Факультет Біомедичної інженерії

Національного технічного університету України

«Київський політехнічний інститут»

Кафедра біомедичної кібернетики

Дисципліна

«Інтелектуальний аналіз даних»

***Лабораторна робота № 4***

**Тема:** Відпрацювання навичок застосування дискримінантного аналізу на прикладі класифікації прогнозу ускладнень після операції на серці

Виконав:

студент групи БС-81

Сєров О. В.

Перевірив:

викладач

доц. каф. БМК

Павлов В.А.

дата \_\_\_\_\_\_\_\_\_\_

підпис \_\_\_\_\_\_\_\_\_

Київ – 2020

**Мета роботи:** застосувати дискримінантний аналіз на навчальній та тестовій вибірках на даних про операцію на сердці.

**Завдання до роботи**

***Частина 1.*** Розробка лінійних класифікаторів прогнозу ускладнень після операції на серці. Відпрацювання застосування параметрів алгоритму для визначення моделей з кращими показниками моделі на навчальній вибірці даних

***Завдання***

*Для одержання класифікаторів (варіація параметрів) керуватися вказівками викладача на занятті та методичним посібником – додаток 4.*

Для вибраного критерію якості класифікації (один з п’яти вибраних вами варіантів) виконати:

1.Для одержаних даних (додаток 4) одержати класифікатор для 4 (3) класів ускладнень операції на серці для значень F за замовчуванням.

2.Знижувати F*вкл* до тих пір поки будуть покращуватися результати класифікації

1. *Виділити окремо навчальну та тестові вібірки однорідно по дисперсії у класах у співвідношенні 70/30-75/25-80/20 (один з варіантів)*
2. *На навчальній послідовності одержувати класифікатори, порівнюючи якість класифікації (чутливість у класах) на тестовій вибірці*
3. Навести найкращі моделі класифікації за найкращим значенням чутливості у класах на тесті.
4. ***Додаткові завдання 1***

Одержати класифікатори для тих же класів з принципу один проти всіх

та провести для них дослідження по п. 1-5

1. ***Додаткові завдання 2***

Провести балансування у класах за рахунок повторення (кратного чи дрібнократного) об’єктів у класах та провести дослідження для класифікаторів один проти всіх по п.1-5.

***Зробити висновки по проведених дослідженнях що до результатів класифікації.***

***Частина 2.*** Розробка нелінійних класифікаторів прогнозу ускладнень після операції на серці. Аналіз моделей різного рівня складності.

***Завдання.***

Розширити поле узагальнених змінних за рахунок функціонального перетворення первинних змінних: додати до первинних «х» - обернені 1/х, та ***прямі та обернені*** множення ***первинних х*** (та хто захоче кореня 2,3) ***другого порядку****. Для множини узагальнених змінних провести дослідження по як в частині 1. Лабораторної роботи.*

**Теоретичні відомості**

Дослідження можливості автоматизованого визначення розвитку післяопераційних ускладнень стану паціента та побудова їх статистичних моделей.

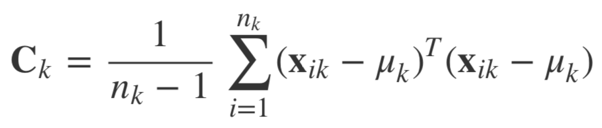
Операція: протезування мітрального клапану. Класифікація станів: норма – гостра серцева недостатність - гостра недостатність мозгового кровообігу - комбіноване ускладнення

Лінійний дискримінантний аналіз (Linear Discriminant Analysis, LDA) є розділом багатовимірного аналізу, який дозволяє оцінювати відмінності між двома і більше групами об'єктів по декількох змінним одночасно. Він реалізує дві тісно пов'язані між собою статистичні процедури:

* інтерпретацію міжгрупових відмінностей, коли потрібно відповісти на питання: наскільки добре використовується набір змінних в змозі сформувати розділяє поверхню для об'єктів навчальної вибірки і які з цих змінних найбільш інформативні?
* класифікацію, тобто передбачення значення группировочного фактора для екзаменованих групи спостережень.

В основі дискримінантного аналізу лежить припущення про те, що опису об'єктів кожного k-го класу представляють собою реалізації багатовимірної випадкової величини, розподіленої за нормальним законом  із середніми 

і ковариационной матрицею



(Індекс m вказує на розмірність простору ознак).

**Критерий отбора переменных “лямбда” Уилкса *Wilks' lambda***

это отношение разброса точек внутри класса от средних в классах (внутригрупповая дисперсия) к общему разбросу точек от общего среднего (общей дисперсии). Для записи формулы критерия определим что внутригрупповой расброс характеризует матрица ковариаций **** :  а межгрупповой расброс - матрица ****: . Матрицу полного расброса  можно вычислить

как или как 

Так как простым скалярным показателем расброса является определитель матрицы расброса то “лямбда” Уилкса” определяют как

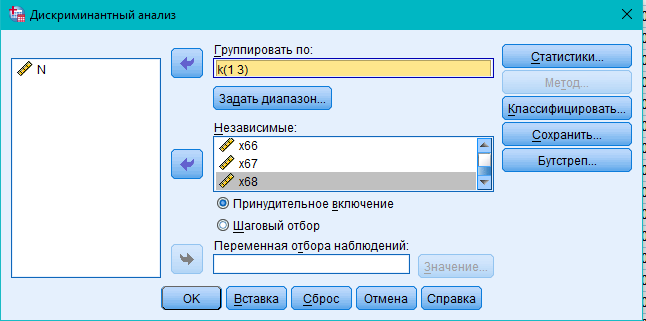
 Отбор переменных в шаговом дискриминантном анализе, для ввода в уравнение осуществляется на основании того, насколько они уменьшают значение "лямбда" Уилкса. На каждом шаге вводится переменная минимизирующая это значение или что то-же - максимизирующая соответствующий F-критерий. Кроме того, SPSS проверяет уже включенные в модель переменные; та из них, которая имеет слишком маленькое значение F исключения, исключается.

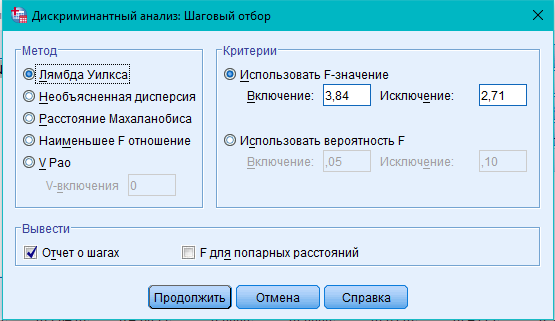
F-значение для изменения в *лямбде* *Уилкса* при включении переменной в модель, содержащую *р* независимых переменных, равно:

 где

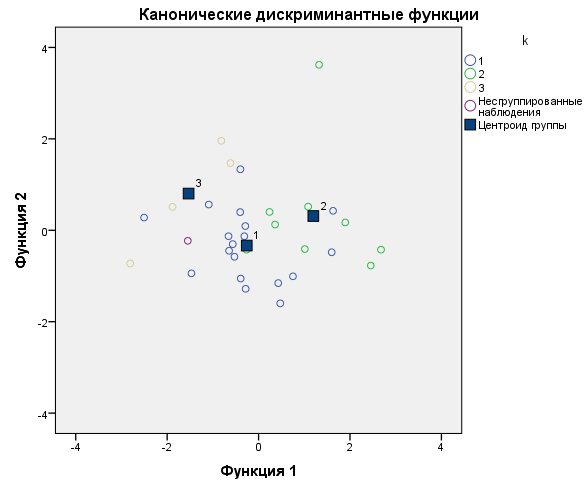
p- текущее значение количества переменных пространства Х,n— общее число наблюдений,   
К — число групп,  *— лямбда* *Уилкса* до включения новой переменной, *—лямбда Уилкса* после включения новой переменной.

**Розрахунки та результати**





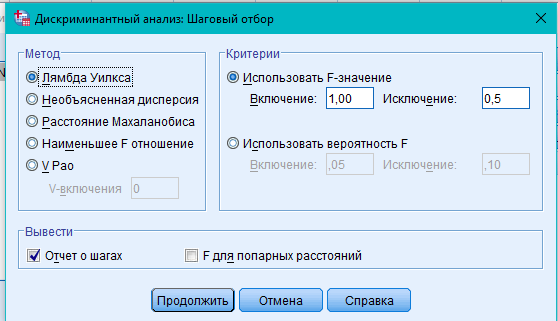
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Коэффициенты классифицирующей функции** | | | | | | |
|  | | k | | | | |
| 1,0000 | | 2,0000 | | 3,0000 |
| x35 | | 1,623 | | 3,034 | | 1,798 |
| x49 | | 38,500 | | 34,835 | | 43,208 |
| x54 | | ,274 | | ,378 | | ,066 |
| (Константа) | | -74,847 | | -70,040 | | -85,629 |
| Линейные дискриминантные функции Фишера | | | | | | |
| **Нормированные коэффициенты канонической дискриминантной функции** | | | | |
|  | Функция | | | |
| 1 | | 2 | |
| x35 | ,580 | | ,798 | |
| x49 | -,849 | | ,258 | |
| x54 | ,711 | | -,482 | |



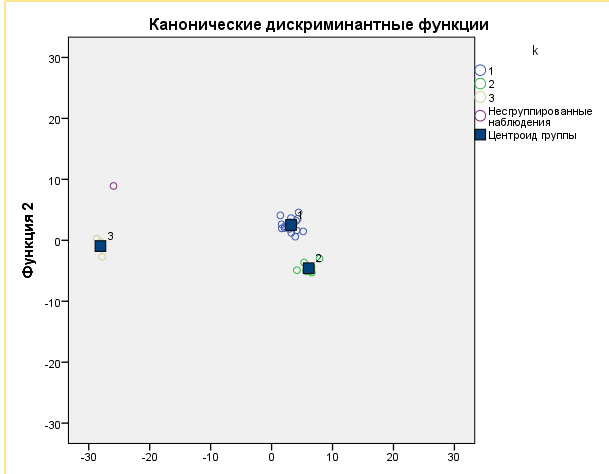
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Результаты классификацииa,c** | | | | | | |
|  | | k | Предсказанная принадлежность к группе | | | Итого |
| 1,0000 | 2,0000 | 3,0000 |
| Исходные | Частота | 1,0000 | 13 | 2 | 3 | 18 |
| 2,0000 | 3 | 6 | 0 | 9 |
| 3,0000 | 0 | 0 | 4 | 4 |
| Несгруппированные наблюдения | 0 | 0 | 1 | 1 |
| % | 1,0000 | 72,2 | 11,1 | 16,7 | 100,0 |
| 2,0000 | 33,3 | 66,7 | ,0 | 100,0 |
| 3,0000 | ,0 | ,0 | 100,0 | 100,0 |
| Несгруппированные наблюдения | ,0 | ,0 | 100,0 | 100,0 |
| Кросс-проверенныеb | Частота | 1,0000 | 13 | 2 | 3 | 18 |
| 2,0000 | 3 | 5 | 1 | 9 |
| 3,0000 | 1 | 0 | 3 | 4 |
| % | 1,0000 | 72,2 | 11,1 | 16,7 | 100,0 |
| 2,0000 | 33,3 | 55,6 | 11,1 | 100,0 |
| 3,0000 | 25,0 | ,0 | 75,0 | 100,0 |
|  | | | | | | |

За отриманими даними більшість (74.2%) первинних згрупованих спостережень класифіковано правильно. 67,7% перехресно-перевірених згрупованих спостережень класифіковано правильно.

Підвищуємо відсоток правильно класифікованих спостережень понижуючи Fвкл та Fвид.



|  |  |  |  |
| --- | --- | --- | --- |
| **Коэффициенты классифицирующей функции** | | | |
|  | k | | |
| 1,0000 | 2,0000 | 3,0000 |
| x0 | -1320,692 | -1350,920 | -1086,854 |
| x3 | 775,814 | 785,995 | 672,754 |
| x7 | 542,487 | 558,915 | 451,069 |
| x10 | 180,472 | 182,405 | 154,742 |
| x11 | -58,273 | -53,860 | -56,700 |
| x13 | 363,027 | 361,286 | 341,377 |
| x14 | 1038,010 | 1235,347 | 570,537 |
| x16 | -15,196 | -13,064 | -17,769 |
| x19 | 47529,475 | 46875,083 | 45035,345 |
| x21 | -78330,294 | -79191,175 | -65853,201 |
| x22 | -10097,321 | -10656,671 | -7955,506 |
| x24 | 115427,414 | 119275,196 | 94507,863 |
| x27 | -30131,662 | -30841,245 | -25174,697 |
| x35 | -2935,045 | -2994,157 | -2388,997 |
| x36 | 6129,615 | 5931,819 | 5554,157 |
| x45 | 1894,585 | 1901,430 | 1712,322 |
| x47 | -89,031 | -86,311 | -87,434 |
| x51 | -387,561 | -394,179 | -333,539 |
| x53 | 368,040 | 370,083 | 325,710 |
| x54 | -45,319 | -44,623 | -44,375 |
| x56 | -134,208 | -129,317 | -131,746 |
| x63 | -275,354 | -281,583 | -230,157 |
| (Константа) | -55693,697 | -55261,426 | -48802,407 |
| Линейные дискриминантные функции Фишера | | | |

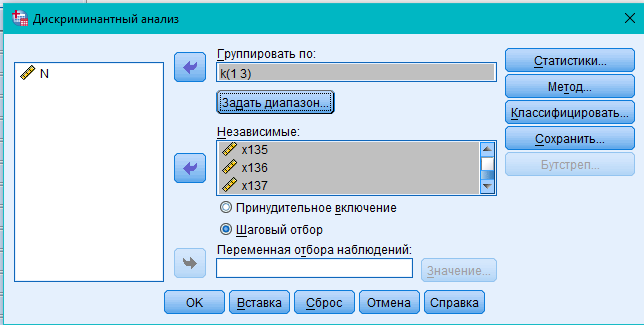


|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Результаты классификацииa,c** | | | | | | |
|  | | k | Предсказанная принадлежность к группе | | | Итого |
| 1,0000 | 2,0000 | 3,0000 |
| Исходные | Частота | 1,0000 | 18 | 0 | 0 | 18 |
| 2,0000 | 0 | 9 | 0 | 9 |
| 3,0000 | 0 | 0 | 4 | 4 |
| Несгруппированные наблюдения | 0 | 0 | 1 | 1 |
| % | 1,0000 | 100,0 | ,0 | ,0 | 100,0 |
| 2,0000 | ,0 | 100,0 | ,0 | 100,0 |
| 3,0000 | ,0 | ,0 | 100,0 | 100,0 |
| Несгруппированные наблюдения | ,0 | ,0 | 100,0 | 100,0 |
| Кросс-проверенныеb | Частота | 1,0000 | 13 | 5 | 0 | 18 |
| 2,0000 | 2 | 7 | 0 | 9 |
| 3,0000 | 0 | 0 | 4 | 4 |
| % | 1,0000 | 72,2 | 27,8 | ,0 | 100,0 |
| 2,0000 | 22,2 | 77,8 | ,0 | 100,0 |
| 3,0000 | ,0 | ,0 | 100,0 | 100,0 |

Тепер усі 100,0% вихідних згрупованих спостережень класифіковано правильно.

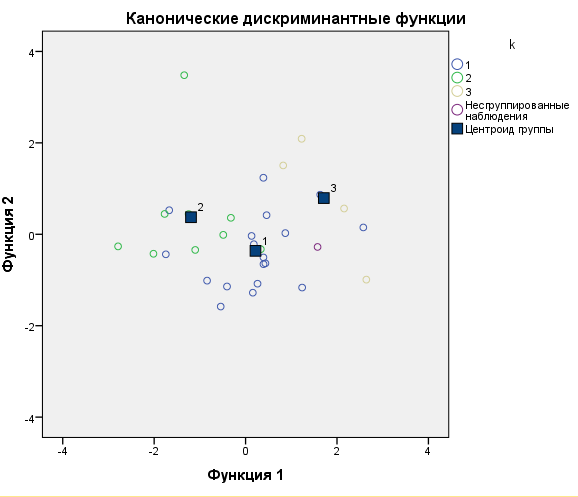
77,4% перехресно-перевірених згрупованих спостережень класифіковано правильно.

Добавимо обернені 1/х змінні – тобто розширимо поле узагальнених змінних.

**

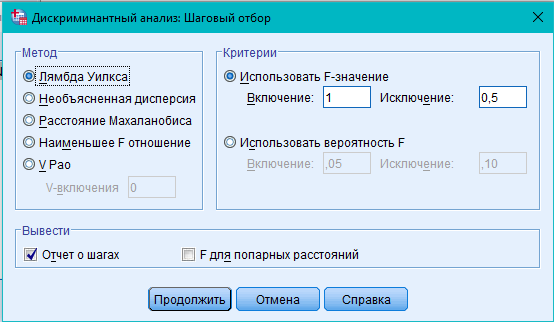
|  |  |  |  |
| --- | --- | --- | --- |
| **Коэффициенты классифицирующей функции** | | | |
|  | k | | |
| 1,0000 | 2,0000 | 3,0000 |
| x35 | -,755 | ,680 | -,680 |
| x49 | 50,555 | 46,961 | 55,372 |
| x123 | 1955,957 | 1814,592 | 2289,048 |
| (Константа) | -116,542 | -104,251 | -144,852 |
| Линейные дискриминантные функции Фишера | | | |

|  |  |  |
| --- | --- | --- |
| **Нормированные коэффициенты канонической дискриминантной функции** | | |
|  | Функция | |
| 1 | 2 |
| x35 | -,566 | ,796 |
| x49 | ,823 | ,152 |
| x123 | ,758 | ,480 |



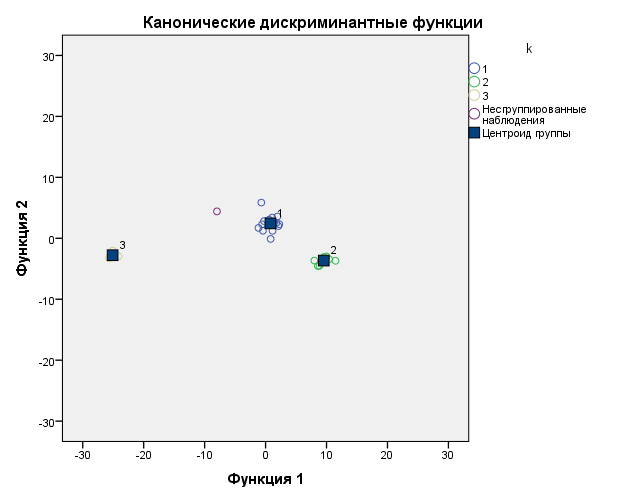
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Результаты классификацииa,c** | | | | | | |
|  | | k | Предсказанная принадлежность к группе | | | Итого |
| 1,0000 | 2,0000 | 3,0000 |
| Исходные | Частота | 1,0000 | 13 | 2 | 3 | 18 |
| 2,0000 | 2 | 7 | 0 | 9 |
| 3,0000 | 0 | 0 | 4 | 4 |
| Несгруппированные наблюдения | 0 | 0 | 1 | 1 |
| % | 1,0000 | 72,2 | 11,1 | 16,7 | 100,0 |
| 2,0000 | 22,2 | 77,8 | ,0 | 100,0 |
| 3,0000 | ,0 | ,0 | 100,0 | 100,0 |
| Несгруппированные наблюдения | ,0 | ,0 | 100,0 | 100,0 |
| Кросс-проверенныеb | Частота | 1,0000 | 13 | 2 | 3 | 18 |
| 2,0000 | 3 | 5 | 1 | 9 |
| 3,0000 | 1 | 0 | 3 | 4 |
| % | 1,0000 | 72,2 | 11,1 | 16,7 | 100,0 |
| 2,0000 | 33,3 | 55,6 | 11,1 | 100,0 |
| 3,0000 | 25,0 | ,0 | 75,0 | 100,0 |
| Тепер 77,4% вихідних згрупованих спостережень класифіковано правильно та 67,7% перехресно-перевірених згрупованих спостережень класифіковано правильно. | | | | | | |

Будемо понижати значення F, для збільшення кількості класифікованих спостережень.



|  |  |  |
| --- | --- | --- |
| **Нормированные коэффициенты канонической дискриминантной функции** | | |
|  | Функция | |
| 1 | 2 |
| x3 | 10,183 | 1,139 |
| x4 | -3,005 | -1,325 |
| x15 | 14,795 | ,598 |
| x16 | -29,790 | -1,367 |
| x28 | 3,409 | -,055 |
| x29 | -2,156 | ,421 |
| x35 | -19,582 | -,485 |
| x37 | 19,823 | -3,761 |
| x46 | -2,939 | -2,508 |
| x51 | -1,916 | 9,088 |
| x53 | 17,103 | 2,159 |
| x56 | 3,470 | -3,197 |
| x71 | -9,391 | -,395 |
| x80 | ,595 | 1,312 |
| x81 | 9,711 | -,178 |
| x85 | -30,419 | ,161 |
| x104 | -15,058 | 1,118 |
| x106 | 16,203 | -4,456 |
| x123 | -2,426 | ,381 |
| x133 | -,375 | -2,676 |
| x136 | 14,146 | 1,552 |
| x137 | 8,408 | 6,214 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Коэффициенты классифицирующей функции** | | | |
|  | k | | |
| 1,0000 | 2,0000 | 3,0000 |
| x3 | -412,821 | -381,381 | -516,152 |
| x4 | 334,786 | 314,429 | 429,737 |
| x15 | -126,575 | -115,556 | -160,446 |
| x16 | 626,975 | 587,937 | 747,578 |
| x28 | -,496 | -,457 | -,612 |
| x29 | 3,875 | 3,621 | 4,510 |
| x35 | 1963,861 | 1789,448 | 2492,355 |
| x37 | -255,938 | -234,506 | -309,861 |
| x46 | 314,106 | 307,771 | 368,590 |
| x51 | 334,767 | 326,086 | 335,021 |
| x53 | -237,476 | -221,082 | -292,106 |
| x56 | -95,026 | -88,742 | -104,245 |
| x71 | 974945,097 | 896578,131 | 1216208,738 |
| x80 | -737,707 | -755,190 | -875,887 |
| x81 | -109641,882 | -99270,077 | -139870,600 |
| x85 | 1261377,176 | 1180477,309 | 1499931,075 |
| x104 | 7893,862 | 7211,356 | 9787,500 |
| x106 | -275336,914 | -250920,034 | -332651,393 |
| x123 | 52886,088 | 48237,902 | 64903,877 |
| x133 | 34956,483 | 36476,846 | 37715,175 |
| x136 | -1009,946 | -921,989 | -1298,435 |
| x137 | 580346,096 | 600638,921 | 437479,468 |
| (Константа) | -32297,907 | -30913,944 | -35527,559 |
| Линейные дискриминантные функции Фишера | | | |



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Результаты классификацииa,c** | | | | | | |
|  | | k | Предсказанная принадлежность к группе | | | Итого |
| 1,0000 | 2,0000 | 3,0000 |
| Исходные | Частота | 1,0000 | 18 | 0 | 0 | 18 |
| 2,0000 | 0 | 9 | 0 | 9 |
| 3,0000 | 0 | 0 | 4 | 4 |
| Несгруппированные наблюдения | 1 | 0 | 0 | 1 |
| % | 1,0000 | 100,0 | ,0 | ,0 | 100,0 |
| 2,0000 | ,0 | 100,0 | ,0 | 100,0 |
| 3,0000 | ,0 | ,0 | 100,0 | 100,0 |
| Несгруппированные наблюдения | 100,0 | ,0 | ,0 | 100,0 |
| Кросс-проверенныеb | Частота | 1,0000 | 15 | 3 | 0 | 18 |
| 2,0000 | 3 | 6 | 0 | 9 |
| 3,0000 | 0 | 0 | 4 | 4 |
| % | 1,0000 | 83,3 | 16,7 | ,0 | 100,0 |
| 2,0000 | 33,3 | 66,7 | ,0 | 100,0 |
| 3,0000 | ,0 | ,0 | 100,0 | 100,0 |

100,0% вихідних згрупованих спостережень класифіковано правильно.

80,6% перехресно-перевіряються згрупованих спостережень класифіковано правильно.

Результати підвищились до максимальних показників по правильності класифікації.

**Висновки**

Сенс дискримінантного аналізу - на підставі навчальних вибірок перетворити багатовимірний масив в одновимірний показник для прогнозування приналежності спостережень до груп, тобто побудувати новий узагальнений показник, значення якого максимально розрізняються для об'єктів, віднесених до різних груп. Навчальна вибірка - це безліч об'єктів, заданих значеннями ознак і приналежність яких до того чи іншого класу достовірно відома.

У цій роботі ми створили всі умови для оптимальної моделі класифікування ускладнень після протезування мітрального клапану. Ми отримали розділення станів на норма, гостра серцева недостатність, гостра недостатність мозгового кровообігу. 100% вихідних класифікацій спостережень виявились правильними згідно тесту.