**Постановка задачі**

У результаті проведення експерименту по вивченню 2-х властивостей було обстежено N об'єктів. Результати вимірів властивостей представлені в таблиці. Необхідно провести

А) ізотропне;

Б) ізотонічне;

В) ізоморфічне

перетворення отриманої сукупності даних. По методу куль розбити досліджувану сукупність об'єктів на підмножини.

Як результат представити:

А) матрицю відстаней , отриману для кожного з видів перетворення;

Б) критичне значення радіуса кулі і сукупність отриманих підмножин.

Вихідні дані приведені в табл. 1.

Таблиця 1.

|  |  |
| --- | --- |
| X | Y |
| 1,2 | 3,2 |
| 1,5 | 3,8 |
| 1,6 | 4 |
| 1,8 | 2,1 |
| 2,1 | 2,3 |
| 2,2 | 4,6 |
| 2,3 | 4,5 |
| 2,6 | 8,2 |
| 2,7 | 5,2 |
| 2,9 | 5,3 |

**Виконання завдання**

А) Ізотропне перетворення

Найчастіше використовуються такі таксономічні методи, за допомогою яких виділяються підмножини, однорідні в змісті ізотропності (об'єкти, що належать тій самій підмножині, мало відрізняються друг від друга за рівнем і структурою значень ознак). Справа в тім, що, як правило, позбутися від одиниць виміру ознак можна шляхом так званої стандартизації:

D:\Навчання\4\Аналіз даних\ТЕМА 1_files\image002.gif (1.1)

де *xij – i-*та реалізація *j-*тої ознаки;

– середнє значення *j-*тої ознаки;

*sj* – середньоквадратичне відхилення *j-*тої ознаки.

У результаті обидві компоненти реалізації ознак (структура і рівень) вирівняні.

Стандартизовані значення:

|  |  |
| --- | --- |
| Стандартизовані дані | |
| Х | У |
| -1,593072 | -0,640308 |
| -1,056081 | -0,297286 |
| -0,877085 | -0,182945 |
| -0,519091 | -1,269182 |
| 0,0178997 | -1,154841 |
| 0,1968965 | 0,160077 |
| 0,3758934 | 0,102907 |
| 0,9128839 | 2,218209 |
| 1,0918808 | 0,503099 |
| 1,4498745 | 0,560269 |

Відстань між об’єктами обчислюється за формулою звичайної евклідової метрики

D:\Навчання\4\Аналіз даних\ТЕМА 1_files\image006.gif (1.2)

Матриця відстаней наведена в табл. 2.

Таблиця 2

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Матриця відстаней | | | | | | | | | | |
|  | d1 | d2 | d3 | d4 | d5 | d6 | d7 | d8 | d9 | d10 |
| d1 | 0,0000 | 0,6372 | 0,8496 | 1,2446 | 1,6911 | 1,9608 | 2,1046 | 3,8014 | 2,9183 | 3,2712 |
| d2 | 0,6372 | 0,0000 | 0,2124 | 1,1104 | 1,3743 | 1,3338 | 1,4868 | 3,1945 | 2,2922 | 2,6486 |
| d3 | 0,8496 | 0,2124 | 0,0000 | 1,1437 | 1,3212 | 1,1274 | 1,2852 | 2,9949 | 2,0851 | 2,4428 |
| d4 | 1,2446 | 1,1104 | 1,1437 | 0,0000 | 0,5490 | 1,5986 | 1,6382 | 3,7699 | 2,3950 | 2,6877 |
| d5 | 1,6911 | 1,3743 | 1,3212 | 0,5490 | 0,0000 | 1,3270 | 1,3077 | 3,4898 | 1,9754 | 2,2343 |
| d6 | 1,9608 | 1,3338 | 1,1274 | 1,5986 | 1,3270 | 0,0000 | 0,1879 | 2,1791 | 0,9585 | 1,3153 |
| d7 | 2,1046 | 1,4868 | 1,2852 | 1,6382 | 1,3077 | 0,1879 | 0,0000 | 2,1824 | 0,8202 | 1,1673 |
| d8 | 3,8014 | 3,1945 | 2,9949 | 3,7699 | 3,4898 | 2,1791 | 2,1824 | 0,0000 | 1,7244 | 1,7427 |
| d9 | 2,9183 | 2,2922 | 2,0851 | 2,3950 | 1,9754 | 0,9585 | 0,8202 | 1,7244 | 0,0000 | 0,3625 |
| d10 | 3,2712 | 2,6486 | 2,4428 | 2,6877 | 2,2343 | 1,3153 | 1,1673 | 1,7427 | 0,3625 | 0,0000 |

Критичне значення радіуса кулі:

|  |
| --- |
| r= |
| 1,724426 |

Таким чином отримаємо першу однорідну підмножину:

A1={d2,d3,d4,d5,d6,d7,d9,d10}.

Вилучивши ці елементи з подальшого розгляду, одержимо матрицю відстаней табл. 3.

Таблиця 3

|  |  |  |
| --- | --- | --- |
| матриця відстаней | | |
|  | d1 | d8 |
| d1 | 0 | 3,801439 |
| d8 | 3,801439 | 0 |

Проаналізувавши її видно, що друга та відповідно третя підмножина буде містити такі об’єкти:

A2={d1};

A3={d8}.

Для аналізу отриманих результатів запишемо найбільше та найменше значення кожної ознаки та середнє значення ознак для кожної однорідної групи об’єктів (табл. 4).

Таблиця 4

|  |  |  |
| --- | --- | --- |
|  | X | Y |
| min | 1,2 | 2,1 |
| max | 2,9 | 8,2 |
| середнє | 2,09 | 4,32 |
| сер для А1 | 2,1375 | 3,975 |
| сер для А2 | 1,2 | 3,2 |
| сер для А3 | 2,6 | 8,2 |

Отже, однорідні групи характеризуються такими властивостями:

А1 Середні витрати на одиницю продукції, низька фондовіддача.

А2 Низькі витрати на одиницю продукції, низька фондовіддача.

А3 Високі витрати на одиницю продукції, висока фондовіддача.

Б) Ізотонічне перетворення:

Перш ніж почати виділення ізотонічних підмножин, потрібно позбутися від одиниць виміру ознак. З цією метою використовуємо перетворення, що виключить зі значень кожної ознаки компонентові структури, але збереже компонентові рівня. Для цього перетворимо ознаки D:\Навчання\4\Аналіз даних\ТЕМА 1_files\image008.gif

до виду

D:\Навчання\4\Аналіз даних\ТЕМА 1_files\image010.gif, (1.3)

за допомогою ізотонічного перетворення (Таблиця 5)

 (1.4)

Таблиця 5- значення кожної ознаки без компонентів структури.

|  |  |
| --- | --- |
| V1 | V2 |
| 0,057416 | 0,074074 |
| 0,07177 | 0,087963 |
| 0,076555 | 0,092593 |
| 0,086124 | 0,048611 |
| 0,100478 | 0,053241 |
| 0,105263 | 0,106481 |
| 0,110048 | 0,104167 |
| 0,124402 | 0,189815 |
| 0,129187 | 0,12037 |
| 0,138756 | 0,122685 |

Отримані в такий спосіб ознаки характеризуються тим, що їхні довжини ( що розуміються як довжини векторів) дорівнюють одиниці, тобто для ознак Vj виконується умова:

D:\Навчання\4\Аналіз даних\ТЕМА 1_files\image014.gif. (1.5)

Вихідні ж ознаки звичайно мають різну довжину.

При одержанні ознак Vj виключаються одиниці виміру вихідних ознак і встановлюються однакові довжини ознак, рівні 1.Таким чином, значення vij характеризують "потенціал", "масштаб", "позицію" або "ранг" даного об'єкта в розглянутій сукупності об'єктів. У свою чергу сума значень

D:\Навчання\4\Аналіз даних\ТЕМА 1_files\image016.gif (1.6)

характеризує внесок значень усіх властивостей даного об'єкта в значення всіх властивостей сукупності, що включає всі об'єктів. Ця величина синтетичним образом представляє "потенціал" або "позицію" j-го об'єкта в сукупності. Вектор D:\Навчання\4\Аналіз даних\ТЕМА 1_files\image018.gif називається ізотонічним показником сукупності даних (таблиця 6).

Таблиця 6 – Ізотонічні показники сукупності даних.

|  |
| --- |
| W=V1+V2 |
| 0,13149 |
| 0,159733 |
| 0,169148 |
| 0,134736 |
| 0,153719 |
| 0,211745 |
| 0,214215 |
| 0,314217 |
| 0,249557 |
| 0,261441 |

Для виділення підмножин, однорідних в ізотонічному відношенні, відстань між точками об’єктами визначається за формулою “міської” метрики:

D:\Навчання\4\Аналіз даних\ТЕМА 1_files\image020.gif. (1.7)

Елементи dіj приймають значення, рівні нулеві, якщо значення показників wi і wj ідентичні (табл. 7).

Таблиця 7 – Матриця відстаней.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Матриця відстаней | | | | | | | | | | |
|  | d1 | d2 | d3 | d4 | d5 | d6 | d7 | d8 | d9 | d10 |
| d1 | 0,0000 | 0,0282 | 0,0377 | 0,0032 | 0,0222 | 0,0803 | 0,0827 | 0,1827 | 0,1181 | 0,1300 |
| d2 | 0,0282 | 0,0000 | 0,0094 | 0,0250 | 0,0060 | 0,0520 | 0,0545 | 0,1545 | 0,0898 | 0,1017 |
| d3 | 0,0377 | 0,0094 | 0,0000 | 0,0344 | 0,0154 | 0,0426 | 0,0451 | 0,1451 | 0,0804 | 0,0923 |
| d4 | 0,0032 | 0,0250 | 0,0344 | 0,0000 | 0,0190 | 0,0770 | 0,0795 | 0,1795 | 0,1148 | 0,1267 |
| d5 | 0,0222 | 0,0060 | 0,0154 | 0,0190 | 0,0000 | 0,0580 | 0,0605 | 0,1605 | 0,0958 | 0,1077 |
| d6 | 0,0803 | 0,0520 | 0,0426 | 0,0770 | 0,0580 | 0,0000 | 0,0025 | 0,1025 | 0,0378 | 0,0497 |
| d7 | 0,0827 | 0,0545 | 0,0451 | 0,0795 | 0,0605 | 0,0025 | 0,0000 | 0,1000 | 0,0353 | 0,0472 |
| d8 | 0,1827 | 0,1545 | 0,1451 | 0,1795 | 0,1605 | 0,1025 | 0,1000 | 0,0000 | 0,0647 | 0,0528 |
| d9 | 0,1181 | 0,0898 | 0,0804 | 0,1148 | 0,0958 | 0,0378 | 0,0353 | 0,0647 | 0,0000 | 0,0119 |
| d10 | 0,1300 | 0,1017 | 0,0923 | 0,1267 | 0,1077 | 0,0497 | 0,0472 | 0,0528 | 0,0119 | 0,0000 |

Узагалі можна стверджувати, що чим вище значення dіj, тим більше відрізняються між собою розглянуті об'єкти за рівнем значень властивостей. Низькі значення dіj свідчать про малі розходження об'єктів за рівнем значень властивостей.

Критичне значення радіуса кулі:

|  |
| --- |
| r= |
| 0,0528 |

Таким чином отримаємо першу підмножину:

А1={d1,d2,d3,d4,d5,d6,d7}.

Вилучивши ці елементи з подальшого розгляду, одержимо матрицю відстаней табл. 8.

Таблиця 8 – Матриця відстаней.

|  |  |  |  |
| --- | --- | --- | --- |
| Матриця відстаней | | | |
|  | d8 | d9 | d10 |
| d8 | 0,0000 | 0,0647 | 0,0528 |
| d9 | 0,0647 | 0,0000 | 0,0119 |
| d10 | 0,0528 | 0,0119 | 0,0000 |

Проаналізувавши її видно, що друга та відповідно третя підмножина буде містити такі об’єкти:

А2={d9,d10};

А3={d8}.

Для аналізу отриманих результатів запишемо найбільше та найменше значення кожної ознаки та середнє значення ознак для кожної однорідної групи об’єктів (табл. 9).

Таблиця 9.

|  |  |  |
| --- | --- | --- |
|  | X | Y |
| min | 1,2 | 2,1 |
| max | 2,9 | 8,2 |
| середнє | 2,09 | 4,32 |
| сер для А1 | 1,814286 | 3,5 |
| сер для А2 | 2,8 | 5,25 |
| сер для А3 | 2,6 | 8,2 |

Отже, за рівнями значень показників об’єкти розподілились так :

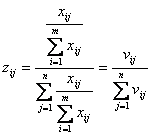
А1 – низькі витрати на одиницю продукції, низька фондовіддача.

А2 – високі витрати на одиницю продукції, середня фондовіддача.

А3 – високі витрати на одиницю продукції, висока фондовіддача.

В) Ізоморфічне перетворення

Такі підмножини характеризуються однорідністю з точки зору структури значень. Тобто, пропорції значень відповідних ознак будуть мало відрізнятись одна від одної. Для їх одержання потрібно провести наступні перетворення

. (1.8)

В результаті одержимо нові точки Ri = (zi1, zi2, …, zin) .Вони характеризуються тим, що D:\Навчання\4\Аналіз даних\ТЕМА 1_files\image024.gif. Відстань між такими точками визначається також за допомогою звичайної евклідової метрики:

D:\Навчання\4\Аналіз даних\ТЕМА 1_files\image026.gif. (1.9)

Мінімальна відстань між об’єктами Ri та Rj спостерігається тоді, коли вихідні вектори Pi та Pj колінеарні. Максимум відстані досягається, якщо вихідні вектори ортогональні. Помітимо, що значення перетворених векторів лежать в межах від 0 до 1. Отже в такому випадку максимальна відстань буде становити D:\Навчання\4\Аналіз даних\ТЕМА 1_files\image028.gif.

Проведемо стандартизацію даних і запишемо їх до табл. 10, а матрицю відстаней – в табл. 11.

Таблиця 10 – Стандартизовані дані.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| V1 | V2 | W=V1+V2 | Z1=V1/W | Z2=V2/W |
| 0,057416 | 0,074074 | 0,13149 | 0,436658 | 0,563342 |
| 0,07177 | 0,087963 | 0,159733 | 0,449314 | 0,550686 |
| 0,076555 | 0,092593 | 0,169148 | 0,452593 | 0,547407 |
| 0,086124 | 0,048611 | 0,134736 | 0,639211 | 0,360789 |
| 0,100478 | 0,053241 | 0,153719 | 0,653649 | 0,346351 |
| 0,105263 | 0,106481 | 0,211745 | 0,497123 | 0,502877 |
| 0,110048 | 0,104167 | 0,214215 | 0,513727 | 0,486273 |
| 0,124402 | 0,189815 | 0,314217 | 0,395911 | 0,604089 |
| 0,129187 | 0,12037 | 0,249557 | 0,517664 | 0,482336 |
| 0,138756 | 0,122685 | 0,261441 | 0,530735 | 0,469265 |

Табдиця 11 – Матриця відстаней.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Матриця відстаней | | | | | | | | | | |
|  | d1 | d2 | d3 | d4 | d5 | d6 | d7 | d8 | d9 | d10 |
| d1 | 0,0000 | 0,0179 | 0,0225 | 0,2865 | 0,3069 | 0,0855 | 0,1090 | 0,0576 | 0,1146 | 0,1330 |
| d2 | 0,0179 | 0,0000 | 0,0046 | 0,2686 | 0,2890 | 0,0676 | 0,0911 | 0,0755 | 0,0967 | 0,1151 |
| d3 | 0,0225 | 0,0046 | 0,0000 | 0,2639 | 0,2843 | 0,0630 | 0,0865 | 0,0802 | 0,0920 | 0,1105 |
| d4 | 0,2865 | 0,2686 | 0,2639 | 0,0000 | 0,0204 | 0,2009 | 0,1775 | 0,3441 | 0,1719 | 0,1534 |
| d5 | 0,3069 | 0,2890 | 0,2843 | 0,0204 | 0,0000 | 0,2214 | 0,1979 | 0,3645 | 0,1923 | 0,1738 |
| d6 | 0,0855 | 0,0676 | 0,0630 | 0,2009 | 0,2214 | 0,0000 | 0,0235 | 0,1431 | 0,0290 | 0,0475 |
| d7 | 0,1090 | 0,0911 | 0,0865 | 0,1775 | 0,1979 | 0,0235 | 0,0000 | 0,1666 | 0,0056 | 0,0241 |
| d8 | 0,0576 | 0,0755 | 0,0802 | 0,3441 | 0,3645 | 0,1431 | 0,1666 | 0,0000 | 0,1722 | 0,1907 |
| d9 | 0,1146 | 0,0967 | 0,0920 | 0,1719 | 0,1923 | 0,0290 | 0,0056 | 0,1722 | 0,0000 | 0,0185 |
| d10 | 0,1330 | 0,1151 | 0,1105 | 0,1534 | 0,1738 | 0,0475 | 0,0241 | 0,1907 | 0,0185 | 0,0000 |

Критичне значення радіуса кулі:

|  |
| --- |
| r= |
| 0,057624 |

Отже, перша підмножина містить такі об’єкти:

А1={d6,d7,d9,d10}.

Перетворена матриця відстаней наведена в табл. 12.

Табдиця 12 – Матриця відстаней.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Матриця відстаней | | | | | | |
|  | d1 | d2 | d3 | d4 | d5 | d8 |
| d1 | 0,0000 | 0,0179 | 0,0225 | 0,2865 | 0,3069 | 0,0576 |
| d2 | 0,0179 | 0,0000 | 0,0046 | 0,2686 | 0,2890 | 0,0755 |
| d3 | 0,0225 | 0,0046 | 0,0000 | 0,2639 | 0,2843 | 0,0802 |
| d4 | 0,2865 | 0,2686 | 0,2639 | 0,0000 | 0,0204 | 0,3441 |
| d5 | 0,3069 | 0,2890 | 0,2843 | 0,0204 | 0,0000 | 0,3645 |
| d8 | 0,0576 | 0,0755 | 0,0802 | 0,3441 | 0,3645 | 0,0000 |

Отже, друга підмножина містить такі об’єкти:

А2={d1,d2,d3}.

Одержимо наступну матрицю відстаней табл. 13.

Табдиця 13 – Матриця відстаней.

|  |  |  |  |
| --- | --- | --- | --- |
| Матриця відстаней | | | |
|  | d4 | d5 | d8 |
| d4 | 0,0000 | 0,0204 | 0,3441 |
| d5 | 0,0204 | 0,0000 | 0,3645 |
| d8 | 0,3441 | 0,3645 | 0,0000 |

Як видно, об’єкти розподіляться по двом групам.

Отже, в результаті маємо 4 однорідні з точки зору структури даних підмножини:

А1={d6,d7,d9,d10},

А2={d1,d2,d3},

А3={d4,d5},

A4={d8}.

Проаналізуємо розподіл об’єктів в такому випадку. Для цього визначимо структуру ознак для кожного об’єкта, поділивши значення ознак. Результати для аналізу наведені в табл. 14.

Таблиця 14 – Структура даних.

|  |  |  |  |
| --- | --- | --- | --- |
| Структура даних | | | |
| n | X | Y | X/Y |
| 1 | 1,2 | 3,2 | 0,375 |
| 2 | 1,5 | 3,8 | 0,394737 |
| 3 | 1,6 | 4 | 0,4 |
| 4 | 1,8 | 2,1 | 0,857143 |
| 5 | 2,1 | 2,3 | 0,913043 |
| 6 | 2,2 | 4,6 | 0,478261 |
| 7 | 2,3 | 4,5 | 0,511111 |
| 8 | 2,6 | 8,2 | 0,317073 |
| 9 | 2,7 | 5,2 | 0,519231 |
| 10 | 2,9 | 5,3 | 0,54717 |

Як видно, першу групу переважно складають об’єкти, в яких значення частки ознаки X менше значення частки ознаки У приблизно в 2 рази. Друга група включає об’єкти, в яких значення частки ознаки X менше значення частки ознаки У приблизно в 2,5 -2,6 рази. Третя підмножина містить об’єкти з приблизно однаковими частками ознак. Четверту групу складають об’єкти, в яких значення часток ознак X та Y приблизно знаходяться у співвідношенні як 1 : 3.