

Adatbányászati technikák 2. zárthelyire gyakorló feladatok 2017. április 27.

1. Az alábbi táblázat arról tartalmaz adatokat, hogy adott korú, súlyú, sportolási státuszú emberek érdeklődtek-e egy adott termék iránt.

	kor	súly	sportol	érdekli
1.	fiatal	alacsony	igen	igen
2.	idős	közepes	nem	nem
3.	középkorú	magas	nem	igen
4.	idős	közepes	igen	nem
5.	fiatal	magas	nem	igen
6.	középkorú	alacsony	nem	nem
7.	idős	alacsony	nem	nem
8.	fiatal	közepes	nem	igen
9.	középkorú	magas	igen	igen
10.	idős	közepes	igen	nem

(a) Bayes osztályozót használva milyen címkét kap egy fiatal, közepes súlyú, sportoló ember?

(b) Mi a döntés, ha Laplace módszert használunk a Bayes eljáráson belül?

2. Bayes osztályozót szeretnék használni egy olyan helyzetben, ahol az A, B, C, D attribútumok alapján akarom jósolni a célváltozó értékét. Az A attribútum folytonos, a tanítóhalmaz $+$ címkéjű soraiban 10, 60, 70, 80, 130, 150 és 200 értékek vannak az A oszlopban, a $-$ címkéjű sorokban pedig 40, 50 és 60.

Hogyan kell kiszámolni az $A=100, B=1, C=0, D=3$ sor címkéjét, ha $P(B=1|+) = 0.3, P(B=1|-) = 0.2, P(C=0|+) = 0.7, P(C=0|-) = 0.2, P(D=3|+) = 0.5, P(D=3|-) = 0.5, P(+)=0.7, P(-)=0.3$?

3. Készítsen perceptront az $A \text{ OR } \text{NOT } B$ függvényhez vagy mutassa meg, hogy ez nem lehetséges. Ez utóbbi esetben készítsen neurális hálózatot.

4. Az alábbi táblázattal adott függvényhez készítsen perceptront vagy mutassa meg, hogy ez nem lehetséges. Ez utóbbi esetben készítsen neurális hálózatot a függvény megvalósítására.

A	B	C	címke
0	0	0	1
0	0	1	0
0	1	0	0
1	0	0	1
0	1	1	1
1	0	1	0
1	1	0	0
1	1	1	0

5. A 2, 3, 4, 7, 8, 10, 12, 14, 20 számokat akarjuk Kmeans algoritmussal három klaszterbe sorolni.
 (a) Írja le, hogy hogyan fut le az algoritmus, ha a 2, 7, 10 kezdő-centroidokat használjuk!
 (b) Írja le, hogy hogyan fut le az algoritmus, ha a 2, 3, 4 kezdő-centroidokat használjuk!

6. Az alábbi mátrix alapján rajzolja fel dendrogramon, hogy a MIN (single link) hierarchikus klaszterező algoritmus hogyan klaszterez. A dendrogram elkészítésének és a hasonlóságmátrix változásának minden fázisát mutassa be.

	p_1	p_2	p_3	p_4	p_5	p_6
p_1	1	0.8	0.3	0.2	0.9	0.1
p_2	0.8	1	0.2	0.25	0.7	0.15
p_3	0.3	0.2	1	0.8	0.1	0.05
p_4	0.2	0.25	0.8	1	0.05	0.6
p_5	0.9	0.7	0.1	0.05	1	0.7
p_6	0.1	0.15	0.05	0.6	0.7	1

7. (a) Az előző mátrix alapján futtassa le a MAX (complete link) hierarchikus klaszterező algoritmust.
 (b) Hogyan néz ki a proximity mátrix az első kör után a fenti mátrixból kiindulva, ha a GROUP AVERAGE módszert használjuk?