Gépi tanulás (VIMIM136) **Pótzárthelyi** **2014. május 20. 9:00 – 10:20 (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. Egy kétosztályos (bináris) osztályozási problémára döntési fát akarunk kialakítani 6000 minta alapján történő tanítással. A mintahalmaz 3000 C_0 és 3000 C_1 osztálybeli elemet tartalmaz. A gyökércsomópontba javasolt teszt 3500 különböző kimenetellel rendelkezik, ennek megfelelően a teszt elvégzése után 3500 gyermek-csomópontot kaptunk: 2500 csomópontba 2-2 minta jutott, 1000 csomópontba pedig 1-1 minta. Azt találtuk, hogy minden gyermekcsomópont homogén, tehát vagy csak C_0 vagy csak C_1 osztályba tartozó elemet/elemeket tartalmaz. Mekkora lesz az információnyereség-arány a teszt nyomán? **(3 pont)**

2. **a.** 4-bemenetű lineáris eszközünk öt valós paraméterrel rendelkezik, ezért ennek az eszköznek a VC dimenziója kisebb mint 5. **a. Igaz Hamis**

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

c. Egy problémában két valós attribútum jellemzi az eseteket: x_1 és x_2; tudjuk, hogy $X_{1\text{Lim1}} \leq x_1 \leq X_{1\text{Lim2}}$ és . $X_{2\text{Lim1}} \leq x_2 \leq X_{2\text{Lim2}}$. A feladat négy – kisebb mintatérben megoldható – feladatra való mechanikus dekomponálása a két attribútumtartomány felező, azaz $(X_{1\text{Lim1}}+X_{1\text{Lim2}})/2$ és $(X_{2\text{Lim1}}+X_{2\text{Lim2}})/2$ mentén biztosan nem könnyíti meg a megoldást. **c. Igaz Hamis**

d. Ha egy N-osztályos osztályozási feladatot kétosztályos részfeladatokra dekomponálunk, akkor $N\cdot(N-1)/2$ részfeladatot kell megoldanunk, és a részeket integrálnunk. **d. Igaz Hamis**

e. Ha rögzített struktúrájú, 10 valós paraméteres, egy-kimenetű függvényt megvalósító eszközt tanítunk, akkor a hipotézistér mérete nagyon nagy, de véges. **e. Igaz Hamis**

f. Két döntési fa készült ugyanazon problémára ugyanazt a tanítóhalmazt felhasználva. Ha mindkét fa konzisztens a tanító halmazzal, akkor nem adhatnak különböző választ egy, a tanításban fel nem használt, ismeretlen mintára. **f. Igaz Hamis**

g. Turbózás (boosting) eljárásnál az ötödik iterációban a tanításnál csökkentjük azoknak a mintáknak a súlyát, amelyeket a negyedik osztályozó jól osztályozott. **g. Igaz Hamis**

h. Egy adott, rögzített mintahalmazon a túltanulás jelentkezése rendszerint csökkenti a *tesztmintákon* elért hibát, miközben a *tanítóminta* halmazon még nő a hiba. **h. Igaz Hamis**

i. Egy kétbemenetű lineáris eszköz VC dimenziója 3. Ez azt jelenti, hogy a kétdimenziós tér minden (akár speciális helyzetű) ponthármasát képes tetszőleges címkézés mellett – megfelelő paraméterezés esetén – a kívánt két osztályba sorolni. **i. Igaz Hamis**

j. Ha egy tanítómintánál hiányzik az egyik attribútumérték, akkor egyszerű lehetőségként ennek az attribútumnak a többi mintán vett átlagával pótolhatjuk. **j. Igaz Hamis**

(≤ 5 *jó válasz:* **0 pont**, 5< *jó válasz:* **(jó válaszok száma-5) pont**, *10 jó válasz:* ***5 pont***)

3. Egy 2-bemenetű 1-kimenetű (MISO) MOE struktúrában <u>nemlineáris</u> kapuzó hálózatot és két lineáris szakértőt használunk. A szakértők paraméterei rendre $\boldsymbol{\theta}_1$=[1 2]^T , $\boldsymbol{\theta}_2$=[0 -1]^T, <u>nincs</u> eltolás (bias). A kapuzó hálózat megfelelő paramétervektorai $\mathbf{v}_1$=[2 1]^T, $\mathbf{v}_2$=[1 1]^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 0,1.

$$g_1 = \frac{e^{\xi_1}}{e^{\xi_1}+e^{\xi_2}} \quad ; \quad g_2 = \frac{e^{\xi_2}}{e^{\xi_1}+e^{\xi_2}} \qquad \xi_1 = \mathbf{v}_1^T\mathbf{x} \qquad \text{és} \qquad \xi_2 = f_2(\mathbf{v}_2^T\mathbf{x}) = (\mathbf{v}_2^T\mathbf{x})^3$$

A. Adja meg a MOE struktúra kimeneti jelét, ha a bemeneti vektor **x**=[0 -1]^T! **(2 pont)**

B. Adja meg a kapuzó hálózat $\mathbf{v}_2$ paramétervektorának új értékét! A tanítást az $\mathbf{x}^{(l)}$=[0 -1]^T $y^{(l)}$= 0 mintával végezzük! **(5 pont)**

4. Egy tanítóminta-halmazunk van, amely C_0, illetve C_1 osztályba tartozó elemeket tartalmaz. Elvégeztünk egy tesztet, amelynek 3 lehetséges kimenetele van (A, B és C), és a következő eredményre jutottunk:

	teszteredmény = A	teszteredmény = B	teszteredmény = C
C_0-beli elemek száma	3200	2600	3200
C_1-beli elemek száma	2800	3400	2800

A χ2 vizsgálatra alapozva mondhatjuk-e azt, hogy 0,1%-nál kisebb az esélye, hogy a teszt irreleváns? **(5 pont)**

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

Jó munkát !

A pót-pótzárthelyin használható összefüggések

$$I(P(v_1),...,P(v_K)) = -\sum_{i=1}^{K} P(v_i)\cdot \log_2\left(P(v_i)\right)$$

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

$$GR = \sum_{k=1}^{K} -\frac{p_k+n_k}{p+n}\cdot log_2(\frac{p_k+n_k}{p+n})$$

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

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

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

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

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

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

Gépi tanulás (VIMIM136) Pótzárthelyi 2014. május 20. 9:00 – 10:20 (80 perc!)

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

$$\Pr(h(\mathbf{x}_{teszt}) \neq d) \quad \leq \quad \Pr(h(\mathbf{x}_{tanító}) \neq d) \ + \ O(\sqrt{\frac{Td_{VC}}{N}})$$

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

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

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

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

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

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

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

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

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

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

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

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

χ^2 táblázat

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