Bài 02 \_Thực hành

(Cơ bản) Một website thương mại điện tử chuyên kinh doanh thiết bị gia dụng thu thập được những đơn hàng như sau:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Đơn hàng*** | **Máy lạnh** | **Máy giặt** | **Tủ lạnh** | **Tivi** | **Bếp điện** |
| *1* | 1 | 0 | 1 | 1 | 1 |
| *2* | 1 | 0 | 1 | 0 | 1 |
| *3* | 0 | 0 | 1 | 1 | 1 |
| *4* | 1 | 0 | 0 | 1 | 0 |
| *5* | 0 | 1 | 0 | 1 | 0 |
| *6* | 0 | 1 | 0 | 0 | 1 |
| *7* | 0 | 0 | 0 | 0 | 0 |
| *8* | 1 | 1 | 1 | 0 | 1 |
| *9* | 0 | 1 | 0 | 1 | 0 |

Công ty dự định xây dựng một hệ thống khuyến nghị dựa trên việc gom cụm những đơn hàng trên thành 3 nhóm có độ tương đồng với nhau. Từ đó gợi ý cho việc mua hàng của khách hàng dựa trên sự tương đồng này. Sinh viên hãy tham gia xây dựng hệ thống khuyến nghị bằng cách thực hiện những yêu cầu sau:

* + 1. Sử dụng thuật toán *k*-Means.
* Với k = 3 ta khởi tạo ma trận phân hoạch như sau:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Đơn hàng 1** | **Đơn hàng 2** | **Đơn hàng 3** | **Đơn hàng 4** | **Đơn hàng 5** | **Đơn hàng 6** | **Đơn hàng 7** | **Đơn hàng 8** | **Đơn hàng 9** |
| Cụm 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| Cụm 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Cụm 3 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 |

* Vecto trọng tâm của từng cụm là
* Cụm V1 = (
* Cụm V2 = (
* Cụm V3 = (
* Tính khoảng cách Eculide từng vị trí lần lượt đến các cụm là

|  |  |  |  |
| --- | --- | --- | --- |
|  | Cụm 1 | Cụm 2 | Cụm 3 |
| Đơn hàng 1 | **0,61** | 2 | 1,66 |
| Đơn hàng 2 | **0,61** | 1,41 | 1,80 |
| Đơn hàng 3 | **0,94** | 1,73 | 1,5 |
| Đơn hàng 4 | 1,54 | 1,41 | **1,12** |
| Đơn hàng 5 | 1,84 | 1,41 | **0,5** |
| Đơn hàng 6 | 1,54 | 1,41 | **1,12** |
| Đơn hàng 7 | 1,70 | **0** | 1,12 |
| Đơn hàng 8 | **0,94** | 2 | 1,66 |
| Đơn hàng 9 | 1,84 | 1,41 | **0,5** |

* Ma trận phân hoạch các điểm thuộc cụm

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Đơn hàng 1** | **Đơn hàng 2** | **Đơn hàng 3** | **Đơn hàng 4** | **Đơn hàng 5** | **Đơn hàng 6** | **Đơn hàng 7** | **Đơn hàng 8** | **Đơn hàng 9** |
| Cụm 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| Cụm 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Cụm 3 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 |

* Như vậy các điểm thuộc cụm không thay đổi, thuật toán dừng lại, với kết quả sau đây:
* Vị trí 1, 2, 3, 8 thuộc cụm 1
* Vị trí 7 thuộc cụm 2
* Vị trí 4, 5, 6, 9 thuộc cụm 3

1. Sử dụng mạng Kohonen với các thông số: 𝑒𝑝𝑜𝑐ℎ𝑠 = 10, 𝑅 = 0 và 𝛼 = 0,8.

* Khởi tạo giá trị của vecto trọng số
* w1 = (
* w2 = (
* w3 = (
* Ở lần lặp thứ nhất
* Xét vecto đầu tiên (đơn hàng 1) x1
* Khoảng cách từ x1 đến w1 là D11 **0,47**
* Khoảng cách từ x1 đến w2 là D12 2
* Khoảng cách từ x1 đến w3 là D13
* Như vậy, nơ ron có trọng số w1 là nơ ron có khoảng cách ngắn nhất đến x1. Tiến hành cập nhật trọng số của vecto w1
* Ta có
* w11 (new) = 2/3 + 0,8\*(1-2/3) = 14/15
* w12(new) = 0 + 0,8\*(0-0) = 0
* w13(new) = 1+0,8\*(1-1)= 1
* w14(new) = 2/3 + 0,8\*(1-2/3) = 14/15
* w15(new) = 1+0,8\*(1-1)= 1
* Lúc này, do R = 0 nên không cần cập nhật vùng lân cận w2, w3 sẽ giữ nguyên giá trị, ta có:
* w1 = (
* w2 = (
* w3 = (
* Giảm độ học đi 1 nửa  **= 0,8/2 = 0,4**
* Ở lần lặp thứ hai
* Xét vecto thứ 2 (đơn hàng 1) x1
* Khoảng cách từ x2 đến w1 là D21 **0,94**
* Khoảng cách từ x2 đến w2 là D22 1,73
* Khoảng cách từ x2 đến w3 là D23
* Như vậy, nơ ron có trọng số w1 là nơ ron có khoảng cách ngắn nhất đến x2. Tiến hành cập nhật trọng số của vecto w1
* Ta có
* w11 (new) = 14/15 + 0,4\*(1-14/15) = 0,96
* w12(new) = 0 + 0,4\*(0-0) = 0
* w13(new) = 1+0,4\*(1-1)= 1
* w14(new) = 14/5 + 0,4\*(0-14/5) = 0,56
* w15(new) = 1+0,4\*(1-1)= 1
* Lúc này, do R = 0 nên không cần cập nhật vùng lân cận w2, w3 sẽ giữ nguyên giá trị, ta có:
* w1 = (
* w2 = (
* w3 = (
* Giảm độ học đi 1 nửa  **= 0,4/2 = 0,2**
* Ở lần lặp thứ ba
* Xét vecto thứ 3 (đơn hàng 3) x3
* Khoảng cách từ x3 đến w1 là D31 **1,056**
* Khoảng cách từ x3 đến w2 là D32 1,73
* Khoảng cách từ x3 đến w3 là D33
* Như vậy, nơ ron có trọng số w1 là nơ ron có khoảng cách ngắn nhất đến x3. Tiến hành cập nhật trọng số của vecto w1
* Ta có
* w11 (new) = 0,96 + 0,2\*(0-0,96) = 0,768
* w12(new) = 0 + 0,2\*(0-0) = 0
* w13(new) = 1+0,2\*(1-1)= 1
* w14(new) = 0,56 + 0,2\*(1-0,56) = 0,648
* w15(new) = 1+0,2\*(1-1)= 1
* Lúc này, do R = 0 nên không cần cập nhật vùng lân cận w2, w3 sẽ giữ nguyên giá trị, ta có:
* w1 = (
* w2 = (
* w3 = (
* Giảm độ học đi 1 nửa  **= 0,2/2 = 0,1**
* Ở lần lặp thứ tư
* Xét vecto thứ 4 (đơn hàng 4) x4
* Khoảng cách từ x4 đến w1 là D41 1,48
* Khoảng cách từ x4 đến w2 là D42 1,41
* Khoảng cách từ x4 đến w3 là D43 **1,16**
* Như vậy, nơ ron có trọng số w3 là nơ ron có khoảng cách ngắn nhất đến x4. Tiến hành cập nhật trọng số của vecto w3
* Ta có
* w31 (new) = 0,4+ 0,1\*(1-0,4) = 0,46
* w32(new) = 0,8 + 0,1\*(0-0,8) = 0,72
* w33(new) = 0,2+0,1\*(0-0,2)= 0,18
* w34(new) = 0,6 + 0,1\*(1-0,6) = 0,64
* w35(new) = 0,4+0,1\*(0-0,4)= 0,36
* Lúc này, do R = 0 nên không cần cập nhật vùng lân cận w1, w2 sẽ giữ nguyên giá trị, ta có:
* w1 = (
* w2 = (
* w3 = (
* Giảm độ học đi 1 nửa  **= 0,1/2 = 0,05**
* Ở lần lặp thứ năm
* Xét vecto thứ 5 (đơn hàng 5) x5
* Khoảng cách từ x5 đến w1 là D51 1,93
* Khoảng cách từ x5 đến w2 là D52 1,41
* Khoảng cách từ x5 đến w3 là D53 **0,76**
* Như vậy, nơ ron có trọng số w3 là nơ ron có khoảng cách ngắn nhất đến x5. Tiến hành cập nhật trọng số của vecto w3
* Ta có
* w31(new) = 0,46+ 0,05\*(0-0,46) = 0,437
* w32(new) = 0,72 + 0,05\*(1-0,72) = 0,734
* w33(new) = 0,18+0,05\*(0-0,18)= 0,171
* w34(new) = 0,64 + 0,05\*(1-0,64) = 0,658
* w35(new) = 0,36+0,05\*(0-0,36)= 0,342
* Lúc này, do R = 0 nên không cần cập nhật vùng lân cận w1, w2 sẽ giữ nguyên giá trị, ta có:
* w1 = (
* w2 = (
* w3 = (
* Giảm độ học đi 1 nửa  **= 0,05/2 = 0,025**
* Ở lần lặp thứ sáu
* Xét vecto thứ 6 (đơn hàng 6) x6
* Khoảng cách từ x6 đến w1 là D61 1,73
* Khoảng cách từ x6 đến w2 là D62 1,41
* Khoảng cách từ x6 đến w3 là D63 **1,08**
* Như vậy, nơ ron có trọng số w3 là nơ ron có khoảng cách ngắn nhất đến x6. Tiến hành cập nhật trọng số của vecto w3
* Ta có
* w31(new) = 0,437+ 0,025\*(0-0,437) = 0,43
* w32(new) = 0,734 + 0,025\*(1-0,734) = 0,74
* w33(new) = 0,171+0,025\*(0-0,171)= 0,17
* w34(new) = 0,658 + 0,025\*(0-0,658) = 0,64
* w35(new) = 0,342+0,025\*(1-0,342)= 0,358
* Lúc này, do R = 0 nên không cần cập nhật vùng lân cận w1, w2 sẽ giữ nguyên giá trị, ta có:
* w1 = (
* w2 = (
* w3 = (
* Giảm độ học đi 1 nửa  **= 0,025/2 = 0,0125**
* Ở lần lặp thứ bảy
* Xét vecto thứ 7 (đơn hàng 7) x7
* Khoảng cách từ x7 đến w1 là D71 1,73
* Khoảng cách từ x7 đến w2 là D72 **0**
* Khoảng cách từ x7 đến w3 là D73 1,13
* Như vậy, nơ ron có trọng số w2 là nơ ron có khoảng cách ngắn nhất đến x7. Tiến hành cập nhật trọng số của vecto w2
* Ta có
* w21(new) = 0+ 0,0125\*(0-0) = 0
* w22(new) = 0+ 0,0125\*(0-0) = 0
* w23(new) = 0+0,0125\*(0-0)= 0
* w24(new) = 0+ 0,0125\*(0-0) = 0
* w25(new) = 0+0,0125\*(0-0)= 0
* Lúc này, do R = 0 nên không cần cập nhật vùng lân cận w1, w3 sẽ giữ nguyên giá trị, ta có:
* w1 = (
* w2 = (
* w3 = (
* Giảm độ học đi 1 nửa  **= 0,0125/2 = 0,00625**
* Ở lần lặp thứ tám
* Xét vecto thứ 8 (đơn hàng 8) x8
* Khoảng cách từ x8 đến w1 là D81 **1,21**
* Khoảng cách từ x8 đến w2 là D82 2
* Khoảng cách từ x8 đến w3 là D83 1,37
* Như vậy, nơ ron có trọng số w1 là nơ ron có khoảng cách ngắn nhất đến x8. Tiến hành cập nhật trọng số của vecto w1
* Ta có
* w11(new) = 0,768+ 0,00625\*(1-0,768) = 0,77
* w12(new) = 0+ 0,00625\*(1-0) = 0,006
* w13(new) = 1+0,00625\*(1-1)= 1
* w14(new) = 0,648+ 0,00625\*(0-0,648) = 0,644
* w15(new) = 1+0,00625\*(1-1)= 1
* Lúc này, do R = 0 nên không cần cập nhật vùng lân cận w2, w3 sẽ giữ nguyên giá trị, ta có:
* w1 = (
* w2 = (
* w3 = (
* Giảm độ học đi 1 nửa  **= 0,00625/2 = 0,003125**
* Ở lần lặp thứ chín
* Xét vecto thứ 9 (đơn hàng 9) x9
* Khoảng cách từ x9 đến w1 là D91 1,93
* Khoảng cách từ x9 đến w2 là D92 1,41
* Khoảng cách từ x9 đến w3 là D93 **0,73**
* Như vậy, nơ ron có trọng số w3 là nơ ron có khoảng cách ngắn nhất đến x9. Tiến hành cập nhật trọng số của vecto w3
* Ta có
* w31(new) = 0,43+ 0,003125\*(0-0,43) = 0,43
* w32(new) = 0,74+ 0,003125\*(1-0,74) = 0,741
* w33(new) = 0,17+0,003125\*(0-0,17)= 0,169
* w34(new) = 0,64+ 0,003125\*(1-0,64) = 0,641
* w35(new) = 0,358+0,003125\*(0-0,358)= 0,357
* Lúc này, do R = 0 nên không cần cập nhật vùng lân cận w1, w2 sẽ giữ nguyên giá trị, ta có:
* w1 = (
* w2 = (
* w3 = (
* Thuật toán dừng lại, với kết quả sau đây:
  + Vị trí 1, 2, 3, 8 thuộc cụm 1
  + Vị trí 7 thuộc cụm 2
  + Vị trí 4, 5, 6, 9 thuộc cụm 3

1. So sánh kết quả thu được từ thuật toán *k*-Means và mạng Kohonen với nhau.

|  |  |
| --- | --- |
| ***k*-Means** | **Mạng Kohonen** |
| Cụm 1 | *Cụm 1* |
| Cụm 2 | *Cụm 3* |
| Cụm 3 | *Cụm 2* |

1. Một khách đang thực hiện đặt hàng trên hệ thống đã mua Máy lạnh, tủ lạnh. Theo em những mặt hàng nào nên được khuyến nghị dựa theo kết quả đã tính được từ các câu trên.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Đơn hàng** | **Máy lạnh** | **Máy giặt** | **Tủ lạnh** | **Tivi** | **Bếp điện** |
| **10** | 1 | 0 | 1 | 0 | 0 |

* Dùng thuật toán K-Means

Ta có khoảng cách Eculide giữa vị trí 10 và v1 là D101 1,247

Ta có khoảng cách Eculide giữa vị trí 10 và v2 là D102 1,41

Ta có khoảng cách Eculide giữa vị trí 10 và v3 là D103 1,47

* Vậy vị trí số 10 thuộc cụm v1
* Nhựng mặt hàng được khuyến nghị: Ti vi, bếp điện