**Aula Prática 11**

**Validação Agrupamento e Regras de Associação**

**Murielly Oliveira Nascimento – 11921BSI222**

1. Dados os dados do exemplo abaixo, avaliar os grupos gerados pelo algoritmo k-means usando a silhueta (completa) e a silhueta simplificada. Use K=2 e as duas primeiras instâncias como os centroides iniciais.

Tabela

Descrição gerada automaticamente

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nome | Febre | Enjôo | Manchas | Dores |
| João | 1 | 1 | 0 | 1 |
| Pedro | 0 | 0 | 1 | 0 |
| Maria | 1 | 1 | 0 | 0 |
| José | 1 | 0 | 1 | 1 |
| Ana | 1 | 0 | 0 | 1 |
| Leila | 0 | 0 | 1 | 1 |

Os centroides iniciais são o João e o Pedro, respectivamente, e a distância usada é a Euclidiana.

|  |  |  |  |
| --- | --- | --- | --- |
| Nome | João (C1) | Pedro(C2) | Cluster |
| Maria | 1 | 1,73 | C1 |
| José | 1,41 | 1,41 | C1 |
| Ana | 1 | 1,73 | C1 |
| Leila | 1,73 | 1 | C2 |

Recalculando os centroides

C1 = João + Maria + José + Ana/4 = 1 0,5 0,25 0,75

C2 = Pedro + Leila/2 = 0 0 1 0,5

Com os novos centroides a matriz de distância é como segue:

|  |  |  |  |
| --- | --- | --- | --- |
| Nome | C1 | C2 | Cluster |
| João | 0,61 | 1,80 | C1 |
| Pedro | 1,53 | 0,5 | C2 |
| Maria | 1,19 | 1,8 | C1 |
| José | 0,93 | 1,11 | C1 |
| Ana | 0,61 | 1,5 | C1 |
| Leila | 1,36 | 0,0625 | C2 |

Cluster 1 = {João, Maria, José, Ana, C1}

Cluster 2 = {Pedro, Leila, C2}

Silhueta Completa

Calculamos a distância euclidiana de cada ponto com os membros do seu cluster e com os do cluster vizinho

**João**

b(João)

* Distância Euclidiana João -> Pedro = 2
* Distância Euclidiana João -> Leila = 1,73
* Distância Euclidiana João -> C2 = 1,80

b(João) = = 1,84

a(João)

* Distância Euclidiana João -> C1 = 0,61
* Distância Euclidiana João -> Maria = 1
* Distância Euclidiana João -> José = 1,41
* Distância Euclidiana João -> Ana = 1

a(João) = = 1,005

s(João) = = 0,45

**Maria**

b(Maria)

* Distância Euclidiana Maria-> Pedro = 1,73
* Distância Euclidiana Maria->Leila = 2
* Distância Euclidiana Maria->C2 = 1,80

b(Maria) = 1,84

a(Maria)

* Distância Euclidiana Maria->João = 1
* Distância Euclidiana Maria->José = 1,73
* Distância Euclidiana Maria->Ana = 1,41
* Distância Euclidiana Maria->C1 = 0,93

a(Maria) = 1,26

s(Maria) = 0,31

**José**

b(José)

* Distância Euclidiana José->Pedro = 1,41
* Distância Euclidiana José->Leila = 1
* Distância Euclidiana José->C2 = 1,11

b(José) = 1,17

a(José)

* Distância Euclidiana José->João = 1,41
* Distância Euclidiana José->Maria = 1,73
* Distância Euclidiana José->Ana = 1
* Distância Euclidiana José->C1 = 0,93

a(José) = 1,26

s(José) = -0,07

**Ana**

b(Ana)

* Distância Euclidiana Ana->Pedro = 1,73
* Distância Euclidiana Ana->Leila = 1,41
* Distância Euclidiana Ana->C2 = 1,5

b(Ana) = 1,54

a(Ana)

* Distância Euclidiana Ana->João = 1
* Distância Euclidiana Ana->Maria = 1,41
* Distância Euclidiana Ana->José = 1
* Distância Euclidiana Ana->C1 = 0,61

a(Ana) = 1,005

s(Ana) = 0,34

**Pedro**

b(Pedro)

* Distância Euclidiana Pedro->João = 2
* Distância Euclidiana Pedro->Maria = 1,73
* Distância Euclidiana Pedro->José = 1,41
* Distância Euclidiana Pedro->Ana = 1,73
* Distância Euclidiana Pedro->C1 = 1,36

b(Pedro) = 1,64

a(Pedro)

* Distância Euclidiana Pedro->Leila = 1
* Distância Euclidiana Pedro->C2 = 0,5

a(Pedro) = 0,75

s(Pedro) = 0,54

**Leila**

b(Leila)

* Distância Euclidiana Leila->João = 1,73
* Distância Euclidiana Leila->Maria = 2
* Distância Euclidiana Leila->José = 1
* Distância Euclidiana Leila->Ana = 1,41
* Distância Euclidiana Leila->C1 = 1,36

b(Leila) = 1,5

a(Leila)

* Distância Euclidiana Leila->Pedro=1
* Distância Euclidiana Leila->C2=0,5

a(Leila) = 0,75

s(Leila) = 0,5

SWC = =0,34

Silhueta Simplificada

**João**

b(João) = 1,80

a(João) = 0,61

s(João) = 0,66

**Maria**

b (Maria) = 1,80

a(Maria) = 0,93

s(Maria) = 0,48

**José**

b(José) = 1,11

a(José) = 0,93

s(José) = 0,16

**Ana**

b(Ana) = 1,5

a(Ana) = 0,61

s(Ana) = 0,59

**Pedro**

b(Pedro) = 1,36

a(Pedro) = 0,5

s(Pedro) = 0,63

**Leila**

b(Leila) = 1,36

a(Leila) = 0,5

s(Leila) = 0,63

SSWC = = 0,52

1. Considere 10 pontos em um espaço de 2 dimensões e considere que você executou o k-means com k = 3 e o resultado produzido pelo agrupamento é mostrado na figura a seguir.

Desenho com traços pretos em fundo branco

Descrição gerada automaticamente com confiança baixa

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Pontos | X | Y | C1 | C2 | C3 | Cluster |
| A | 1 | 7 | 1,41 | 4,24 | 4,12 | C1 |
| B | 2 | 7 | 1 | 3,60 | 3,16 | C1 |
| C | 6 | 6 | 4,12 | 3,60 | 1,41 | C3 |
| D | 3 | 5 | 1,41 | 1,41 | 2,23 | C2 |
| E | 4 | 5 | 2,23 | 2,23 | 1,41 | C3 |
| F | 3 | 4 | 1,41 | 3,16 | 1,0 | C3 |
| G | 7 | 3 | 5,09 | 5,09 | 2,23 | C3 |
| H | 1 | 2 | 1,41 | 4,24 | 4,12 | C1 |
| I | 6 | 2 | 4,12 | 4,12 | 1,41 | C3 |
| J | 3 | 1 | 1,41 | 3.16 | 2,23 | C1 |
| C1 | 2 | 3 | 0 | 3,60 | 3,16 | C3 |
| C2 | 4 | 7 | 3,60 | 0 | 1,41 | C3 |
| C3 | 5 | 4 | 3,16 | 1,41 | 0 | C2 |

Calcule a silhueta e silhueta simplificada

Silhueta

(A)

* Distância Euclidiana A->B = 1
* Distância Euclidiana A->H = 5
* Distância Euclidiana A->J = 6,32
* Distância Euclidiana A->C1 = 4,12

(A) = 4,11

1. Procure na internet por pelo menos um exemplo de problema real em que seria interessante aplicar regras de associação, exceto o exemplo do supermercado.

Na World Wide Web é preciso identificar os documentos com maior semelhança, facilitando a sua recuperação toda vez que o usuário faz uma consulta. Tradicionalmente isso é feito comparando as palavras usadas na consulta com os termos de indexação dos documentos. Este processo pode ser mais eficiente usando as técnicas de agrupamento que reunirá documentos com maior semelhança entre si e os retornará quando o seu grupo for consultado.

1. Dado a base de dados de transações

|  |  |
| --- | --- |
| Id Transação | Itens comprados |
| 1 | {a,d,e} |
| 24 | {a,b,c,e} |
| 12 | {a,b,d,e} |
| 31 | {a,c,d,e} |
| 15 | {b,c,e} |
| 22 | {b,d,e} |
| 29 | {c,d} |
| 40 | {a,b,c} |
| 33 | {a,d,e} |
| 38 | {a,b,e} |

1. O suporte dos itemsets {e}, {b,d}, {b,d,e}

Suporte do itemset {e} =

Suporte do itemset {b,d} =

Suporte do itemset {b,d,e} =

1. A confiança das regras: {b,d} -> {e}, {e} ->{b,d}. A confiança é uma medida simétrica?

Confiança de {b,d} -> {e} = = 2/2 = 1

Confiança de {e}->{b,d} = = 2/8 = ¼

A confiança não é uma medida simétrica, implicando , por exemplo, não significa que implica em .