I Neha Moolchandani declare that I have completed this assignment completely and entirely on my own, without any consultation with others.  I understand that any breach of the UAB Academic Honor Code may result in severe penalties.

**Chap 2:**

**PART I: Exercises 2.6 (20 pts), 2.8 (50 pts (30 pts for (a) and 20 pts for (b).) For the normalization step in (b), you should normalize each value by dividing it by the length (Euclidean norm) of that data point. Do not forget to normalize the query point as well.Text, letter

Description automatically generated**

**a)** d = √[ (p1 – q1 )^2 + (p2 – q2 )^2].

d = √[ (22-20)^2 + (1-0)^2 + (42-36)^2 + (10-8)^2 ]

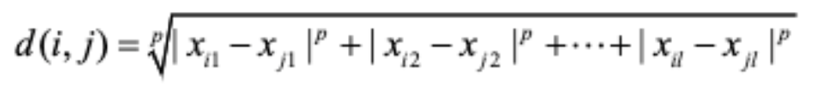
d= √[ (2)^2 + (1)^2 + (6)^2 + (2)^2 ]

d=√45 => 6.7082

**b)** d =  **|**p1 – q2| + |p1 – q2|

d= **|**22-20| + |1-0| + |42-36| +|10-8| = 11

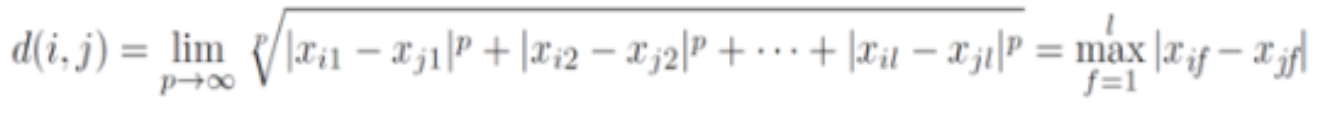
**c)**



d= 3√[ |22-20|^3 + |1-0|^3 + |42-36|^3 + |10-8|^3 ]

d=√233 = 6.15434

**d)**



d=|42-36| => 6

**Table

Description automatically generated**

**a)**

**Euclidean:**

d = √[ (p1 – q1 1 )^2 + (p2 – q2 )^2].

X1: d = √[ (1.4-1.5)^2 + (1.6-1.7)^2] = 0.141

X2: d = √[ (1.4-2)^2 + (1.6-1.9)^2 = 0.671

X3: d = √[ (1.4-1.6)^2 + (1.6-1.6)^2 = 0.283

X4: d = √[ (1.4-1.2)^2 + (1.6-1.5)^2 = 0.224

X5: d = √[ (1.4-1.5)^2 + (1.6-1.0)^2 = 0.608

Ranking based on **Euclidean**: x1,x4,x3,x5,x2

**Manhattan:**

d =  **|**p1 – q2| + |p1 – q2|

X1: d= **|**1.4-1.5| + |1.6-1.7| = 0.2

X2: d= **|**1.4-2| + |1.6-1.9| = 0.9

X3: d= **|**1.4-1.6| + |1.6-1.8| = 0.4

X4: d= **|**1.4-1.2| + |1.6-1.5| = 0.3

X5: d= **|**1.4-1.5| + |1.6-1.0| = 0.7

Ranking based on **Manhattan**: x1, x4,x3,x5,x2

**Supremum:**

X1: d = |1.4-1.5| or |1.6-1.7| => 0.1

X2: d = |1.4-2| or |1.6-1.9| => 0.6

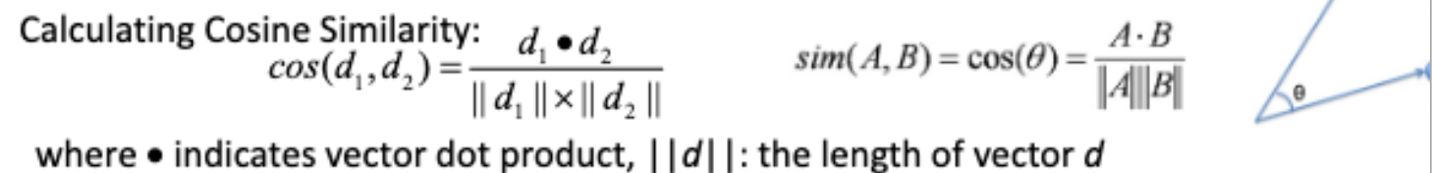
X3: d = |1.4-1.6| or |1.6-1.8| => 0.2

X4: d = |1.4-1.2| or |1.6-1.5| => 0.2

X5: d = |1.4-1.5| or |1.6-1.0| => 0.6

Ranking based on **Supremum**: x1, x4,x3,x5,x2

**Cosine Similarity:**



x: (1.4,1.6)

y: (1.5, 1.7) => (1.4x1.5 + 1.6x1.9) / (√[ ((1.4)^2+(1.6)^2 x √[ ((1.5^2)+(1.7^2))

* 4.189/4.791 => 0.99

x: (1.4,1.6)

y: (2.0, 1.9) => (1.4x2 + 1.6x1.7) / (√[ ((1.4)^2+(1.6)^2 x √[ ((2.0^2)+(1.9^2))

* 5.84/5.89 => 0.9849

x: (1.4,1.6)

y: (1.6, 1.8) => (1.4x1.6 + 1.6x1.8) / (√[ ((1.4)^2+(1.6)^2 x √[ ((1.6^2)+(1.8^2))

* 5.12056/5.12001 => 0.9999

x: (1.4,1.6)

y: (1.2, 1.5) => (1.4x1.6 + 1.2x1.5) / (√[ ((1.4)^2+(1.6)^2 x √[ ((1.2^2)+(1.5^2))

* 4.04/4.08396 => 0.992

x: (1.4,1.6)

y: (1.5, 1.0) => (1.4x1.6 + 1.5x1.0) / (√[ ((1.4)^2+(1.6)^2 x √[ ((1.5^2)+(1.0^2))

* 3.74/3.83275 => 0.975799

Ranking based on **Cosine Similarity**: x1, x4,x3,x2,x5

**b) Normalizing the data to make norm of each data equal to 1.**

Diagram

Description automatically generated

Range of Values: Lengh of a query point (1.4,1.6) use squareroot(x1sqr+x2sqr) , divide the resulting length with each point to get normalized form of that point.

Length of Query Point: (1.4,1.6) = √[1.4^2 + 1.6^2] = 2.126

|  |  |  |
| --- | --- | --- |
|  | A1 | A2 |
| X1 | 1.5/2.23 = 0.6726 | 1.7/2.23 = 0.76233 |
| X2 | 2/2.193 = 0.9119 | 1.9/2.193 = 0.6884 |
| X3 | 1.6/2.41 = 0.6639 | 1.8/2.41 = 0.7469 |
| X4 | 1.2/1.92 = 0.625 | 1.5/1.92 = 0.78125 |
| X5 | 1.5/1.8027 = 0.8321 | 1/1.92 = 0.5208 |
| X(1.4,1.6) | 1.4/2.126 = 0.6585 | 1.6/1.8027 = 0.7526 |

Using Euclidean Distance: Respect to Query Point: (1.4,1.6)

d = √[ (p1 – q1 )^2 + (p2 – q2 )^2].

X1: d =√[ (0. 6639-0. 6726)^2 + (0. 76233-0. 7526)^2] = 0.0786

X2: d =√[ (0. 9119-0.6585)^2 + (0. 6884-0. 7526)^2] = 0.0299

X3: d =√[ (0. 6639-0.6585)^2 + (0. 8664-0. 7526)^2] = 0.0818

X4: d = √[ (0. 625-0.6585)^2 + (0. 7469-0.7526)^2] = 0.1487

X5: d = √[ (0. 8321-0.6585)^2 + (0. 78125-0.7526)^2] = 0.26286

Ranking: x2, x3, x1, x4, x5

**PART II: For the same data from 2.8, do z-score normalization (see Lecture 2: Mean Absolute Deviation approach) for each feature dimension (A1 and A2, respectively). Please include the new point *x*(1.4, 1.6) in the calculation of mean and mean absolute deviation. Use Euclidean distance on the normalized data to rank the data points in the ascending order of their distance values. (10 pts for getting the z-scores correctly, and 10 points for getting the rank correctly.)**

**Diagram, schematic

Description automatically generated**

Mean: (1.0+ 1.2+1.5+1.5+1.5+1.6+1.7+1.8+1.9+2.0+1.4+1.6) / 12 = 1.55

Finding all distances per point:

((1.5 - 1.53) + ( 2 - 1.53) + ( 1.6 - 1.53) + ( 1.2 - 1.53) + ( 1.5 - 1.53) + ( 1.4 - 1.53) +

A2: (( 1.0 - 1.583) + ( 1.5 - 1.583) + ( 1.7 - 1.583) + ( 1.8 - 1.583) + ( 1.9 - 1.583) + ( 1.6 - 1.583)) / 6

Add all Distances up: MAD

MAD: A1: 0.1778

Mean A1: 1.533

MAD A2: 0.22

Mean A2: 1.583

(MEAN,MAD) and (A1,A2)

**Z-Score: Score-Mean/Deviation**

|  |  |  |
| --- | --- | --- |
|  | A1 | A2 |
| X1 | (1.5-1.53)/0.1778 = -0.169 | (1.7-1.58)/0.22 = 0.5454 |
| X2 | (2-1.53)/ 0.1778 = -2.64 | (1.9-1.58)/0.22 = 1.4545 |
| X3 | (1.6-1.53)/ 0.1778 = 0.3937 | (1.8-1.58)/0.22 = 1.0 |
| X4 | (1.2-1.53)/ 0.1778 = -1.856 | (1.5-1.58)/0.22 = -0.3636 |
| X5 | (1.5-1.53)/ 0.1778 = -0.169 | (1.0-1.58)/0.22 = -2.636 |

MAD: A1: 0.1778

Mean A1: 1.533

MAD A2: 0.22

Mean A2: 1.583

(MEAN,MAD) and (A1,A2)

MAD Together: 0.2083

MEAN Together: 1.558

(MEAN,MAD) (A1,A2)

Using Euclidean Distance: Respect to Query Point: (1.4,1.6)

d = √[ (p1 – q1 )^2 + (p2 – q2 )^2].

X1: d =√[ (1.558+0. 169)^2 + (0. 2083-0. 5454)^2/] = 1.759

X2: d =√[ (1.558+2.64)^2 + (0. 2083-0. 4545)^2/] = 4.207

X3: d =√[ (1.558-0.3937)^2 + (0. 2083-1)^2] = 1.4099

X4: d = √[ (1.558+1.859)^2 + (0. 2083+0. 3636)^2] = 3.4646

X5: d = √[ (1.558 +0. 169)^2 + (0. 2083+2.636)^2] = 3.328

Ranking: x3, x1, x5, x4, x2

**PART III: Read Section 2.4.6 and calculate the distance between (X1, X2) and that between (X1, X3) according to the data in the table below. Please include the details of your calculation. (20 pts)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A1 (NOMINAL) | A2 (NUMERIC) | A3  (ORDINAL) | A4 (Asymmetric Binary) | A5 (Asymmetric Binary) |
| X1 | A | 100 | Small | 0 | 1 |
| X2 | B | 20 | Medium | 1 | 0 |
| X3 | C | 50 | Large | 1 | 1 |

**Nominal Attributes(A1) P-M/P**

P is attributes and M is Matches

P=1

A picture containing text, clock

Description automatically generatedD(I,j) = 0 (if objects i and j match) otherwise 1

**Numeric Attributes (A2)**

Max = 100 and min = 20

D(i,j) = (i-j)/max-min

vi-min/max-min

**Ordinal Attributes(A3)**

3 states thus M(f) =3

Small = 1, M=2, L=3

Normalize: x-xMin/xmin-xMax

Small = 1-1/1-1 = 0

Medium = 2-1/1-3 = 0.5

Large = 3-1/3-1 = 1

Calculate using Manhattan:

|x2-x1| + |y2-y1|

D(x1,x2) = |0-0.5| = 0.5

D(x2,x3) = |0.5-1| = 0.5

D(x1,x3) = |0-1| = 1

A picture containing calendar

Description automatically generated

**Assymetric Binary (A4)**

Graphical user interface, text, application

Description automatically generated

Computing Dissimilarity for all attributes: A picture containing text

Description automatically generated

Text, letter

Description automatically generated

**Text, letter

Description automatically generated**Text, letter

Description automatically generated