Intelligens orvosi műszerek (VIMIA023)
Gyakorló feladatok, megoldással (2016 ősz)

Régi zárthelyi- és vizsgafeladatok, egyéb feladatok – megoldással. Nem jelenti azt, hogy pontosan ezek, vagy pontosan ilyenek lesznek a mostani pótzh-n, vizsgazh-n, vizsgán, csak felkészülési segítség kíván lenni!

1. Jelünket egy determinisztikus 117 ms periódusidejű periodikus zaj zavarja, amely 13 különböző szinuszos komponensből áll (és a komponensei közt nem tartalmaz konstanst, azaz egyik szinusz se elfajult, nem 0 frekvenciájú). Milyen mozgóablak átlagolást használjunk ahhoz, hogy a zaj hatását teljesen meg tudjuk szüntetni? (Válaszát természetesen indokolja!)

Megoldás:

Tudjuk, hogy a periodikus jel periódusideje a szinuszos komponenseinek egész számú többszöröse. Tehát ha a periódusidőt választjuk átlagolási ablakhossznak, akkor az összes szinuszos komponensből egész számú periódust átlagolunk. Viszont ha egy szinuszos jelet teljes periódusaira (tetszőleges számú periódusára) átlagolunk, akkor a kapott érték 0 lesz. Tehát a helyes válasz T_{MA}= 117 ms-os ablakkal kell az átlagolást végeznünk.

Kiegészítés:

Fontos, hogy ne legyen konstans komponense a jelnek, mert a konstans egy elfajult szinusz: végtelen a periódusideje. Ezért semmilyen véges ablakban nem ad nullát, ha összegezzük. Tehát például a következő jel periodikus:

$$x(k) = 5 + \sin\left(\frac{2 * \pi * k}{17}\right)$$

Ezt a jelet egy 17 (vagy 34, vagy 51,… vagy 187, …, vagy 34272,…) hosszúságú ablakban átlagolva a második komponens eltűnik, de az 5 megmarad.

2. Egy nyolcmilliós célcsoport szűrésének lehetőségét vizsgáljuk. Az adott betegségben a népesség 0,4%-a szenved. A szűrés költsége fejenként 1.700 forint, tehát ez mindenkinél jelentkezik. Ugyanakkor, ha az egészségeseket is betegnek mutatja a vizsgálat, ezt a tévedést csak egy újabb 8.200 Ft-os vizsgálattal lehet korrigálni. Ha idejében felismerjük a betegséget a szűrés, majd kontrollvizsgálat eredményeképpen, akkor a kezelés átlagosan 28.400 forintba kerül. Ha nem ismerjük fel idejében a betegséget, és később kezdjük el kezelni, akkor az átlagosan 273.000 forintba kerül a kezelés. Az előzetes vizsgálatot egy 23.576 fős mintán végeztük, akik közül 7.471 volt a beteg, a többi egészséges. A vizsgálat során 6891 beteget helyesen diagnosztizáltunk a szűrésre javasolt eljárással, a többi betegnél nem ismertük fel, hogy betegek. Viszont az egészségesek közül 312 esetet tévesen betegnek diagnosztizáltunk.
Az előzetes vizsgálatunk alapján mekkora az alkalmazni javasolt eljárásunk *szenzitivitása és specificitása*?

Megoldás:

Ez egy kismértékben beugratós kérdés. Ugyanis több adat van megadva, mint amennyire a válaszhoz szükség van. Általában is azt szeretném elérni, hogy ne alakuljanak ki a fejekben formális mintázatok: ha formailag így néz ki a kérdés, ezek vannak megadva, akkor ez és ez a válasz. Olvassák el, gondolkodjanak el, hogy mi a kérdés, és azokat az adatokat használják, amikre szükség van. Az élet is ilyen: ki kell szelektálnunk a sok és sokszor zavaros információból, hogy mi fontos az adott megoldandó probléma esetén, és mi nem fontos.

Formailag ez olyan leírás, mint egy Bayes-döntési probléma, adottak költségek, az a priori valószínűségek (betegek, egészségesek aránya a népességben) és kiszámíthatók a valószínűségek (P(0|0)=Spec és P(1|1)=Szenz). De a kérdés csak az előzetes vizsgálat alapján megbecsülhető valószínűségekre vonatkozik!
Tehát összesen erre a részre van sükségünk:

„Az előzetes vizsgálatot egy 23.576 fős mintán végeztük, akik közül 7.471 volt a beteg, a többi egészséges. A vizsgálat során 6891 beteget helyesen diagnosztizáltunk a szűrésre javasolt eljárással, a többi betegnél nem ismertük fel, hogy betegek. Viszont az egészségesek közül 312 esetet tévesen betegnek diagnosztizáltunk."

- 23.567 a teljes tesztmintánk, aminek alapján P(0|0)=Spec és P(1|1)=Szenz becslését elvégezzük. Ebből 7.471 beteg, tehát 23.576-7.471=16.105 az egészséges.
- A 7.471 betegből 6.891-nél felismertük, hogy beteg, tehát TP=6.891. Innen a szenzitivitás: P(1|1)=Szenz=TP/BetegekSzáma=6.891/7.471=0,92.
- A tesztmintában az egészségesek (16.105 fő) közül 312 esetben tévesen betegnek gondoltuk a résztvevőt, ezek a FP esetek (FP=312). A valódi negatív eset a többi, amikor az egészségest nem gondoltuk betegnek (=egészségesnek diagnosztizáltuk); TN=16.105-312=15.793. Ezek alapján P(0|0)=Spec=TN/EgészségesekSzáma=15.793/16.105=0,98.

3. Egy tanulási folyamat során a rendelkezésre álló – ismert kívánt válasszal rendelkező – mintáink közül 12.370-et tanításra, 1.349-et tesztelésre használunk. Amikor az 12.370 tanítómintával végrehajtjuk a tanítást, utána lemérjük az eszköznek a tanítóhalmazon és a teszthalmazon mutatott átlagos hibáját: ezt így együtt nevezzük egy-egy tanítási menetnek. Majd tovább tanítjuk az eszközt (második menet, harmadik menet és így tovább). A következőt tapasztaltuk:

eddigi menetek száma	10	20	30	40	50	60	70	80	90
átl. hiba a tanítómintán	0,5	0,4	0,38	0,29	0,23	0,17	0,14	0,11	0,09
átl. hiba a tesztmintán	0,55	0,41	0,36	0,38	0,41	0,37	0,36	0,4	0,42

A táblázat alapján hány menet után tapasztalunk túltanulást? (Természetesen indoklással!)

Megoldás:
Tipikusan túltanulásra utal, amikor a tanítómintákon még javul a teljesítmény, a tesztmintákon viszont romlani kezd vagy nem javul. (A kiadott mintapéldákat már szinte fejből tudja a hallgató, a kezdőszavakból emlékszik a jó válaszra, de a vizsgán esetleg ugyanolyan kezdőszavak után más lesz a kérdés – ezért rosszul válaszol. Mert túlzottan megtanulta a kiadott példák formáját, szavait, nem a lényegét.)
Azt látjuk, hogy itt a tanítómintán végig javul a teljesítmény: csökken a hiba, ahogy tovább és tovább tanítunk. A tesztmintán is javul a 30-dik menetig, utána viszont nem fejlődik, sőt némi ingadozással romlik. Tehát kb. a 30-dik menet környékén kezdődik a túltanulás.

4. Egy diagnosztikai eljárást vizsgáltunk meg 8.006 személy segítségével, akik közül 2.629 beteg volt (bár tudjuk, hogy a népességben ez a betegség csak 3,5%-ban fordul elő). A diagnosztika vizsgálata során egy z paramétert mértünk (a szemléletesség kedvéért legyen a testhőmérséklet), ennek különböző értékeinél tekintjük betegnek, illetve egészségesnek a pácienst. Azt tapasztaltuk, hogy a következő tartományokba esnek a *z* értékek a táblázatban megadott számú betegnél és egészségesnél.

$[z_{jmin}, z_{jmax})$	**[35,36)**	**[36,37)**	**[37,38)**	**[38,39)**	**[39,40)**	**[40,41]**
Betegek, akiknek $[z_{jmin}, z_{jmax})$ – ba esik a mért z paramétere	1874	538	124	54	32	7
Egészségesek, akiknek $[z_{jmin}, z_{jmax})$ – ba esik a mért z paramétere	52	168	469	725	1542	2421

A. Ezen adatok alapján olcsóbb-e, ha a [37,38) tartományba eső értékek esetén betegnek tekintjük az embereket, vagy az az olcsóbb, ha egészségesnek? A diagnosztikai eljárás 2.300 Ft-ba kerül személyenként, a tévesen egészségesnek diagnosztizált betegek kezelése 213.000 Ft-ba kerül, az időben diagnosztizált betegeké viszont csak 12.300 Ft-ba. A tévesen betegnek tekintetteknél átlagosan 5.000 Ft-ba kerül, mire kiderül, hogy valójában egészségesek.

B. Önmagában a szűrés költsége mennyiben befolyásolja, hogy az A. pontban vizsgált csoportot egészségesnek vagy betegnek célszerű-e tekinteni?

Megoldás:

A. Ez egy tipikus Bayes-döntési feladat. A Lim_1 =37 és Lim_2=38 közé eső csoportra a döntésünk legyen akkor „beteg”, ha $Lim_1 \le z < Lim_2$ esetén:

$$P_1 \cdot (C_{01} - C_{11}) \cdot P(z|1) > P_0 \cdot (C_{10} - C_{00}) \cdot P(z|0)$$

különben – ha a jobboldal a nagyobb – az „egészséges” döntést célszerű hozunk.

Ugyanis az egyenlőtlenség bal oldalán az a költségnövekedés szerepel, amikor a *z* értéket produkáló beteget tévesen egészségesnek vesszük, ahhoz képest mintha jól diagnosztizáltuk volna: betegnek.
Itt $P_1 \cdot P(z|1)$ annak a valószínűsége, hogy valaki beteg (P_1) és az adott tartományba eső *z* paraméterértéket produkálja ez a beteg $P(z|1)$. A kétféle döntésünk közti költségkülönbség pedig: $(C_{01} - C_{11})$.
Az egyenlőtlenség jobb oldalán pedig az a költségnövekedés szerepel, amikor a *z* értéket produkáló egészségest tévesen betegnek vesszük, ahhoz képest mintha jól diagnosztizáltuk volna: egészségesnek.
Itt $P_0 \cdot P(z|0)$ annak a valószínűsége, hogy valaki egészséges (P_0) és az adott tartományba eső *z* paraméterértéket produkálja ez az egészséges ember $P(z|0)$. A kétféle döntésünk közti költségkülönbség pedig: $(C_{10} - C_{00})$.

Majd ha bejön egy *z* értéket produkáló új személy, őt a nála mért *z* alapján vagy egészségesnek vesszük majd vagy betegnek. Lényegében azt választhatjuk meg, hogy átlagosan az egyenlőtlenség bal oldalán jelentkező többletköltséget szenvedjük el, vagy a jobb oldalon jelentkezőt.

Ha minden ilyen személyt betegnek veszünk, akkor a bal oldalon látható többletköltséget nem szenvedjük el, csak a jobboldalit, hiszen C_{01} költségű döntésünk nem lesz (nem döntünk úgy, hogy egészséges). A jobb oldalon látható költségnövekedés viszont megjelenik, mert mindenkinél betegre döntünk, tehát az egészségeseknél C_{10} lesz a költség C_{00} helyett.
Ha minden ilyen személyt egészségesnek veszünk, akkor a jobb oldalon látható többletköltséget nem szenvedjük el, csak a baloldalit, hiszen C_{10} költségű döntésünk nem lesz (nem döntünk úgy, hogy beteg). A bal oldalon látható költségnövekedés viszont megjelenik, mert mindenkinél egészségesre döntünk, tehát az betegeknél C_{01} lesz a költség C_{11} helyett.
Ha az egyenlőtlenség baloldala volt nagyobb, akkor átlagosan többe kerül nekünk, ha a beteget egészségesnek vesszük, tehát érdemes ezt elkerülnünk, és mindenkit, aki ezt a z értéket produkálja betegnek tekintenünk. Ha a jobb oldal a nagyobb, akkor viszont többet vesztünk, ha betegnek tekintjük az egészségest, tehát ezt el kell kerülnünk – tekintsük egészségesnek.

Egészítsük ki a táblázatunkat azzal, hogy milyenek az egyes tartományokba eső $P(z|1)$ és $P(z|0)$ értékek (a 8006 vizsgált személyből 2629 beteg, a többi 8006-2629=5377 egészséges):

$[z_{jmin}, z_{jmax})$	**[35,36)**	**[36,37)**	**[37,38)**	**[38,39)**	**[39,40)**	**[40,41]**
Betegek, akiknek $[z_{jmin}, z_{jmax})$ – ba esik a mért z paramétere	1874	538	124	54	32	7
$P(z\|1)$	$\frac{1874}{2629} = 0{,}71$	$\frac{538}{2629} = 0{,}20$	$\frac{124}{2629} = 0{,}05$	$\frac{54}{2629} = 0{,}02$	$\frac{32}{2629} = 0{,}01$	$\frac{7}{2629} = 0{,}003$
Egészségesek, akiknek $[z_{jmin}, z_{jmax})$ – ba esik a mért z paramétere	52	168	469	725	1542	2421
$P(z\|0)$	$\frac{52}{5377} = 0{,}01$	$\frac{168}{5377} = 0{,}03$	$\frac{469}{5377} = 0{,}09$	$\frac{725}{5377} = 0{,}13$	$\frac{1542}{5377} = 0{,}29$	$\frac{2421}{5377} = 0{,}45$

Fontos, tipikus hiba szokott lenni: P_0-t és P_1-et nem a tesztminta alapján határozzuk meg, adott a szövegben! P_1= 0,035 (3,5%), P_0=1-P_1=0,965.

Az egyenlőtlenség bal oldalán kiszámítható várható (átlagos) költségnövekedés:
$$P_1 \cdot (C_{01} - C_{11}) \cdot P(z|1) = 0{,}035 * (213.000 + 2.300 - 12.300 - 2.300) * 0{,}05 = 351{,}23$$
(belevettük a szűrés költségét, 2.300 Ft-ot mindkét esetbe).
Az egyenlőtlenség jobb oldalán kiszámítható várható költségnövekedés pedig:
$$P_0 \cdot (C_{10} - C_{00}) \cdot P(z|0) = 0{,}965 * (5.000 + 2.300 - 2.300) * 0{,}09 = 434.25$$
Tehát nagyobb átlagos költségnövekedést okoz, ha betegnek vesszük az ilyen paramétert produkáló embereket, tehát az a jó döntés, ha egészségesnek tekintjük őket!

B. Önmagában a szűrés költsége mennyiben befolyásolja, hogy az A. pontban vizsgált csoportot egészségesnek vagy betegnek célszerű-e tekinteni?

Látható a fenti számításból is: nincs hatása a döntésünkre. Hiszen mindenkinél jelentkezik, aki a szűrésben részt vesz, akárhogy is döntünk arról, hogy beteg-e az illető agy egészséges.

Persze az, hogy egyáltalán érdemes-e szűrnünk, az függ a szűrés költségétől, és persze az összköltség vagy a egy főre jutó várható költség is.

5. Egy diagnosztikai eljárást vizsgáltunk meg 3.131 személy segítségével, akik közül 1.883 beteg volt. A diagnosztika során egy z paramétert mértünk, ennek egy adott határ alatti értékénél tekintjük betegnek a pácienst. Azt tapasztaltuk, hogy a következő tartományokba esnek a z értékek a táblázatban megadott számú betegnél és egészségesnél.

$[z_{jmin}, z_{jmax})$	[-3,-2)	[-2,-1)	[-1,0)	[0,+1)	[+1,+2)	[+2,+3)
Betegek, akiknek [*zjmin, zjmax*) – ba esik a mért z paramétere	874	538	211	134	67	59
Egészségesek, akiknek [*zjmin, zmax)* – ba esik a mért z paramétere	23	68	169	225	342	421

Rajzolja fel a mért z alapján a ROC görbét (Receiving Operating Characteristic görbe)!

Megoldás:

A mintánkban összesen 1883 a beteg, és 3131-1883=1248 a egészséges.

$[z_{jmin}, z_{jmax})$	[-3,-2)	[-2,-1)	[-1,0)	[0,+1)	[+1,+2)	[+2,+3)
Betegek, akiknek [*zjmin, zjmax*) – ba esik a mért z paramétere	**874**	**538**	**211**	**134**	**67**	**59**
Akiket a betegek közül betegnek diagnosztizálunk, mert *zjmax*-nál kisebb értéket produkálnak (TP)	**874**	874+538= **1412**	874+538 +211= **1623**	ugyanígy: **1757**	ugyanígy: **1824**	ugyanígy: **1883**
Akiket a betegek közül egészségesnek diagnosztizálunk, mert *zjmax*-nál nagyobb értéket produkálnak (FN)	1883-874 = **1009**	1883-1412 = **471**	1883-1623 = **260**	1883-1757 = **126**	1883-1824 = **59**	**0**
Egészségesek, akiknek [*zjmin, zmax)* – ba esik a mért z paramétere	23	68	169	225	342	421
Akiket az egészségesek közül betegnek diagnosztizálunk, mert *zjmax*-nál kisebb értéket produkálnak (FP)	**23**	23+68= **91**	23+68+ 169=**260**	ugyanígy: **485**	ugyanígy: **827**	ugyanígy: **1248**
Akiket az egészségesek közül egészségesnek diagnosztizálunk, mert *zjmax*-nál nagyobb értéket produkálnak (TN)	1248-23 =**1225**	1248- 91 =**1157**	1248-260 = **988**	1248-485 = **763**	1248- 827 = **421**	**0**

Megjegyzés: Az egészségesnek *diagnosztizáltakat* (harmadik és hatodik sor) úgy is számolhattuk volna, hogy a z_{jmax} fölé eső megfelelő számokat összegezzük. Pl. a betegek közül, ha z_{jmax} =0, akkor e fölé eső betegcsoportok létszámösszege: 134+67+59=260, ami természetesen ugyanaz, mint amit az összes betegszámból való kivonással kaptunk. Másik példa: az egészségesek közül, ha z_{jmax} =-1, akkor az e fölé eső egészséges csoportok létszámának összege: 169+225+342+421=1157, ami természetesen ugyanaz, mint amit az egészségesek összlétszámából való kivonással kaptunk.

Ezek után a keresett szenzitivitás és specificitás értékek:

z_{jmax}	-3	-2	-1	0	+1	+2	+3
Akiket a betegek közül betegnek diagnosztizálunk, mert *zjmax*-nál kisebb értéket produkálnak (TP)	**0**	**874**	**1412**	**1623**	**1757**	**1824**	**1883**
SZ=TP/(TP+FN)	**0**	874/1883 = **0,46**	1412/1883 = **0,75**	1623/1883 = **0,86**	1757/1883 = **0,93**	1824/1883 = **0,97**	1883/1883 = **1**
Akiket az egészségesek közül betegnek diagnoszti-zálunk, mert *zjmax*-nál kisebb értéket produkálnak (FP)	**0**	**23**	**91**	**260**	**485**	**827**	**1248**
1-SP=FP/(FP+TN)	**0**	23/1248= **0,018**	91/1248 = **0.073**	260/1248 = **0,21**	485/1248 = **0,39**	827/1248 = **0,66**	1248/1248 = **1**

Ennek alapján a ROC görbe (kiegészítve az elejét mindkét jellemzőnél 0-val, ha z_{jmax}=-3, akkor mind a TP=0 és FP=0, mert senkit se tartunk betegnek):