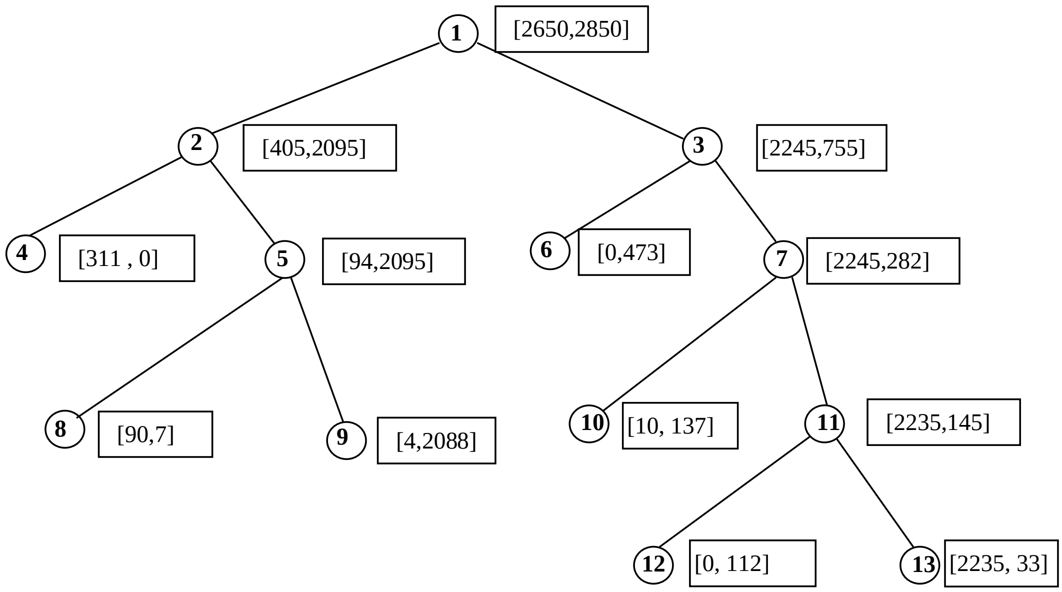
Gépi tanulás (VIMIM136) **Pótzárthelyi** **2014. május 9. 8:30 – 9:50 (80 perc!)**

Minden válaszhoz rövid, de áttekinthető indoklást is kérek, kivéve az IGAZ/HAMIS feleletválasztós kérdéseket (2. feladat)! A megfelelt szint: 40%.

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

Aláírás:..

1. Az alábbi ábrán egy 5500 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 2-es csomópont tesztje nyomán? **(5 pont)**

B. Két kollégánk fogadott, hogy az 2-es vagy a 7-es csomópontot kell-e előbb bezárni, ha a hibaarány-komplexitás kompromisszumon alapuló becslést használjuk. A komplexitást a levelek számával mérjük. Melyiküknek lesz igaza? **(3 pont)**

2. **a.** Kétbemenetű lineáris eszközünk öt paraméterrel rendelkezik, a paraméterek csak a 0, +1 vagy +2 értékeket vehetik fel, ezért ennek az eszköznek a VC dimenziója kisebb mint 5. **a. Igaz Hamis**

b. Egy bináris osztályozási feladatnál a $P\left(d^{(tanító)} \neq y^{(tanító)}\right)$ valószínűség az általánosító képességet adja meg. **b. Igaz Hamis**

c. A mintatér minden előzetes ismeret nélküli részekre bontása is segíthet egy bonyolult probléma egyszerűbb részfeladatokra való dekomponálásánál. **c. Igaz Hamis**

d. Egy döntési fa csomópontjában elvégzett attribútum vizsgálat relevanciáját nézzük meg χ^2 teszttel. A döntési fában elvégzett attribútum vizsgálat akkor releváns, ha a χ^2 tesztben kiszámított valószínűségi változó nulla közelében van (kisebb egy adott limitnél). **d. Igaz Hamis**

e. Ha 10-bemenetű (egy kimenetű) logikai függvényt megvalósító eszközt tanítunk, akkor a hipotézistér mérete 2^{10}=1024. **e. Igaz Hamis**

f. Két döntési fa készült ugyanazon problémára. A két fa *különböző attribútumokat* tesztel, de ettől még lehetnek konzisztensek ugyanazzal a mintahalmazzal. **f. Igaz Hamis**

(*≤3 jó válasz* ***0 pont****,* *< jó válasz* ***(jó válaszok száma-3) pont****,* *6 jó válasz* ***3 pont***)

3. 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 $\mathbf{v}_1=[2\ 0\ 0]^T$, $\mathbf{v}_2=[1\ 0\ 1]^T$, itt is az *első* paraméter tartozik az eltoláshoz. A tanítás során a két szakértő kimenetéhez nulla várhatóértékű Gauss zajt asszociálunk, az egyes szórások rendre $\sigma_1=0{,}5$ és $\sigma_2= 2$.

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

3.B Adja meg a kapuzó hálózat paramétervektorainak új értékét, miután egy lépésben tanítottuk az $\mathbf{x}^{(\ell)} = [+1 - 1]^T$, $y^{(\ell)} = -0{,}2$ mintával! (A tanítás során a bátorsági faktor 0,05). **(4 pont)**

4. 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 legfeljebb 48% osztályozási hibát vét. Hány turbózási lépést hajtsunk végre, hogy az eredő eszközzel az 57 313 876 tanítómintán biztosan elérjünk 90% pontosságot? **(5 pont)**

5. 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)**

Jó munkát !

A pótzárthelyin használható összefüggések

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

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

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

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

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

$$h(\mathbf{x}) = sign\left(\sum_{t=1}^{T}\alpha_t h_t(\mathbf{x})\right) \qquad \alpha_t = \frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$$

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

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

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

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

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

$$\Pr(h(\mathbf{x}_{teszt})\ne d) \le \Pr(h(\mathbf{x}_{tanító})\ne d) + O\left(\sqrt{\frac{Td_{VC}}{N}}\right)$$

$$m(\mathbf{x}_k, d_k) = \frac{d_k\sum_t \alpha_t h_t(\mathbf{x}_k)}{\sum_t \alpha_t} = \frac{d_k\cdot h'(\mathbf{x}_k)}{\sum_t \alpha_t}$$

$$\Pr(h(\mathbf{x}_{teszt})\ne d_{teszt}) \le \Pr(|m(\mathbf{x}_{tanító}, d_{tanító})| \le \theta) + O\left(\sqrt{\frac{d_{VC}}{N\theta^2}}\right)$$

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

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

$$\frac{\partial L}{\partial \boldsymbol{\Theta}_i} = h_i^{(\ell)} \cdot \frac{(y^{(\ell)} - \boldsymbol{\mu}_i^{(\ell)})\mathbf{x}^{(\ell)}}{\sigma_i^2}$$ Gauss eloszlású asszociált zaj, lineáris szakértő

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

$$E\left\{z_i^{(\ell)}(t+1)\right\} = \hat{P}\left(i \middle| x^{(\ell)}, \boldsymbol{\theta}(t+1)\right) = \frac{\frac{1}{\hat{\sigma}_i(t)} e^{-\frac{(x^{(\ell)}-\hat{\mu}_i(t))^2}{2\cdot\hat{\sigma}_i(t)^2}} \hat{P}_i(t)}{\sum_{j=1}^{K} \frac{1}{\hat{\sigma}_j(t)} e^{-\frac{(x^{(\ell)}-\hat{\mu}_j(t))^2}{2\cdot\hat{\sigma}_j(t)^2}} \hat{P}_j(t)}$$

$$\hat{\mu}_i(t+1) = \frac{\sum_{\ell=1}^{L} E\left\{z_i^{(\ell)}(t+1)\right\} \cdot x^{(\ell)}}{\sum_{\ell=1}^{L} E\left\{z_i^{(\ell)}(t+1)\right\}} \qquad \hat{\sigma}_i(t+1)^2 = \frac{\sum_{\ell=1}^{L} E\left\{z_i^{(\ell)}(t+1)\right\} \cdot (x^{(\ell)} - \hat{\mu}_i(t+1))^2}{\sum_{\ell=1}^{L} E\left\{z_i^{(\ell)}(t+1)\right\}}$$

$$\hat{P}_i(t+1) = \frac{1}{L} \sum_{\ell=1}^{L} \hat{P}\left(i \middle| x^{(\ell)}, \boldsymbol{\theta}(t+1)\right) = \frac{1}{L} \sum_{\ell=1}^{L} E\left\{z_i^{(\ell)}(t+1)\right\}$$

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