

Основи роботи з сучасними програмними комплексами

Лабораторна робота № __ «Кластерний аналіз у ППП STATISTICA»

Мета роботи: дослідити процедури класифікування методами кластерного аналізу.

Робоче завдання: виконати розрахунки згідно індивідуального завдання і набути вмінь кластерного аналізу сукупності об'єктів.

Завдання на підготовку. Студент повинен мати уявлення про характеристики кластеру і родини методів кластерного аналізу. Розглянути типи відстаней між кластерами та базові алгоритми кластеризування. Розглянути оцінювання якості кластеризування та ступеню впливу ознак.

Хід виконання роботи

Згідно індивідуального завдання виконати кластеризування об'єктів за двома ознаками у середовищі STATISTICA.

Задача 1. Виконати кластеризування об'єктів методом «близького сусіди» для вихідних і нормованих даних.

Задача 2. Виконати кластеризування об'єктів методом «далекого сусіди» для вихідних і нормованих даних. Тип відстані використати такий самий як і для задачі 1.

Задача 3. Виконати кластеризування об'єктів методом «близького сусіди» або «далекого сусіди» (для вихідних або нормованих даних) з використанням іншого типу відстані ніж у задачах 1 та 2.

Задача 4. Виконати кластеризування об'єктів методом k-середніх для заданої кількості кластерів – двох, трьох і чотирьох. За допомогою критерію якості кластеризування визначити найкращу кількість кластерів для заданого набору об'єктів. Для цієї визначеної кількості кластерів (тобто для найякіснішої кластеризації) виконати оцінювання ступеню впливу ознак.

Продемонструвати розрахунки викладачу та оформити протокол лабораторної роботи.

Оформлення звіту та порядок його подання

Протокол лабораторної роботи повинен містити

1. титульну сторінку;
2. загальне завдання;
3. індивідуальне завдання;
4. короткі теоретичні відомості (2-3 стор.);
5. хід розв'язування задач 1-4, у тому числі:
 - для агломеративних методів –
дендрограми та матриці відстаней
(вказати використаний тип відстані, зробити висновки);
 - для методу k-середніх –
 - ◆ для заданої кількості кластерів: описові статистики для кожного кластеру, значення середніх, відстані, результати дисперсійного аналізу, члени кластерів з відстанями до центру кластеру;
 - ◆ хід оцінювання якості кластеризування,
 - ◆ хід оцінювання ступеню впливу ознак для найякіснішої кластеризації (зробити висновки).
6. загальні висновки.

Повний та охайний протокол подається викладачу для захисту виконаної лабораторної роботи.

Орієнтовний перелік контрольних запитань

1. Математичні характеристики кластеру.
2. Родини методів кластерного аналізу.
3. Який об'єкт називають спірним?
4. Як визначити евклідову та квадратичну евклідову відстані між двома об'єктами?
5. Як визначити відстань міських кварталів між двома об'єктами?
6. Коротко опишіть хід кластеризування за алгоритмом «близького сусіди».
7. Коротко опишіть хід кластеризування за алгоритмом «далекого сусіди».
8. Коротко опишіть хід кластеризування за методом k-середніх.
9. Критерій якості кластеризування.
10. Як визначити ступінь впливу ознак на результати кластеризування?

Варіант 1		Варіант 2		Варіант 3		Варіант 4	
X	Y	X	Y	X	Y	X	Y
3,73	133,23	6,38	265,90	12,30	399,30	15,88	528,15
4,30	140,96	8,80	272,43	11,05	416,11	14,55	528,12
3,96	137,78	6,37	270,34	10,02	409,00	12,55	543,60
3,80	140,82	8,57	278,92	9,69	411,99	15,79	541,52
4,36	140,25	8,84	261,10	11,48	401,79	16,88	557,29
3,03	131,28	6,88	270,62	9,90	406,69	14,97	542,33
7,93	176,99	8,65	276,20	23,18	523,98	14,95	535,22
8,84	171,38	6,74	265,10	24,27	512,67	15,35	537,03
7,09	174,56	8,51	263,34	22,23	531,43	14,61	551,98
8,92	176,37	8,70	277,14	24,36	533,84	15,07	529,69
8,37	177,67	6,51	272,82	22,13	538,64	13,47	555,18
11,14	217,62	8,87	275,09	34,36	659,12	14,07	542,36
12,75	216,95	7,27	275,30	36,85	648,36	16,65	548,32
11,73	211,58	8,67	267,79	33,10	655,53	12,54	532,39
11,54	212,37	22,99	431,91	36,43	644,07	46,78	865,54
12,60	214,06	23,02	421,10	36,53	651,69	48,95	851,14
12,34	212,67	22,80	423,12	33,46	635,65	48,29	873,93
11,06	220,81	22,54	429,84	33,69	640,74	45,84	862,84
11,78	218,59	23,25	426,73	35,53	656,88	45,09	847,88
12,81	213,65	22,81	423,15	36,67	647,26	48,34	861,28

Варіант 5		Варіант 6		Варіант 7		Варіант 8	
X	Y	X	Y	X	Y	X	Y
19,56	697,41	22,98	826,23	25,55	979,20	28,04	1110,16
18,45	673,20	21,47	782,52	28,58	958,40	31,44	1096,23
20,88	659,44	22,13	798,31	27,30	911,64	27,89	1086,36
20,94	659,06	24,04	793,15	28,30	929,92	25,75	1079,05
15,33	655,09	18,05	780,37	24,85	978,76	29,15	1084,85
16,65	676,94	19,69	816,61	27,90	956,75	29,18	1043,70
35,22	890,90	23,83	797,84	51,41	1196,18	26,65	1063,17
35,79	894,73	22,84	815,89	49,39	1248,90	30,48	1094,96
39,40	864,88	18,11	803,02	54,41	1253,40	29,79	1049,65
35,52	880,62	24,09	818,81	50,68	1204,55	25,46	1071,75
40,41	873,16	19,32	813,11	52,62	1199,26	24,26	1104,10
58,87	1084,98	20,37	795,62	78,96	1486,14	29,07	1094,74
55,05	1071,95	19,67	783,99	82,51	1514,98	30,66	1091,32
60,88	1053,87	24,73	815,49	82,21	1530,23	30,41	1069,35
58,44	1089,58	71,30	1267,34	78,94	1537,27	89,11	1758,09
56,91	1087,94	70,57	1284,42	84,73	1527,70	92,83	1720,39
56,19	1068,54	68,38	1287,80	79,02	1523,03	92,13	1686,27
57,71	1056,24	68,31	1288,95	82,41	1533,21	91,69	1716,97
57,15	1089,39	69,50	1268,64	77,56	1483,57	92,52	1760,72
60,93	1064,46	67,45	1268,65	83,19	1477,27	93,96	1691,69

Варіант 9		Варіант 10		Варіант 11		Варіант 12	
X	Y	X	Y	X	Y	X	Y
32,69	1178,51	35,60	1314,96	41,32	1482,40	38,81	1680,03
27,60	1252,32	36,55	1324,26	41,63	1430,08	39,20	1603,43
33,81	1242,42	38,27	1312,08	35,40	1501,89	48,12	1669,03
27,75	1255,92	32,82	1389,19	44,62	1478,86	38,02	1652,01
36,47	1199,39	33,22	1333,02	38,03	1519,85	44,61	1629,21
31,14	1248,49	35,72	1348,31	33,83	1536,09	45,44	1600,89
65,23	1615,70	31,82	1330,74	84,32	1947,02	42,32	1583,89
72,25	1569,97	32,18	1341,28	86,73	1872,53	38,05	1593,61
69,54	1566,52	38,93	1331,63	87,85	1933,72	44,33	1592,23
70,34	1554,12	31,58	1373,42	78,77	1875,70	37,35	1633,68
63,46	1538,31	34,10	1333,81	86,43	1909,36	41,63	1616,01
100,10	1936,07	30,47	1342,70	132,65	2420,98	43,16	1603,79
107,14	1936,09	40,16	1382,75	129,58	2348,24	46,99	1578,55
100,66	1959,24	31,03	1305,17	127,38	2373,12	44,54	1678,61
107,98	1961,57	114,96	2144,71	132,62	2350,70	136,56	2597,95
104,08	1924,45	111,60	2197,73	127,28	2361,72	134,92	2632,57
103,88	1970,96	110,06	2108,58	125,12	2375,21	141,49	2587,83
108,89	1910,70	118,19	2138,48	131,71	2386,68	143,19	2551,83
108,24	1915,31	111,30	2181,59	125,65	2412,77	133,99	2530,57
107,03	1913,27	116,38	2114,91	130,15	2390,41	133,58	2636,07

Варіант 13		Варіант 14		Варіант 15		Варіант 16	
X	Y	X	Y	X	Y	X	Y
43,90	1718,62	45,33	1942,46	48,95	2085,58	60,56	2086,99
50,39	1748,62	56,47	1890,25	54,90	2041,21	55,36	2120,02
48,82	1697,01	46,00	1887,87	45,48	2036,88	60,80	2086,51
46,18	1771,74	45,57	1891,92	47,29	2017,70	64,04	2222,88
50,69	1727,75	46,19	1921,58	47,56	1988,66	57,24	2108,98
46,85	1732,09	48,43	1935,13	50,30	2039,57	63,95	2216,80
97,55	2315,64	45,69	1876,74	108,79	2614,66	63,73	2180,92
96,44	2247,90	45,48	1885,07	106,10	2581,70	55,64	2088,87
100,02	2282,10	49,00	1829,34	116,01	2604,52	62,23	2166,45
99,81	2322,44	49,11	1931,71	119,97	2667,19	58,28	2140,78
91,76	2278,94	47,65	1954,08	106,50	2609,73	63,09	2121,64
144,43	2756,05	51,86	1860,78	170,08	3250,95	51,03	2090,77
148,68	2757,40	51,60	1920,95	170,04	3172,41	64,12	2156,45
154,39	2791,18	55,93	1931,91	168,33	3245,15	52,39	2084,69
144,86	2762,29	158,66	2974,86	179,96	3219,24	177,03	3499,93
154,38	2764,81	163,00	3037,33	170,71	3214,49	191,14	3436,04
151,73	2793,85	155,45	3041,88	167,82	3203,34	176,83	3399,36
143,23	2736,31	160,47	2963,74	178,15	3276,70	189,56	3373,37
156,80	2766,76	163,32	3007,53	175,83	3239,63	182,46	3384,58
154,28	2792,94	163,70	3027,97	180,03	3155,87	188,69	3499,40

Варіант 17		Варіант 18		Варіант 19		Варіант 20	
X	Y	X	Y	X	Y	X	Y
68,20	2210,68	58,95	2502,49	62,71	2516,29	70,89	2763,77
55,01	2380,66	66,36	2404,72	58,57	2544,39	61,01	2773,90
61,14	2333,34	58,07	2471,79	75,85	2588,02	73,73	2649,62
53,19	2354,63	54,50	2450,57	63,00	2513,76	75,98	2706,00
60,49	2218,42	63,23	2350,30	69,08	2587,68	72,77	2616,19
61,23	2242,29	58,21	2398,02	59,33	2630,93	73,83	2743,43
125,35	3024,48	69,34	2495,11	138,09	3231,43	77,98	2678,96
120,03	2946,36	54,44	2379,68	148,38	3382,49	71,69	2658,18
135,89	3024,33	66,88	2343,88	137,95	3278,94	70,51	2628,07
120,75	2955,38	61,53	2415,84	134,89	3412,11	76,16	2674,76
120,83	2924,02	72,30	2430,99	138,64	3260,57	77,62	2635,69
195,35	3621,55	71,17	2497,20	221,71	4133,08	76,85	2706,08
187,61	3678,24	61,11	2517,19	226,10	4103,67	62,60	2662,35
198,23	3736,48	64,67	2352,82	221,00	4137,35	75,65	2643,75
199,97	3662,00	200,04	3859,20	222,54	4163,51	236,84	4250,79
189,65	3697,30	211,55	3867,75	220,07	4141,64	230,19	4381,03
201,54	3737,48	200,39	3880,89	225,51	4118,19	230,13	4291,92
190,60	3653,27	211,58	3915,57	226,77	3995,86	236,73	4312,71
189,54	3615,13	207,43	3925,96	222,56	4050,16	240,95	4250,93
194,59	3732,19	210,80	3900,55	218,88	4056,07	220,09	4258,89