Gépi Tanulás (vimim136)
Zárthelyi

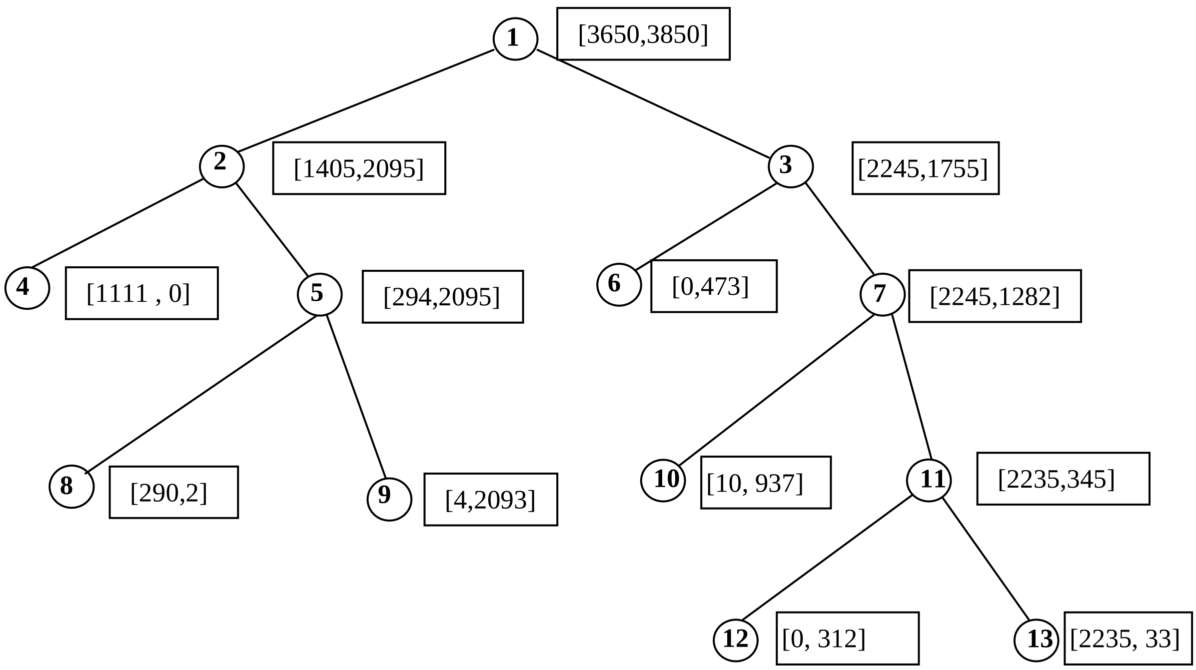
2012. március 22. **8:30 - 9:50 (80 perc !)**

Minden válaszhoz rövid, de áttekinthető indoklást is kérek! Mnidkét részből el kell érni a 40-40 %-ot!

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

Aláírás:..

1. Az alábbi ábrán egy 7500 minta alapján tanítással előállított döntési fa látható. A csomópontok mellett 2 szám látható egymás mellett bekeretezve szögletes zárójelben, ez a C_1, illetve C_2 osztályba jutó tanítóminták száma az adott csomópontban.



A. Mekkora lesz az információnyereség az 3-as csomópont tesztje nyomán? (**5 pont**)

B. Két kollegánk fogadott, hogy az 5-ös vagy a 7-es csomópontot kell-e előbb bezárni, ha a hibaarány-komplexitás kompromisszumon alapuló becslést használjuk. Viszont a komplexitást *nem a levelek számával*, *hanem a részfa összes* (levél és belső) *csomópontjának számával* mérik. (Tehát pl. a teljes fa komplexitása 13, a 12-es levél csomóponté 1.) Melyiküknek lesz igaza? (**5 pont**)

2. Kétosztályos (bináris) osztályozási feladatot oldunk meg. 7 paramétert mértünk minden mintán, és a véletlen módon kiválasztott, mért és minősített (ismert osztályba sorolású) 8792 minta alapján egy egyszerű lineáris döntési határt megvalósító eszközt (súlyozott összegző + egy ugrásfüggvény) tanítottunk. A súlyok 2 különböző előjelű hatványai lehetnek, -4, -2, 0, +2 és +4. A tanítás végén 100%-os pontosságot értünk el.

Kollegánk azt állította, hogy a kapott eszköz az új 23571 mintánkon is legalább 95%-os pontosságot fog elérni, ebben ő legalább 99%-ban biztos. Indokolt-e a véleménye? (**5 pont**)

3. A döntési fák csomópontjában alkalmazott teszt kiválasztásánál mi a különbség az *információ nyereség arány* használata és az *információ nyereség* használata közt? Mutasson be egy példát, amikor előnyösebb a *nyereség arány* használata? (**5 pont**)

****

4. Szakértő együttest tanítunk, az *i*-dik szakértő kimenetét jelölje y_i, az eredő kimenetet az egyes hálók válaszainak súlyozott lineáris összegeként kapjuk: $y=\sum_{i=1}^{K} w_i \cdot y_i$, ahol $\sum_{i=1}^{K} w_i = 1$. Vezesse le az órán tanult összefüggést: hogyan írható fel a teljes rendszer átlagos négyzetes hibája az egyes hálók átlagos négyzetes hibája és a végső kimenettől való eltérése alapján! Mit mutat ez az összefüggés? (**5 pont**)

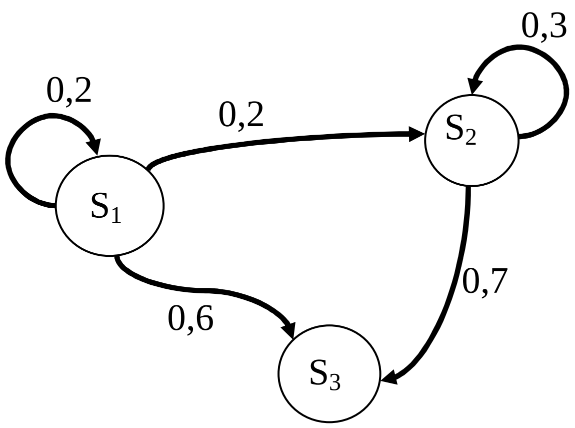
5. Egy bináris osztályozási problémát Adaboost eljárással akarunk megoldani. Matematikus barátunk bebizonyította, hogy a felhasznált alap eljárás az adott probléma bármelyik mintahalmazára, bármilyen – az Adaboostban használt – súlyozás esetén legalább 53% osztályozási pontosságot elér. Hány turbózási lépést hajtsunk végre, hogy az eredő eszközzel a 13876 tanítómintán biztosan elérjünk 90% pontosságot? (**5 pont**)

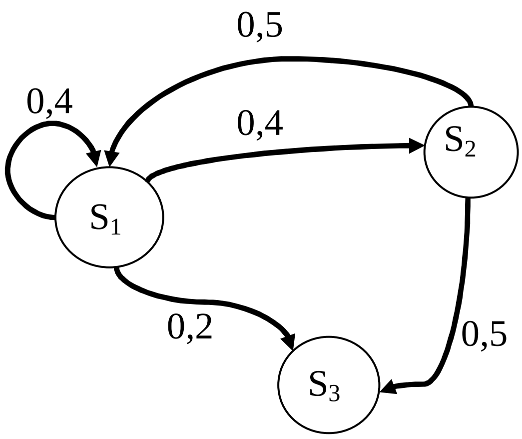
6. Egy 2 bemenetű 1 kimenetű (MISO) MOE struktúrában <u>nemlineáris kapuzó hálózatot</u> és két <u>lineáris szakértőt</u> használunk. A kapuzóhálózat nemlinearitását az okozza, hogy a g_1 és g_2 súlyok számításánál $\xi_1 = \left(\mathbf{v}_1^T \mathbf{x}\right)^3$ és $\xi_2 = \mathbf{v}_2^T \mathbf{x}$. A szakértők paraméterei rendre $\boldsymbol{\theta}_1$=[-1 2]^T , $\boldsymbol{\theta}_2$=[0 1]^T, nincs eltolás (*bias*). A kapuzó hálózat megfelelő paramétervektorai $\mathbf{v}_1$=[0,5 -1]^T, $\mathbf{v}_2$=[-1 3]^T, itt sincs eltolás. A szakértők kimenetéhez nulla várható értékű, Gauss zajt asszociálunk, az első szakértőnél σ_1=2, a másodiknál σ_2=1. A tanítás során a bátorsági faktor 0,01.

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

B. Adja meg a kapuzó hálózat $\mathbf{v}_2$ paramétervektorának új értékét! A tanítást az $\mathbf{x}^{(l)}$=[2 1]^T, $y^{(l)}$=0 mintával végezzük! (**3 pont**)

7. Egy szekvenciális döntési problémában 3 állapot alkotja az állapotteret, S_3 a végállapot. Minden állapotban kétféle cselekvés (a_1 és a_2) közt választhatunk. Az alábbi baloldali ábrán láthatók az a_1-hez tartozó állapot-átmenet valószínűségek, a jobboldali ábrán az a_2 cselekvéshez tartozók. A leszámítolási tényező γ=0,1.





Az egyes állapotokhoz tartozó jutalmak $R(S_1)$ = - 1 ; $R(S_2)$ = +1,4 ; $R(S_3)$=+2. Értékiterációs algoritmust alkalmazunk, a kiinduló hasznosságok $U_0(S_1) = U_0(S_2)$ = 1, $U_0(S_3) = R(S_3)$= 2. Mi lesz a következő (első) iterációs ciklusban az S_2 állapothoz tartozó hasznosság kiszámított értéke? (**5 pont**)

Jó munkát !

Megadott paraméterek, összefüggések

$$I(P(\nu_1),..., P\nu_K) = -\sum_{i=1}^{K} P(\nu_i)\cdot \log_2\left(P(\nu_i)\right)$$

$$M(A) = \sum_{i=1}^{N} \frac{n_{i,1}+...+n_{i,N}}{n_1+...+n_N}\cdot I(P(\nu_{i1}),...,P(\nu_{iK}))$$

$$D = \sum_{i=1}^{k}\left(\frac{(p_i-\hat{p}_i)^2}{\hat{p}_i} + \frac{(n_i-\hat{n}_i)^2}{\hat{n}_i}\right)$$

$$R(T_n) + \alpha\cdot|T_n| = R(\{T_n\}) + \alpha\cdot|\{T_n\}|$$

$$\Pr(y(\mathbf{x}_{teszt}) \neq d_{teszt}) \;\leq\; \Pr(y(\mathbf{x}_{\tan ító}) \neq d_{\tan ító}) \;+\; O(\sqrt{\frac{Td_{VC}}{N}})$$

$$U^{\pi}(s) = R(s) + \gamma \max_a \sum_{s'} T(s,a,s')U^{\pi}(s')$$

$$\pi^*(s) = \arg\max_a \sum_{s'} T(s,a,s')U(s')$$

$$\frac{1}{\varepsilon}\left(\ln(|H|) - \ln(\delta)\right) \leq N$$

$$\varepsilon_t = \sum_{\substack{i=1 \\ y_t(\mathbf{x}_i)\neq d_i}}^{N} D_t(i)$$

$$\alpha_t = \frac{1}{2}\ln(\frac{1-\varepsilon_t}{\varepsilon_t})$$

$$D^*_{t+1}(i) = \begin{cases} D_t(i)\cdot\exp(-\alpha_t) & \text{ha } y_t(\mathbf{x}_i) = d_i \\ D_t(i)\cdot\exp(+\alpha_t) & \text{ha } y_t(\mathbf{x}_i) \neq d_i \end{cases}$$

$$D_{t+1}(i) = \frac{D^*_{t+1}(i)}{\sum_{k=1}^{N} D^*_{t+1}(k)}$$

$$y(\mathbf{x}) = sign(\sum_{t=1}^{T}\alpha_t y_t(\mathbf{x}))$$

$$m(\mathbf{x}_k, d_k) = \frac{d_k\sum_t \alpha_t y_t(\mathbf{x}_k)}{\sum_t \alpha_t}$$

$$\Pr(y(\mathbf{x}_{teszt}) \neq d_{teszt}) \;\leq\; \Pr(|m(\mathbf{x}_{tanító}, d)| \leq \theta) \;+\; O(\sqrt{\frac{d_{VC}}{N\theta^2}})$$

$$\Pr(y(\mathbf{x}) \neq d) \leq \prod_t [2\sqrt{\varepsilon_t(1-\varepsilon_t)}\;] = \prod_t \sqrt{1-4\gamma_t^2} \leq e^{-2\sum_t \gamma_t^2}$$

$$\Pr(y(\mathbf{x}) \neq d) \leq e^{-2T\gamma^2}$$

$$y = \mu = \sum_{k=1}^{K} g_k\mu_k$$

$$g_i^{(\ell)} = \frac{e^{f_i(\mathbf{v}_i^{\mathbf{T}}\mathbf{x}^{(\ell)})}}{\sum_{k=1}^{K} e^{f_k(\mathbf{v}_k^{\mathbf{T}}\mathbf{x}^{(\ell)})}}$$

$$h_i^{(\ell)} = \frac{g_i^{(\ell)}P(y^{(\ell)}|\mathbf{x}^{(\ell)},\mathbf{\Theta}_i)}{\sum_j g_j^{(\ell)}P(y^{(\ell)}|\mathbf{x}^{(\ell)},\mathbf{\Theta}_j)}$$

$$\frac{\partial L}{\partial \mu_i} = \frac{g_i^{(\ell)}}{\sum_j g_j^{(\ell)}P(y^{(\ell)}|\mathbf{x}^{(\ell)},\theta_j)}\cdot\frac{\partial}{\partial\mu_i}P(y^{(\ell)}|\mathbf{x}^{(\ell)},\theta_i)$$

$$\frac{\partial L}{\partial \mathbf{\Theta}_i} = h_i^{(\ell)}\cdot\frac{(y^{(\ell)} - \mu_i^{(\ell)})\mathbf{x}^{(\ell)}}{\sigma_i^2}$$

$$\frac{\partial L}{\partial \mathbf{v}_i} = (h_i^{(\ell)} - g_i^{(\ell)})\cdot\mathbf{x}^{(\ell)}$$

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