

Adatbányászati technikák 2. zárthelyire gyakorló feladatok 2016. május 9.

1. Osztályozza kNN osztályozóval $k = 5$ értéket használva a következő tanítóhalmaz segítségével a 0.9 értékű pontot.

(a) Használjon többségi szavazást!

(b) Használja a távolságokat is figyelembe vevő módszert a címke meghatározására!

0.1(-), 0.2(+), 0.4(+), 0.6(+), 0.8(-), 1.1(-), 1.7(+), 1.9(+),

2. Készítsen perceptront az A OR NOT B függvényhez vagy mutassa meg, hogy ez nem lehetséges. Ez utóbbi esetben készítsen neurális hálózatot.

3. 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

4. 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!

5. 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

6. (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?

7. Tekintsük az alábbi 15 tranzakciót az 123456 elemhalmazon: 123, 13456, 1235, 146, 136, 1234, 2346, 2456, 246, 346, 356, 456, 46, 5, 56. Az Apriori algoritmussal határozza meg a gyakori elemhalmazokat, $min_sup = 0.2$ küszöbvel! Mutassa be, hogy hogyan alakulnak a C_k és F_k halmazok az algoritmus futása során.
8. Az előző feladat tranzakcióit használva határozza meg az órán tanult algoritmussal, hogy melyek a 246 gyakori elemhalmazból kapható nagy megbízhatóságú szabályok, ha $min_conf = 0.7$.