Bài 01 \_ Thực hành

(Cơ bản) Trong quá trình thống kê doanh thu, một công ty kinh doanh chuỗi cửa hàng pizza phát hiện những vị trí có nhu cầu cao nhưng chưa có cửa hàng trong khu vực. Tọa độ tương đối của những vị trí tiềm năng đó được cho trong bảng sau:

|  |  |  |  |
| --- | --- | --- | --- |
| **Vị trí** | ***Tọa độ x*** | ***Tọa độ y*** | ***Số đơn đặt gần vị trí trong năm*** |
| **1** | 8 | 4 | 200 |
| **2** | 8 | 6 | 350 |
| **3** | 9 | 7 | 650 |
| **4** | 10 | 5 | 400 |
| **5** | 11 | 4 | 320 |
| **6** | 11 | 8 | 250 |
| **7** | 12 | 6 | 600 |
| **8** | 12 | 7 | 300 |
| **9** | 14 | 5 | 200 |

Công ty này muốn xây dựng 3 cửa hàng pizza mới trong những khu vực trên nhằm mở rộng chuỗi cửa hàng và phục vụ tốt hơn cho những khu vực này. Sinh viên hãy giúp công ty bằng cách tìm ra vị trí đặt 3 cửa hàng sao cho thuận lợi nhất và chỉ ra cụm khách hàng của từng cửa hàng với những yêu cầu cụ thể sau:

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

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Vị trí 1** | **Vị trí 2** | **Vị trí 3** | **Vị trí 4** | **Vị trí 5** | **Vị trí 6** | **Vị trí 7** | **Vị trí 8** | **Vị trí 9** |
| Cụm 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| Cụm 2 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| Cụm 3 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |

* 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 |
| Vị trí 1 | **106,68** | 200.02 | 375.03 |
| Vị trí 2 | 43,39 | 50,05 | **33,55** |
| Vị trí 3 | 343,34 | **250,01** | 266,68 |
| Vị trí 4 | 93,34 | **1,05** | 16,87 |
| Vị trí 5 | **13,40** | 80,03 | 63,41 |
| Vị trí 6 | **56,80** | 150.01 | 133,34 |
| Vị trí 7 | 293,35 | **200,01** | 216,67 |
| Vị trí 8 | **7,55** | 100,01 | 83,33 |
| Vị trí 9 | **106,76** | 200,04 | 183,36 |

* Ma trận phân hoạch các điểm thuộc cụm

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Vị trí 1** | **Vị trí 2** | **Vị trí 3** | **Vị trí 4** | **Vị trí 5** | **Vị trí 6** | **Vị trí 7** | **Vị trí 8** | **Vị trí 9** |
| Cụm 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| Cụm 2 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 |
| Cụm 3 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

* Vecto trọng tâm của từng cụm là
* Cụm V1 = (11.2
* Cụm V2 = (
* Cụm V3 = (
* Tính khoảng cách Eculide từng vị trí lần lượt đến các cụm là

|  |  |  |  |
| --- | --- | --- | --- |
|  | Cụm 1 | Cụm 2 | Cụm 3 |
| Vị trí 1 | **54,12** | 350,01 | 150,01 |
| Vị trí 2 | 96,05 | 200,01 | **0** |
| Vị trí 3 | 396,01 | **100,01** | 300,00 |
| Vị trí 4 | 146,01 | 150,00 | **50,05** |
| Vị trí 5 | 66,02 | 230,01 | **30,22** |
| Vị trí 6 | **4,67** | 300,01 | 100,06 |
| Vị trí 7 | 346.00 | **50,02** | 250,03 |
| Vị trí 8 | **46,03** | 250,01 | 50,17 |
| Vị trí 9 | **54,08** | 350,02 | 150,12 |

* Ma trận phân hoạch các điểm thuộc cụm

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Vị trí 1** | **Vị trí 2** | **Vị trí 3** | **Vị trí 4** | **Vị trí 5** | **Vị trí 6** | **Vị trí 7** | **Vị trí 8** | **Vị trí 9** |
| Cụm 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |
| Cụm 2 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| Cụm 3 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |

* Vecto trọng tâm của từng cụm là
* Cụm V1 = (11.2
* Cụm V2 = (
* Cụm V3 = (
* Tính khoảng cách Eculide từng vị trí lần lượt đến các cụm là

|  |  |  |  |
| --- | --- | --- | --- |
|  | Cụm 1 | Cụm 2 | Cụm 3 |
| Vị trí 1 | **37,69** | 425,01 | 156,68 |
| Vị trí 2 | 112,5 | 275,01 | **6,94** |
| Vị trí 3 | 412,51 | **25,05** | 293,34 |
| Vị trí 4 | 162,51 | 225,01 | **43,33** |
| Vị trí 5 | 82,52 | 305,01 | **36,70** |
| Vị trí 6 | **12,66** | 375,00 | 106,72 |
| Vị trí 7 | 362,50 | **25,05** | 243,35 |
| Vị trí 8 | 62,51 | 325,00 | **56,75** |
| Vị trí 9 | **37,61** | 425,02 | 156,72 |

* Ma trận phân hoạch các điểm thuộc cụm

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Vị trí 1** | **Vị trí 2** | **Vị trí 3** | **Vị trí 4** | **Vị trí 5** | **Vị trí 6** | **Vị trí 7** | **Vị trí 8** | **Vị trí 9** |
| Cụm 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| Cụm 2 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| Cụm 3 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |

* Vecto trọng tâm của từng cụm là
* Cụm V1 = (11
* Cụm V2 = (
* Cụm V3 = (
* Tính khoảng cách Eculide từng vị trí lần lượt đến các cụm là:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Cụm 1 | Cụm 2 | Cụm 3 |
| Vị trí 1 | **17,02** | 425,01 | 142,53 |
| Vị trí 2 | 133,37 | 275,01 | **7,85** |
| Vị trí 3 | 433,34 | **25,05** | 307,51 |
| Vị trí 4 | 183,34 | 225,01 | **57,50** |
| Vị trí 5 | 103,35 | 305,01 | **22,56** |
| Vị trí 6 | **33,414** | 375,00 | 92,54 |
| Vị trí 7 | 383,34 | **25,05** | 257,51 |
| Vị trí 8 | 83,35 | 325,00 | **42,56** |
| Vị trí 9 | **16,95** | 425,02 | 142,55 |

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

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

* Dữ liệu đầu vào là thông tin của 9 vị trí, mỗi thông tin có thể xem là một vecto trong không gian n = 3 chiều.
* Bán kính vùng lân cận R = 0, điều này có nghĩa là khi thay đổi trọng số của một nơron thì những nơron
* Số lần lặp của thuật toán epochs = 10
* Tốc độ học là
* Vì mục đích bài toán là gom thành 3 cụm nên mạng kohonen ở đầu ra gồm 3 nơron.
* Khởi tạo giá trị của vecto trọng số
* w1 = (11
* w2 = (
* w3 = (
* Ở lần lặp thứ nhất
* Xét vecto đầu tiên (vị trí 1) x1
* Khoảng cách từ x1 đến w1 là D11 **17,02**
* Khoảng cách từ x1 đến w2 là D12 425,01
* 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) = 11 + 0,8\*(8-11) =8,6
* w12(new) = 17/3 + 0,8\*(4-17/3) = 13/3
* w13(new) = 650/3+0,8\*(200-650/3)= 610/3
* Lúc này, do R = 0 nên không cần cập nhật vùng lân cận w2, w3 sẽ giữ nguyên giá trị, ta có:
* w1 = (8.6;13/3;610/3
* w2 = (
* w3 = (
* Giảm độ học đi 1 nửa  **= 0,8/2 = 0,4**
* Ở lần lặp thứ hai
* Xét vecto thứ 2 (vị trí 2) x2
* Khoảng cách từ x2 đến w1 là D21 146,68
* Khoảng cách từ x2 đến w2 là D22 275,01
* Khoảng cách từ x2 đến w3 là D23
* Như vậy, nơ ron có trọng số w3 là nơ ron có khoảng cách ngắn nhất đến x2. Tiến hành cập nhật trọng số của vecto w3
* Ta có
* w31 (new) = 10,25 + 0,4\*(8-10,25) = 9,35
* w32(new) = 5,5 + 0,4\*(6-5,5) = 5,7
* w33(new) = 342,5+0,4\*(350-342,5)= 345,5
* Lúc này, do R = 0 nên không cần cập nhật vùng lân cận w1, w2 sẽ giữ nguyên giá trị, ta có:
* w1 = (8.6;13/3;610/3
* w2 = (
* w3 = (
* Giảm độ học đi 1 nửa  **= 0,4/2 = 0,2**
* Ở lần lặp thứ ba
* Xét vecto thứ 3 (vị trí 3) x3
* Khoảng cách từ x3 đến w1 là D31 446,67
* Khoảng cách từ x3 đến w2 là D32 **25,05**
* Khoảng cách từ x3 đến w3 là D33
* Như vậy, nơ ron có trọng số w2 là nơ ron có khoảng cách ngắn nhất đến x3. Tiến hành cập nhật trọng số của vecto w2
* Ta có
* w21 (new) = 10,5 + 0,2\*(9-10,5) = 10,2
* w22(new) = 6,5 + 0,2\*(7-6,5) = 6,
* w23(new) = 625+0,2\*(650-625)= 630
* Lúc này, do R = 0 nên không cần cập nhật vùng lân cận w1, w3 sẽ giữ nguyên giá trị, ta có:
* w1 = (8.6;13/3;610/3
* w2 = (
* w3 = (
* Giảm độ học đi 1 nửa  **= 0,2/2 = 0,1**
* Ở lần lặp thứ tư
* Xét vecto thứ 4 (vị trí 4) x4
* Khoảng cách từ x4 đến w1 là D41 196,67
* Khoảng cách từ x4 đến w2 là D42 230,01
* Khoảng cách từ x4 đến w3 là D43
* 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) = 9,35 + 0,1\*(10-9,35) = 9,415
* w32(new) = 5,7 + 0,1\*(5-5,7) = 5,63
* w33(new) = 345,5+0,1\*(400-345,5)= 350,95
* Lúc này, do R = 0 nên không cần cập nhật vùng lân cận w1, w2 sẽ giữ nguyên giá trị, ta có:
* w1 = (8.6;13/3;610/3
* w2 = (
* w3 = (
* Giảm độ học đi 1 nửa  **= 0,1/2 = 0,05**
* Ở lần lặp thứ năm
* Xét vecto thứ 5 (vị trí 5) x5
* Khoảng cách từ x5 đến w1 là D51 116,69
* Khoảng cách từ x5 đến w2 là D52 310,01
* Khoảng cách từ x5 đến w3 là D53
* 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) = 9,415 + 0,05\*(11-9,415) 9,49
* w32(new) = 5,63+ 0,05\*(4-5,63) = 5,5485
* w33(new) = 350,95+0,05\*(320-350,95)= 349,4025
* Lúc này, do R = 0 nên không cần cập nhật vùng lân cận w1, w2 sẽ giữ nguyên giá trị, ta có:
* w1 = (8.6;13/3;610/3
* w2 = (
* w3 = (
* Giảm độ học đi 1 nửa  **= 0,05/2 = 0,025**
* Ở lần lặp thứ sáu
* Xét vecto thứ 6 (vị trí 6) x6
* Khoảng cách từ x6 đến w1 là D61 **46,87**
* Khoảng cách từ x6 đến w2 là D62 380,00
* Khoảng cách từ x6 đến w3 là D63
* Như vậy, nơ ron có trọng số w1 là nơ ron có khoảng cách ngắn nhất đến x6. Tiến hành cập nhật trọng số của vecto w1
* Ta có
* w11 (new) = 8,6 + 0,025\*(11-8,6) 8,66
* w12(new) = 13/3+ 0,025\*(8-13/3) = 4,425
* w13(new) = 610/3+0,025\*(250-610/3)= 204,5
* Lúc này, do R = 0 nên không cần cập nhật vùng lân cận w2, w3 sẽ giữ nguyên giá trị, ta có:
* w1 = (8.66;4,425;204,5
* w2 = (
* w3 = (
* Giảm độ học đi 1 nửa  **= 0,025/2 = 0,0125**
* Ở lần lặp thứ bảy
* Xét vecto thứ 7 (vị trí 7) x7
* Khoảng cách từ x7đến w1 là D71 395,52
* Khoảng cách từ x7 đến w2 là D72 **30,06**
* Khoảng cách từ x7đến w3 là D73 250,61
* 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) = 10,2 + 0,0125\*(12-10,2) 10,22
* w22 (new) = 6,6+ 0,0125\*(6-6,6) = 6,59
* w23 (new) = 630+0,0125\*(600-630)= 629,63
* Lúc này, do R = 0 nên không cần cập nhật vùng lân cận w1, w3 sẽ giữ nguyên giá trị, ta có:
* w1 = (8.66;4,425;204,5
* w2 = (
* w3 = (
* Giảm độ học đi 1 nửa  **= 0,0125/2 = 0,00625**
* Ở lần lặp thứ tám
* Xét vecto thứ 8 (vị trí 8) x8
* Khoảng cách từ x8 đến w1 là D81 95,59
* Khoảng cách từ x8 đến w2 là D82 329,64
* Khoảng cách từ x8 đến w3 là D83 **49,49**
* Như vậy, nơ ron có trọng số w3 là nơ ron có khoảng cách ngắn nhất đến x8. Tiến hành cập nhật trọng số của vecto w3
* Ta có
* w31 (new) = 9,49 + 0,00625\*(12-9,49) 9,51
* w32 (new) = 5,5485+ 0,00625\*(7-5,5485)
* w33 (new) = 349,4025+0,00625\*(300-349,4025)= 349,1
* Lúc này, do R = 0 nên không cần cập nhật vùng lân cận w1, w2 sẽ giữ nguyên giá trị, ta có:
* w1 = (8.66;4,425;204,5
* w2 = (
* w3 = (
* Giảm độ học đi 1 nửa  **= 0,00625/2 = 0,003125**
* Ở lần lặp thứ chín
* Xét vecto thứ 9 (vị trí 9) x9
* Khoảng cách từ x9 đến w1 là D91  **7,01**
* Khoảng cách từ x9 đến w2 là D92 429,65
* Khoảng cách từ x9 đến w3 là D93 149,17
* Như vậy, nơ ron có trọng số w1 là nơ ron có khoảng cách ngắn nhất đến x9. Tiến hành cập nhật trọng số của vecto w1
* Ta có
* w11 (new) = 8,66 + 0,003125\*(14-8,66) 8,68
* w12 (new) = 4,425+ 0,003125\*(5-4,425)
* w13 (new) = 204,5+0,003125\*(200-204,5)= 204,49
* Lúc này, do R = 0 nên không cần cập nhật vùng lân cận w2, w3 sẽ giữ nguyên giá trị, ta có:
* w1 = (8.68;4,43;204,49
* w2 = (
* w3 = (
* Vì n = 10 nên kết thúc thuật toán
* Thuật toán dừng lại, với kết quả sau đây:
* Vị trí 1, 6, 9 thuộc cụm 1
* Vị trí 3, 7 thuộc cụm 2
* Vị trí 2, 4, 5, 8 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 2* |
| Cụm 3 | *Cụm 3* |

1. Một vị trí tiềm năng mới xuất hiện tại tọa độ:

|  |  |  |  |
| --- | --- | --- | --- |
| **Vị trí** | ***Tọa độ x*** | ***Tọa độ y*** | ***Số đơn đặt gần vị trí trong năm*** |
| 10 | 11 | 6 | 450 |

Sinh viên hãy giúp công ty xác định cửa hàng nào trong 3 cửa hàng trên sẽ phục vụ cho vị trí này tốt nhất.

* Dùng thuật toán K-Means

Ta có khoảng cách Eculide giữa vị trí 10 và v1 là D101 233,33

Ta có khoảng cách Eculide giữa vị trí 10 và v2 là D102 175,00

Ta có khoảng cách Eculide giữa vị trí 10 và v3 là D103 107,50

* Vậy vị trí số 10 thuộc cụm v3