

Istraživanje podataka - zadaci za vežbe 10

1. Algoritmom K-sredina identifikovati 3 klastera u sledećim podacima. Pri tom, koristiti euklidsko rastojanje. Za polazne centroide uzeti prve tri instance. Kolika je kvadratna greška ovog klasterovanja?

X	Y	Z
1	0	2
2	0	0
-3	-1	1
-4	-2	2
0	4	9
1	5	9

2. Data je matrica sličnosti skupa podataka. Izvršiti hijerarhijsko klasterovanje korišćenjem min i max veze. Rezultat prikazati dendrogramom.

	p1	p2	p3	p4	p5
p1	1.00	0.10	0.41	0.55	0.35
p2	0.10	1.00	0.64	0.47	0.98
p3	0.41	0.64	1.00	0.44	0.85
p4	0.55	0.47	0.44	1.00	0.76
p5	0.35	0.98	0.85	0.76	1.00

3. Klasterovati predmete prema broju studenata koji su upisali predmet, broju studenata koji su ga položili i prosečnoj oceni.
4. Izvršiti klasterovanje studenata koji studiraju o trošku budžeta. Kriterijum za klasterovanje su prosečna ocena na položenim obaveznim predmetima u toku studija i prosečna ocena na položenim uslovnim predmetima u toku studija. Prosečna ocena na položenim obaveznim predmetima se deli u 7 mogućih intervala, a prosečna ocena na položenim uslovnim predmetima se deli na 6 mogućih intervala. Formirati 5 klastera u kojima kao kriterijumi klasterovanja učestvuju obe prosečne ocene. Odrediti naziv smera koji je dominantan u klasteru kome pripada najveći broj studenata. Klasterovanje izvršiti proizvoljnom metodom.