

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
Кафедра штучного інтелекту

Звіт з виконання завдань
комп'ютерного практикуму № 5
Кореляційно-регресійний аналіз у середовищі STATISTICA
з кредитного модуля «Багатовимірний статистичний аналіз»

Звіт склав
студент гр. КІ-01
Копцов В.О.
Прийняла: *Ірина Джигирей*

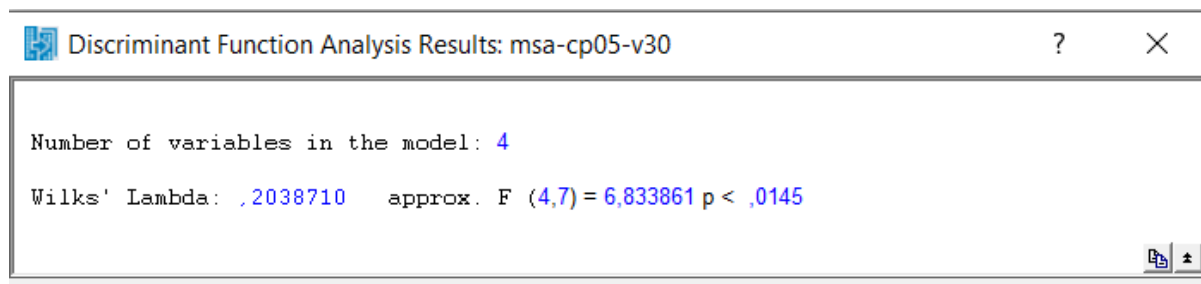
Мета роботи. Опанувати процедури покрокового дискримінантного аналізу, засвоїти поняття класу і дискримінантної функції, набуті досвіду застосування дискримінантного аналізу для пошуку змінних, які дають змогу віднести аналізовані об'єкти до певної групи, інтерпретувати функції дискримінації. Опанувати можливості використання дискримінантного аналізу для класифікування спостережень, набуті досвіду проведення дискримінантного аналізу для класифікування об'єктів.

Завдання. За даними індивідуального завдання (Додаток А) визначити приналежність об'єктів до одного з класів за допомогою дискримінантного аналізу. Розрахунки провести у середовищі STATISTICA.

Варіант №30

Хід виконання завдань практикуму

1.



Оскільки лямбда Уїлкса близька до нуля, можна сказати, що якість моделі висока. В модель увійшли всі 4 змінні, що означає, що кожна змінна якимось чином впливає на клас об'єкту.

Discriminant Function Analysis Summary (msa-cp05-v30)						
No. of vars in model: 4; Grouping: x5 (2 grps)						
Wilks' Lambda: .20387 approx. F (4,7)=6,8339 p< .0145						
N=12	Wilks' Lambda	Partial Lambda	F-remove (1,7)	p-value	Toler.	1-Toler. (R-Sqr.)
x1	0,217682	0,936553	0,47422	0,513223	0,013331	0,986669
x2	0,513691	0,396875	10,63781	0,013834	0,040830	0,959170
x3	0,600091	0,339733	13,60440	0,007773	0,022088	0,977912
x4	0,258887	0,787491	1,88899	0,211701	0,012490	0,987510

Для другої і третьої змінної лямбда Уїлкса більша відносно першої і третьої. Високе значення лямбди Уїлкса означає, що центроїди груп по цій змінній співпадають або по ній відсутні групові розходження. Це позначає, що “найбільше” на класифікацію вплинули перша і четверта змінні.

Відстані між групами (Квадратична відстань Махаланобіса)

Squared Mahalanobis Distances (msa-cp05-v30)				
x5	G_1:0	G_2:1		
G_1:0	0,00000	13,38879		
G_2:1	13,38879	0,00000		

Класифікаційна матриця

Classification Matrix (msa-cp05-v30)			
Rows: Observed classifications			
Columns: Predicted classifications			
Group	Percent Correct	G_1:0 p=,41667	G_2:1 p=,58333
G_1:0	100,0000	5	0
G_2:1	100,0000	0	7
Total	100,0000	5	7

Результати апостеріорної класифікації кожного об'єкту

Classification of Cases (msa-cp05-v30)			
Incorrect classifications are marked with *			
Case	Observed Classif.	1 p=,41667	2 p=,58333
1	G_2:1	G_2:1	G_1:0
2	G_2:1	G_2:1	G_1:0
3	G_2:1	G_2:1	G_1:0
4	G_2:1	G_2:1	G_1:0
5	G_2:1	G_2:1	G_1:0
6	G_2:1	G_2:1	G_1:0
7	G_2:1	G_2:1	G_1:0
8	G_1:0	G_1:0	G_2:1
9	G_1:0	G_1:0	G_2:1
10	G_1:0	G_1:0	G_2:1
11	G_1:0	G_1:0	G_2:1
12	G_1:0	G_1:0	G_2:1

Отримали 2 групи. До класу G_1 відносяться об'єкти зі значенням групувальної функції 0, тобто об'єкти під номерами 8-12. До класу G_2 відносяться об'єкти зі значенням групувальної функції 1, тобто об'єкти під номерами 1-7. По матриці видно, що результат співпадає з значеннями групувальної змінної.

Відстані до центрів класів.

Squared Mahalanobis Distances from Group Centroids (msa-cp05-v30)				
Incorrect classifications are marked with *				
Case	Observed Classif.	G_1:0 p=,41667	G_2:1 p=,58333	
1	G_2:1	29,01011	4,93636	
2	G_2:1	17,65618	5,63156	
3	G_2:1	6,18926	1,77928	
4	G_2:1	18,00398	4,25751	
5	G_2:1	7,97580	1,33038	
6	G_2:1	15,80374	1,53846	
7	G_2:1	26,82837	8,27237	
8	G_1:0	2,99163	26,68920	
9	G_1:0	1,81944	6,36729	
10	G_1:0	3,38111	13,61067	
11	G_1:0	2,64294	22,79307	
12	G_1:0	1,41897	9,73779	

Тут напівжирним виділено відстань об'єктів до “свого” класу. Одразу можна зрозуміти, що об'єкти 9, 3, 5, 12 знаходяться між двома класами оскільки недалеко від центрів як і “свого” так і “чужого” класу, а об'єкт 7, хоч і знаходиться відносно далеко від центру свого класу, але знаходиться на ще більшій відстані від центру G_1.

Значення апостеріорних ймовірностей.

Posterior Probabilities (msa-cp05-v30)			
Incorrect classifications are marked with *			
Case	Observed Classif.	G_1:0 p=,41667	G_2:1 p=,58333
1	G_2:1	0,000004	0,999996
2	G_2:1	0,001746	0,998254
3	G_2:1	0,073002	0,926998
4	G_2:1	0,000739	0,999261
5	G_2:1	0,025107	0,974893
6	G_2:1	0,000570	0,999430
7	G_2:1	0,000067	0,999933
8	G_1:0	0,999990	0,000010
9	G_1:0	0,874072	0,125928
10	G_1:0	0,991660	0,008340
11	G_1:0	0,999941	0,000059
12	G_1:0	0,978604	0,021396

Тут напівжирним виділено апостеріорну ймовірність, що об'єкт належить до “свого” класу.

Видно, що для 1-7 вірогідність більша для G_2, для 8-12 для G_1. Об'єкти помічені в попередньому абзаці 9, 3, 5, 12 мають незначно меншу ймовірність що вони належать “своєму” класу, але все одно дуже велику.

2. Класифікаційні функції

Variable	Classification Functions; grouping: x5 (msa-cp05-v30)			
	G_1:0 p=,41667	G_2:1 p=,58333		
x1	-0,03684	-0,06371		
x2	-0,10624	0,09286		
x3	0,07657	-0,04519		
x4	0,38315	1,33056		
Constant	-4,22633	-5,41276		

$$Q_{k(G1)} = -0,03684x_1 - 0,10624x_2 + 0,07657x_3 - 0,38315x_4 - 4,22633$$

$$Q_{k(G2)} = -0,06371x_1 + 0,09286x_2 - 0,04519x_3 + 1,33056x_4 - 5,41276$$

Об'єкт 13

13	G_1:0	G_2:0
119,32	-4,3957488	-7,6018772
77,46	-8,2293504	7,1929356
157,06	12,0260842	-7,0975414
8,22	3,149493	10,9372032
	-4,22633	-5,41276
Qk	-1,675852	-1,9820398

Об'єкт складно однозначно віднести до одного з класів, але оскільки значення функції для G_1 більше за значення для G_2. Можна віднести об'єкт до класу G_1.

Об'єкт 14

14	G_1:0	G_2:0
67,84	-2,4992256	-4,3220864
83,04	-8,8221696	7,7110944
114,04	8,7320428	-5,1534676
8,76	3,356394	11,6557056
	-4,22633	-5,41276
Qk	-3,4592884	4,478486

Об'єкт однозначно відноситься до класу G_2.

Об'єкт 15

15	G_1:0	G_2:0
92,86	-3,4209624	-5,9161106
110,4	-11,728896	10,251744
188,38	14,4242566	-8,5128922
8,76	3,356394	11,6557056
	-4,22633	-5,41276
Qk	-1,5955378	2,0656868

Об'єкт однозначно відноситься до класу G_2.

Об'єкт 16

16	G_1:0	G_2:0
216,88	-7,9898592	-13,8174248
117,6	-12,493824	10,920336
330,94	25,3400758	-14,9551786
14,16	5,425404	18,8407296
	-4,22633	-5,41276
Qk	6,0554666	-4,4242978

Об'єкт однозначно відноситься до класу G_1.

Об'єкт 17

17	G_1:0	G_2:0
77,02	-2,8374168	-4,9069442
92,76	-9,8548224	8,6136936
393,04	30,0950728	-17,7614776
4,98	1,908087	6,6261888
	-4,22633	-5,41276
Qk	15,0845906	-12,8412994

Об'єкт однозначно відноситься до класу G_1.

Об'єкт 18

18	G_1:0	G_2:0
147,04	-5,4169536	-9,3679184
144,24	-15,3240576	13,3941264
150,22	11,5023454	-6,7884418
17,76	6,804744	23,6307456
	-4,22633	-5,41276
Qk	-6,6602518	15,4557518

Об'єкт однозначно відноситься до класу G_2.

Висновки:

В процесі виконання комп'ютерного практикуму №5 я виконав розрахунки згідно індивідуального завдання і набув вмінь дискримінантного аналізу в Statistica. Було дуже зручно отримувати результати дискримінування за допомогою додатку. Присутньо багато зручних функцій, але чомусь в практикумі вибір куди віднести об'єкт ми робили не за допомогою Statistica, тому, наскільки я розумію, такої функції там немає, але якщо б вона була це ще більше збільшила зручність додатку.