Gépi tanulás (vimim136) Pótzárthelyi 2012. április 23.

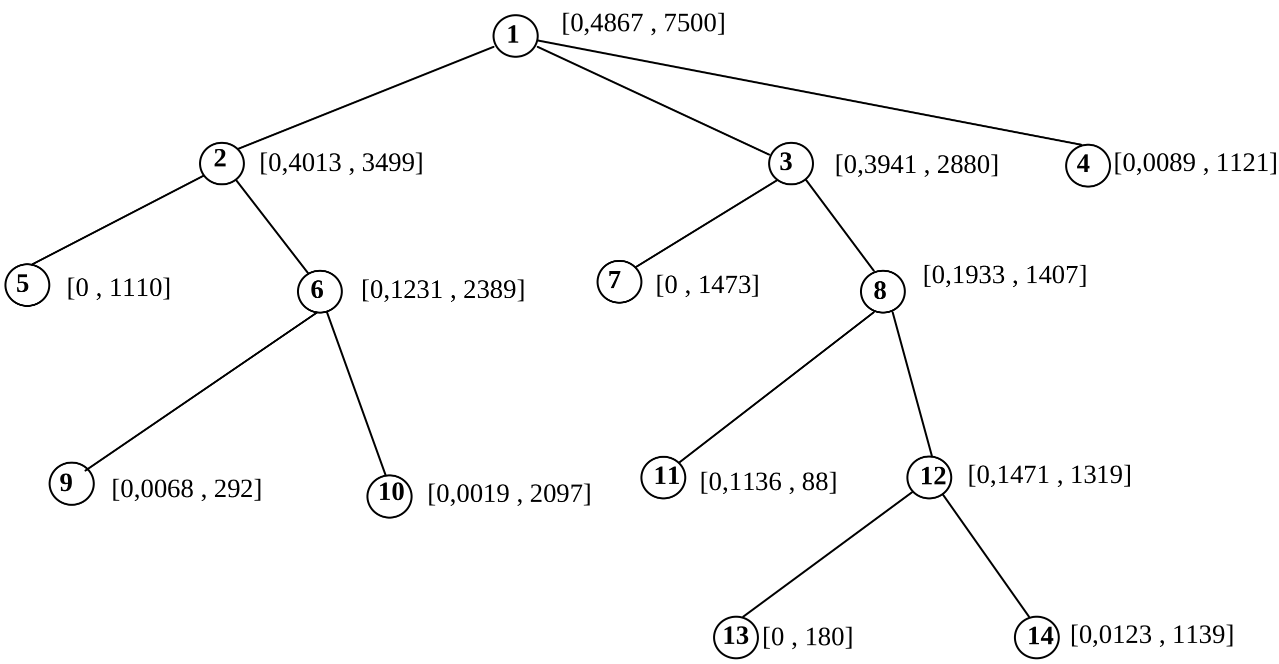
16:15 - 17:35 (80 perc !)

Minden válaszhoz rövid, de áttekinthető indoklást is kérek! Mindkét 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 gyökér csomópontban a C_1 osztályhoz 3650 minta tartozott, a C_2 osztályhoz 3850). A csomópontok mellett 2 szám látható egymás mellett szögletes zárójelben, ez az adott csomópontban mérhető hibaarány, illetve az összes idejutó tanítóminták száma.



A. Mekkora lesz az információnyereség az 1-es csomópont tesztje nyomán? (**5 pont**)

B. A χ^2 szignifikancia vizsgálat alapján megérte-e elvégezni a tesztet az 1-es csomópontban? (**5 pont**)

C. Az 1-es csomópontban mekkora az ábrán látható teszt eredményeképp kialakuló *információ nyereség arány*? (**5 pont**)

2. Kétosztályos (bináris) osztályozási feladatot oldunk meg. 17 paramétert mértünk minden mintán, és egy olyan 17 bemenetű, 1 kimenetű eszközt tanítunk, amelynek 8, a tanítás során állítható paramétere van. Ezek közül 4 paraméter a {-4,-2,-1,0,1,2,4} értékeket veheti fel, a másik 4 paraméter viszont csak a {-1,0,1} értékeket.

Az a piaci igény, hogy legalább 97%-os pontosságot érjen el az eszközünk, az ésszerű kockázatvállaláshoz ezt legalább 95% biztonsággal kell tudnunk garantálni. Az eszközünket 300 mintán tanítottuk, ezen 100% pontosságot ért el. Elég-e ez ahhoz, hogy a megfelelő biztonsággal vállaljuk a kívánt pontosságot? (**5 pont**)

****

3. Szakértő együtteseknél az a célunk, hogy egymástól minél inkább különböző, de minél pontosabb szakértőket hozzunk létre. Milyen módokon próbáljuk elérni, hogy a szakértőink különbözőek legyenek? (**3 pont**)

4. Egy osztályozásra használt eszközt turbózással (adaboost) javítunk. A kiinduló mintahalmazunkat az első, kiinduló eszköz így osztályozta (C_i az *i*-edik bemeneti minta valós osztálybasorolása, Y_i az *i*-edik bemeneti mintának az eszközünk által adott osztálybasorolása):

i	1	2	3	4	5	6	7	8	9
C_i	+1	+1	-1	+1	+1	-1	-1	-1	+1
Y_{1i}	+1	+1	+1	+1	-1	-1	-1	-1	-1

A. Hogyan kell súlyoznunk a mintákat a következő neuronháló tanításához? Ha az algoritmusunk nem tud közvetlenül mintasúlyokkal dolgozni (pl. egy tipikus MLP neuronháló), akkor hogyan oldhatjuk meg, hogy mégis két tizedesjegy pontossággal figyelembe vegyük a minták súlyait? (**4 pont**)

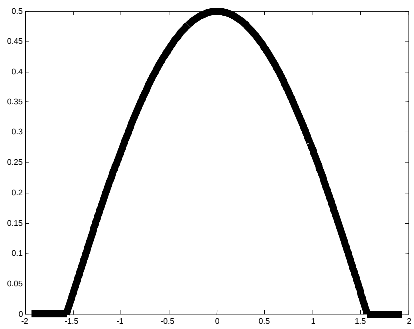
B. Tegyük fel, hogy a turbózás során kapott második eszköz által adott eredmény:

i	1	2	3	4	5	6	7	8	9
C_i	+1	+1	-1	+1	+1	-1	-1	-1	+1
Y_{2i}	+1	+1	-1	+1	-1	-1	-1	-1	-1

Rajzolja fel a kialakuló eredő rendszert, amennyiben e két háló kialakítása után befejezzük a turbózási eljárást! Adja meg az egyes hálókhoz tartozó súlyokat! (**2 pont**)

5. Egy 2 bemenetű 1 kimenetű MOE struktúrában *lineáris* kapuzó hálózatot és két *lineáris* szakértőt használunk. A szakértők paraméterei rendre $\boldsymbol{\theta}_1=[0\ \ -0{,}3\ \ 0{,}2]^T$, $\boldsymbol{\theta}_2=[0.1\ \ 0.2\ -0.2]^T$. Az *első* paraméter tartozik az eltoláshoz. A kapuzó hálózat megfelelő paramétervektorai $\mathbf{v}_1=[1\ 0\ 0]^T$, $\mathbf{v}_2=[0\ 1\ 1]^T$, itt is az *első* paraméter tartozik az eltoláshoz. A tanítás során a két szakértő kimenetéhez nem Gauss zajt asszociálunk, hanem a következő sűrűségfüggvénnyel rendelkező zajt:

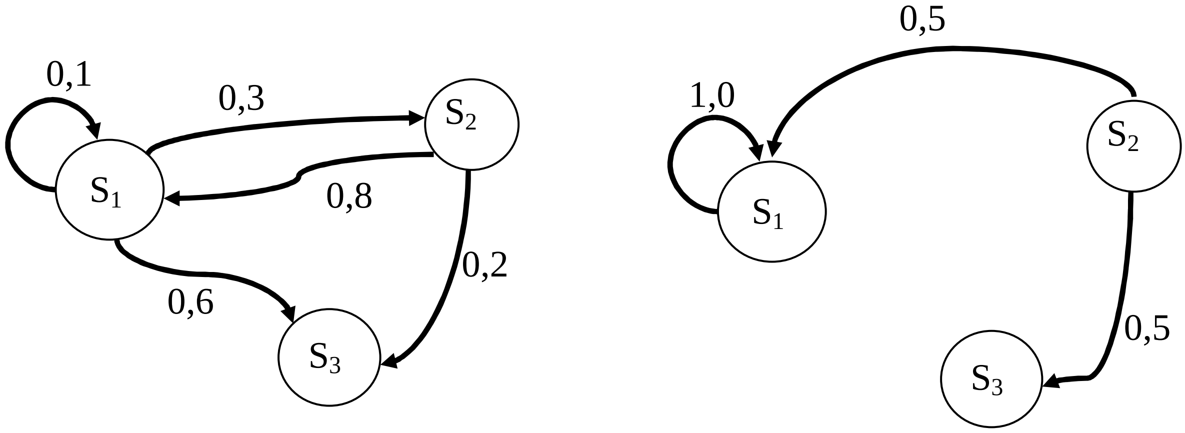
$$P(y-\mu)=\begin{cases}\cos(y-\mu)/2 & \text{ha} \quad -\pi/2 \le (y-\mu) \le \pi/2 \\ 0 & \text{különben}\end{cases}$$



A. Adja meg a MOE struktúra kimeneti jelét, ha a bemeneti vektor $\mathbf{x} = [+1\ \ 0]^T$! (**2 pont**)

B. Adja meg a második szakértő paramétervektorának új értékét, miután egy lépésben tanítottunk az $y^{(\ell)} = 0{,}3 \quad \mathbf{x}^{(\ell)} = [+1 \quad 0]^T$ mintával! (A tanítás során a bátorsági faktor 0,1). (**4 pont**)

6. 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. A $T(s \rightarrow s', a_i)$ állapot-átmenet valósszínűségeket a következő két táblázat tartalmazza:



Az egyes állapotokhoz tartozó jutalmak $R(S_1)$ = - 1 ; $R(S_2)$ = +0,4 ; $R(S_3)$=+2. Értékiterációs algoritmust alkalmazunk, a kiinduló hasznosságok $U_0(S_1) = U_0(S_2) = 0$, $U_0(S_3)= R(S_3)= 2$. Mi lesz a *következő két* (*első és második*) iterációs ciklusban az S_2 állapothoz tartozó hasznosság kiszámított értéke? (**5 pont**)

Jó munkát !

Megadott paraméterek, használható összefüggések

χ^2 táblázat

SzF	$\mathbf{P_{0,5}}$	$\mathbf{P_{0,2}}$	$\mathbf{P_{0,1}}$	$\mathbf{P_{0,05}}$	$\mathbf{P_{0,01}}$	$\mathbf{P_{0,005}}$	$\mathbf{P_{0,001}}$
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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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