Gyakorló feladatok

Gépi tanulás (vimim136) 2015 tavaszi félév

Ahol lehet, ott konkrét számértékeket várok nem puszta egyenleteket. (Azok egy részét amúgyis megadom.) NEM SZABAD AZT HINNI, HOGY PONTOSAN EZEK LESZNEK A ZH FELADATOK! Még az se igaz, hogy feltétlenül ilyenek lesznek csak más paraméterekkel! Ezek a gyakorlópálya a Kékesen, de a versenytúra a Himaláján lesz!

1. Egy bináris osztályozási feladatra tanított döntési fa építése során a 17-es csomópontba eljutott mintáink: 1213 tartozik az C1 osztályba és 3785 a C2 osztályba. A teszt elvégzése után keletkező három csomópontba a következő minták jutottak:

17/A csomópont C1: 403 C2: 1205

17/B csomópont C1: 396 C2: 1537

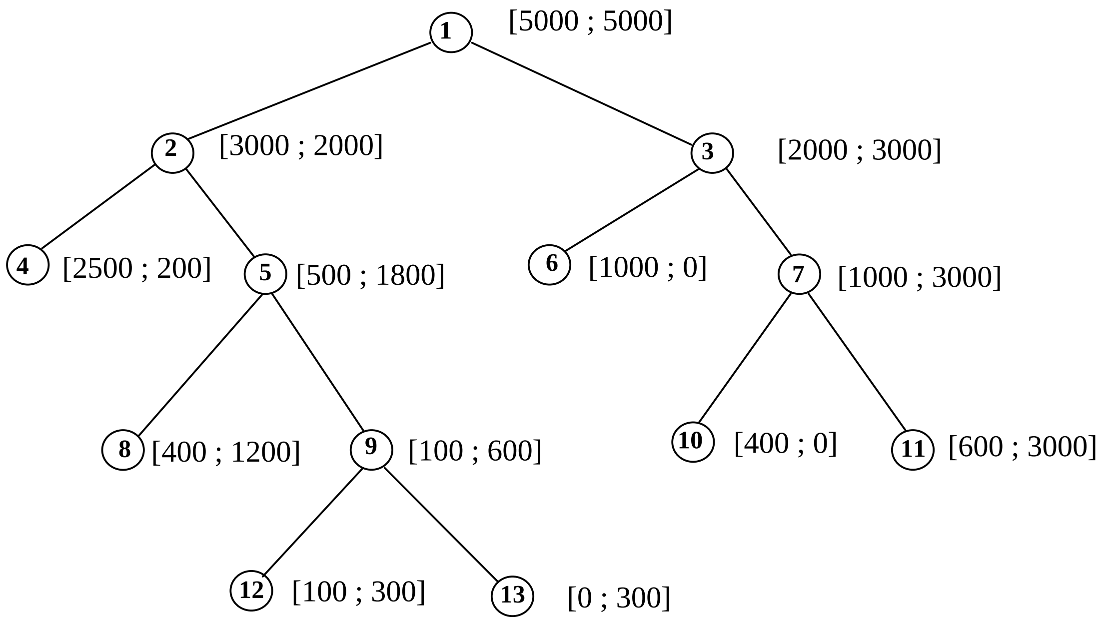
17/C csomópont C1: X C2: Y

Úgy döntöttünk, hogy ha a teszt valószínűleg (legalább 80% valószínűséggel) irreleváns, akkor bezárjuk a csomópontot, azaz a tesztet már nem végezzük el. A khi-négyzet szignifikancia teszt alapján mit fogunk csinálni?

χ^2 táblázat:

SzF	$P_{0,5}$	$P_{0,2}$	$P_{0,1}$	$P_{0,05}$	$P_{0,01}$	$P_{0,005}$	$P_{0,001}$
1	0,46	1,64	2,71	3,84	6,64	7,88	10,83
2	1,39	3,22	4,61	5,99	9,21	10,60	13,82
3	2,37	4,64	6,25	7,82	11,35	12,84	16,27
4	3,36	5,99	7,78	9,49	13,28	14,86	18,47
5	4,35	7,29	9,24	11,07	15,09	16,75	20,52
6	5,35	8,56	10,65	12,59	16,81	18,55	22,46
7	6,35	9,80	12,02	14,07	18,48	20,28	24,32
8	7,34	11,03	13,36	15,51	20,09	21,96	26,12
9	8,34	12,24	14,68	16,92	21,67	23,59	27,88
10	9,34	13,44	15,99	18,31	23,21	25,19	29,59

2. 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 azt mutatja, hogy abba a csomópontba hány *A*, illetve *B* osztálybeli tanítóminta jutott el. 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 eljárás vizsgálatot a kettő közül melyik csomópontból kiinduló részfa metszése a jobb? (Természetesen válaszát számítással támassza alá!)



3. Kétosztályos (bináris) osztályozási feladatot oldunk meg. 10 paramétert mértünk minden mintán, és a véletlen módon kiválasztott, mért és minősített (ismert osztályba sorolású) 7.000 minta alapján egy egyszerű perceptront tanítottunk, amelynél van eltolás is. (Az egyszerű perceptron lineáris elválasztó felületet képes megvalósítani, az eltolás miatt nem csak az origón átmenő hipersíkokat képes megtanulni.) Implementációs okokból a használt perceptronok mindegyik súlya vagy nulla vagy egy max. 3 biten ábrázolt kettő hatvány (magyarán 0, 1, 2, 4). A tanítás végén 100%-os pontosságot értünk el.
Kollegánk azt állította, hogy a kapott perceptron az új mintákon is legalább 95%-os pontosságot fog elérni, ebben ő legalább 90%-ban biztos. Jogos-e a véleménye, ha a mintaszám korlátot elfogadjuk pontos, jó becslésként (bár rendszerint nem az)?

4. Egy tanulási folyamat során a rendelkezésre álló – ismert kívánt válasszal rendelkező – mintáink közül 1370-et tanításra, 349-et tesztelésre használunk. Amikor az 1370 tanítómintával egyenként végrehajtjuk a tanítási lépést, utána lemérjük az eszköznek a tanítóhalmazon és a teszthalmazon mutatott átlagos hibáját: ezt így együtt nevezzük egy-egy tanítási epochnak. Majd újrakezdjük a folyamatot (második epoch, harmadik epoch) és így tovább. A következőt tapasztaltuk:

eddigi epochok száma	100	200	300	400	500	600	700	800	900
átl. hiba a tanítómintán	0,5	0,4	0,38	0,29	0,2	0,17	0,14	0,11	0,09
átl. hiba a tesztmintán	0,55	0,41	0,36	0,33	0,35	0,37	0,36	0,4	0,42

A. A táblázat alapján hány epoch után tapasztalunk túltanulást?
B. Azt javasolták, hogy a tanítás leállításakor vizsgáljuk meg a nagy hibájú mintákat, és a nagy hibájú tesztminták tulajdonságai alapján módosítsuk az eszközt, majd ezzel a módosított eszközzel tanítsunk újra ugyanezzel a módszerrel. Jó-e ez a javaslat?

5. **A.** Egy 3-dimenziós (bemeneti paraméterek: x1, x2, x3) háromosztályos osztályozási feladatot úgy oldottunk meg, hogy kétosztályos részfeladatokat tanítunk, és a részproblémák megoldásait integráljuk. Rajzolja fel a szakértőkből álló moduláris megoldást, a végső eredményt adó rendszerintegráló blokkokkal együtt!
B. Adja meg, hogy a harmadik (C_3) osztályba eső minta esetén a moduláris megoldás egyes blokkjainak kimenetein milyen értéket vár!

6. Van egy minták alapján tanítható eszközünk, amelyet kipróbáltunk egy adott kétosztályos osztályozási probléma esetén, és a – nagyon nagy – tanítóminta halmazon 0,47 hibaarányt értünk el. Van-e módszer rá, hogy ennek az eszköznek, és ennek a tanítóminta halmaznak a felhasználásával olyan összetett rendszert alakítsunk ki, amely 0,05 hibaarány elérésére képes ezen a tanítóhalmazon?

7. Kétbemenetű lineáris eszközünk három paraméterrel rendelkezik (van eltolás), és a paraméterek csak a 0 és +1 értékeket vehetik fel. Igaz-e, hogy ennek az eszköznek a VC dimenziója legalább 3?

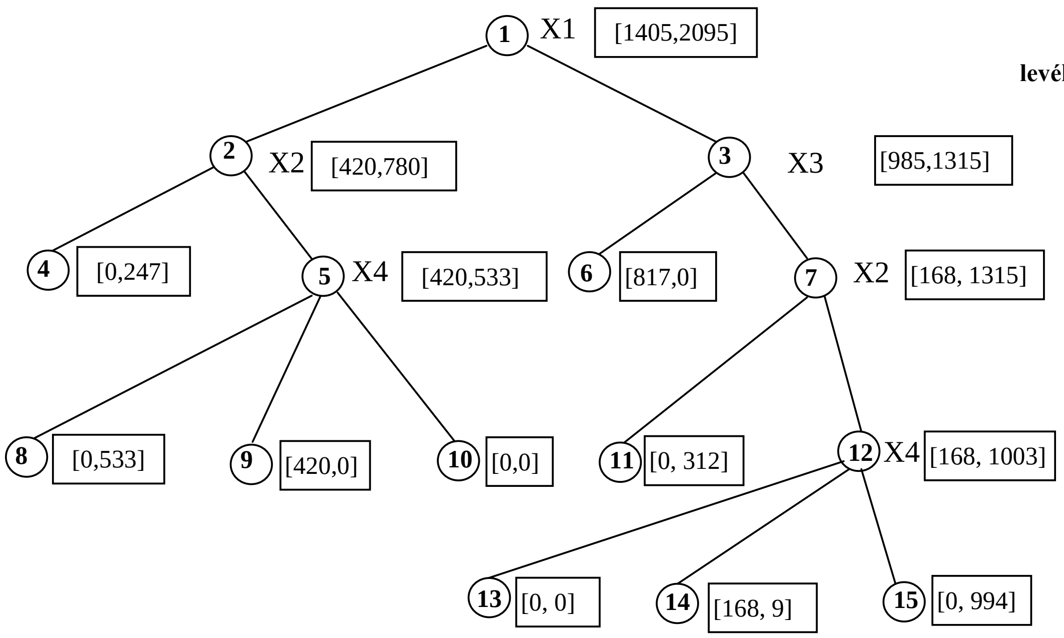
8. Egy osztályozási feladatnál a $P(y^{(teszt)} \neq d^{(teszt)})$ valószínűség mit jelent? (Az eszköz pontossága, általánosító képessége, hibája stb.)

9. 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**
$példák_i$ ← {a példák azon elemei, amelyre a *legjobb* attribútum értéke =v_i }
részfa ← DÖNTÉSI-FA-TANULÁS($példák_i$,*attribútumok*-*legjobb*,*m*)
a *fa* dötési fához adjunk egy v_i címkéjű ágat és a *részfa* részfát
endfor
return *fa* **%⇒ D**
endif

Az 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.



levél: besor. ; kijárat

4: …... ; ……

6: …... ; ……

8: …... ; ……

9: …... ; ……

10:….. ; ……

11:….. ; ……

13:….. ; ……

14:….. ; ……

15:….. ; ……

A. Adja meg, hogy melyik levéltől való visszatéréskor a pszeudókód melyik „kijáratánál” (A, B, C vagyD) léptünk ki, és melyik levélhez milyen besorolás tartozik (az odaérő ismeretlen mintákra mit fog „mondani” a megtanított döntési fa)! (9 levél van: az ábra mellett jobbra felsorolva. Ennél az alkérdésnél nem kérek indoklást!)

B. A gyökér (1) csomópontban hányszor fog meghívódni ez a DÖNTÉSI-FA-TANULÁS függvény?

C. Összesen hányszor fog meghívódni ez a függvény, mire előáll a felrajzolt döntési fa?

D. A gyökér (1) csomópontból melyik kijáraton tér vissza a függvény?

E. Más döntési fát alakított volna ki ez a tanuló függvény (DÖNTÉSI-FA-TANULÁS), ha a 8-as csomópontba jutó minták eloszlása [10,523], a 9-esbe jutóké [410,10] lett volna, és mindegyik másik csomópontba az ábrán látható számú minta jutott volna mindkét osztályból? (Ha mást, akkor mi lenne a különbség, ha nem mást, akkor miért ugyanez lenne?)

10. Kétosztályos osztályozási feladatunknál (C1 és C2 osztályok) a mintákat 8 paraméter jellemzi: az x1,…,x7 paraméterek binárisak, az x8 valós, értékkészlete [-20,120]. Nagyszámú ismert besorolású mintánk van a tanítóhalmazban, a fele az egyik osztályba tartozik, a fele a másikba. Azt tapasztaljuk, hogy a C1-be tartozó minták felére x8<75, a C2-be jutók felére az igaz, hogy x8>75 (pontosan 75 értéket egyik halmaz egyet-len mintájánál se találtunk). Mennyi lesz a tanítóminta halmazon az x8<75 teszt információ nyeresége?

11. Két szakértőből felépített MOE struktúrát használunk, a paramétereket egy tanító mintahalmaz segítségével az előadáson tanult módszerrel állítottuk elő. A mintákat 2 attribútum jellemzi (x1, x2). A kapuzó hálózat a $g_i = \dfrac{e^{\mathbf{v}_i^T \mathbf{x}}}{\sum_{k=1}^{2} e^{\mathbf{v}_k^T \mathbf{x}}}$ i=1,2 összefüggést valósítja meg a $\mathbf{v}_i$, i=1,2 paramétervektorok megtanítása után. Lineáris-e a két szakértő dominanciáját jellemző területeket elválasztó határvonal?

12. Két eszközünk van, egy lineáris és egy nemlineáris. A lineáris eszköznek 3 szabad paramétere van, a nemlineáris eszköznek 5. A lineáris eszköz paramétereit 4 biten ábrázoljuk (16 féle értéket vehetnek fel), a nemlineáris eszközét 2 biten (4-féle érték). Melyiket tekintjük nagyobb komplexitásúnak, ha a nemlinearitás jellege kötött, nemlineáris esetben is csak a paraméterek változhatnak?

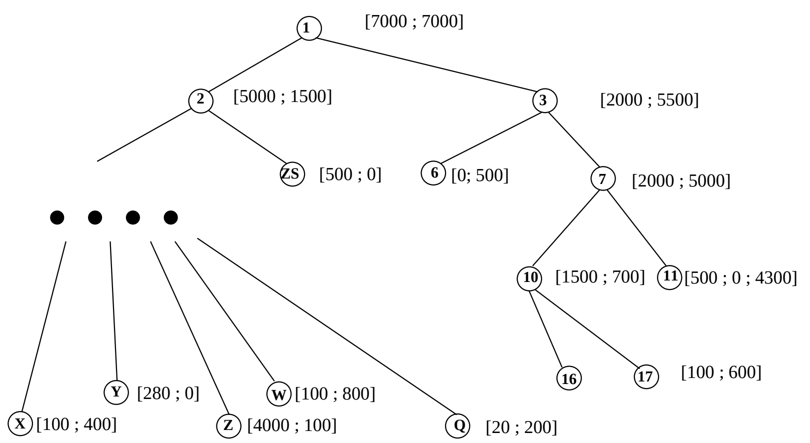
13. Két döntési fa készült ugyanazon problémára. A két fa különböző sorrendben teszteli az attribútumokat. Lehet-e a két fa konzisztens ugyanazzal a mintahalmazzal?

14. A mintatér egyforma méretű részekre bontása is segíthet egy bonyolult probléma egyszerűbb részfeladatokra való dekomponálásánál?

15. MOE struktúrát használunk, a MOE paramétereit egy tanító mintahalmaz segítségével az előadáson tanult módszerrel állítjuk elő. Milyen szerepet játszik a megtanított rendszer ismeretlen, új mintákra való alkalmazásánál az asszociált zaj?

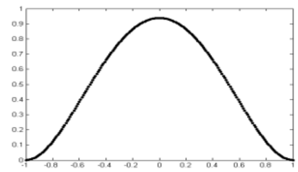
16. A következő – a minták két osztályba sorolását végző – döntési fát egy mintahalmaz alapján tanítottuk. A minták az *A* és *B* osztályba tartoznak, eredetileg 7000-7000 tanítóminta volt mindkét osztályból. Az egyes csomópontok mellett (jobbra) található két szám azt mutatja, hogy abba a csomópontba hány *A,* illetve *B* osztálybeli tanítóminta jutott el.

Mekkora az információnyereség a 2-es csomópontból kiinduló részfa tesztjeinek eredményeképp? (Figyelem: *nem a hibaarányt* kérdezem, mint a metszésnél, hanem az információnyereséget, és a teljes részfára!) A részfát nem rajzoltam be az ábrába, gyökere a 2-es csomópont, levelei az *X, Y, Z, W, Q, ZS* csomópontok.)



17. Egy 2 bemenetű 1 kimenetű MOE struktúrában nemlineáris kapuzó hálózatot és két lineáris szakértőt használunk. A szakértők paraméterei rendre $\theta 1=[0\ \ -0{,}3\ \ 0{,}2]^T$, $\theta 2=[0.1\ \ -0.2\ -0.2]^T$. A harmadik paraméter tartozik az eltoláshoz. A kapuzó hálózat megfelelő paramétervektorai $v1=[1\ 0\ 0]^T$, $v2=[1\ 1\ 1]^T$, itt is a harmadik paraméter tartozik az eltoláshoz. Az egyik kapuzó kimenet (g1) előállítása nem közvetlenül történik $\xi 1$-ből (a szokásos exponenciális összefüggéssel), hanem az $f(\xi_1)=(\mathbf{v}_1^T\mathbf{x})^2$ nemlinearitáson keresztül. Az első szakértő kimenetéhez egységnyi szórású Gauss zajt asszociálunk, de a második szakértőhöz <u>nem Gauss zajt asszociálunk</u>, hanem a következő sűrűségfüggvénnyel rendelkező zajt:

$$P(y-\mu)=\begin{cases}0.9375*(\mathrm{y}-\mu)^4-1.8750(\mathrm{y}-\mu)^2+0.9375 & \text{ha } -1\le(\mathrm{y}-\mu)\le 1\\ 0 & \text{különben}\end{cases}$$



A. Adja meg a MOE struktúra kimeneti jelét, ha a bemeneti vektor $\mathbf{x}=[+2\ -1]^T$!

B. Adja meg annak a görbének az egyenletét, illetve rajzolja is fel az x_1-x_2 síkon a görbét, amely elválasztja azt a térfelet, amelyiken az első szakértő kap nagyobb súlyt, attól, amelyiken a második szakértő (jelölje be, hogy melyik térrészben melyik dominál)!

C. Adja meg a második szakértő paramétervektorának új értékét, miután egy lépésben tanítottunk az $y^{(\ell)}=0$

$\mathbf{x}^{(\ell)}=[+2\quad -1]^T$ mintával! (A tanítás során a bátorsági faktor 0,1).