 (60 perc !)

***Minden kérdés két pontot ér (kapható 0, 1 vagy 2 pont), minden válaszhoz rövid, de áttekinthető indoklást is kérek! El kell érni az 50%-ot!***

Név (nyomatott betűvel):.....Neptun-kód:.....

Aláírás:.....

1. Több tízezen indulnak informatikus diplomát szerezni. Az útjuk, életük minden részletét egészen a diploma megszerzéséig vagy a halálukig egy központi adatbázisba gyűjtjük. Ennek felhasználásával igyekszünk informatikus diploma szerzésére irányuló optimális életstratégiát kialakítani. Milyen jellegű tanulási problémával van dolgunk? (Ellenőrzött, nemellenőrzött, megerősítéses, féligellenőrzött, induktív, deduktív, transzduktív...)
2. Kétbemenetű, lineáris eszközünk VC dimenziója 3. Az eszköz a 2-dimenziós tér összes (akár speciális helyzetű) ponthármását tetszőleges címkézéssel melletti képes – megfelelő paraméterezés esetén –szeparálni?
3. Milyen két fő csoportba osztjuk a klaszterezési eljárásokat?
4. Egy osztályozási feladatnál a  $P(y^{(\text{tanító})} = d^{(\text{tanító})})$  valószínűség mit jelent? (Az eszköz pontossága, általánosító képessége, hibája stb.)
5. A mintatér különböző méretű kisebb részekre bontása is segíthet egy bonyolult kétosztályos probléma egyszerűbb részfeladatokra való dekomponálásánál?
6. A k-átlagképző (k-means) eljárásunknál a 12 kezdeti középpontnak egy-egy tanító mintapontot választunk. Előfordulhat-e, hogy a felvett 12 középpont valamelyike az tanítási iterációk végén egy üres halmazt „reprezentál”?
7. Kétosztályos osztályozási feladatunknál ( $C_1$  és  $C_2$  osztályok) a mintákat 13 paraméter jellemzi: az  $x_1, \dots, x_{12}$  paraméterek binárisak, az  $x_{13}$  valós, értékészlete  $[0,100]$ . Nagyszámú ismert besorolású mintánk van a tanítóhalmazban, kb. a fele az egyik osztályba tartozik, a többi a másikba. Azt tapasztaljuk, hogy a  $C_1$ -be tartozó mintákra  $x_{13} < 73$ , a  $C_2$ -be jutókra nézve  $x_{13} > 75$ . Mennyi lesz a tanítóminta halmazon az  $x_{13} < 75$  teszt információ nyeresége?
8. MOE struktúrát használunk, a paramétereket egy tanító mintahalmaz segítségével az előadáson tanult módszerrel állítjuk elő. Milyen szerepet játszik a megtanított rendszer ismeretlen, új mintákra való használatánál az asszociált zajnak?
9. Két eszközünk van, egy lineáris és egy nemlineáris. A lineáris eszköznek 13 szabad paramétere van, a nemlineáris eszköznek 7. A paramétereket 8 biten ábrázoljuk (256 féle értéket vehetnek fel) Melyiket tekintjük nagyobb komplexitásúnak?
10. Generatív klaszterezésnél Gauss eloszlású kell legyen az egyes mintaosztályokat generáló modell?
11. Aktív megerősítéses tanuláskor a cselekvésvérték tanulás modellmentes eljárás?
12. Kétosztályos féligellenőrzött osztályozási feladatnál gráf-alapú (MINCUT) eljárást alkalmazunk. Ha kimerítő (teljes) keresést alkalmazunk, akkor a megvizsgálandó esetek száma hogyan függ az ismert besorolású minták  $N_L$  és az ismeretlen besorolású minták  $N_U$  számától?

***Jó munkát !***