**Name: Nikhil Jadhav**

**Student ID: 801075504**

Chapter 2 Exercise 19

Exercise 19: For the following vectors, x and y, calculate the indicated similarity or distance measures.

(a) x = (1 ,1,1,1), y = (2 ,2,2,2) cosine, correlation, Euclidean

(b) x = (0 ,1,0,1), y = (1 ,0,1,0) cosine, correlation, Euclidean, Jaccard

(c) x = (0 ,−1,0,1), y = (1 ,0,−1,0) cosine, correlation, Euclidean

(d) x = (1 ,1,0,1,0,1), y = (1 ,1,1,0,0,1) cosine, correlation, Jaccard

(e) x = (2 ,−1,0,2,0,−3), y =(−1,1,−1,0,0,−1) cosine, correlation

Answers:

1. **Cosine**

x ● y = 1\*2 + 1\*2 + 1\*2 + 1\*2 = 8

||x|| = sqrt(1\*1 + 1\*1 + 1\*1 + 1\*1) = sqrt (4) = 2

||y|| = sqrt(2\*2 + 2\*2 + 2\*2 + 2\*2) = sqrt (16) = 4

cos(x,y) = (x ● y) / (||x||\*||y||) = (8)/ (2\*4)

cos(x,y) = 1

**Correlation**

corr(x, y) = [covariance(x,y)] / [standard deviation(x) \* standard deviation(y)]

Mean of x =(1+1+1+1) / 4 = 1

Mean of y = (2+2+2+2) / 4 = 2

covariance(x,y) = 1/(4 -1) [(1-1)(2-2) + (1-1)(2-2) + (1-1)(2-2) + (1-1)(2-2)] = 0

Standard deviation (x) = sqrt[((1/(4-1))) \* {(1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2}] = sqrt[(1/3) \* 0] = 0

Standard deviation (y) = sqrt[((1/(4-1))) \* {(2-2)^2 + (2-2)^2 + (2-2)^2 + (2-2)^2}] = sqrt[(1/3) \* 0] = 0

corr(x,y) = 0/0 = undefined

**Euclidean**

d(x, y) = sqrt((1-2)^2 + (1-2)^2 + (1-2)^2 + (1-2)^2)

Euclidean distance = 2

1. **Cosine**

x ● y = 0\*1 + 1\*0 + 0\*1 + 1\*0 = 0

||x|| = sqrt(0\*0 + 1\*1 + 0\*0 + 1\*1) = sqrt (2)

||y|| = sqrt(1\*1 + 0\*0 + 1\*1 + 0\*0) = sqrt (2)

cos(x,y) = (x ● y) / (||x||\*||y||) = (0)/ (sqrt (2) \*sqrt (2))

cos(x,y) = 0

**Correlation**

corr(x, y) = [covariance(x,y)] / [standard deviation(x) \* standard deviation(y)]

Mean of x = (0+1+0+1) / 4 = ½ = 0.5

Mean of y = (1+0+1+0) / 4 = ½ = 0.5

Covariance(x,y) = 1/(4-1) \* [(0-½)(1-½) + (1-½)(0-½) +(0-½)(1-½) +(1-½)(0-½) ]

Covariance(x,y) = (1/3) \* [(-1/4) + (-1/4) + (-1/4) + (-1/4)]

Covariance(x,y) = -1/3

Standard\_deviation (x) = sqrt[((1/(4-1))) \* {(1-1/2)^2 + (0-1/2)^2 + (1-1/2)^2 + (0-1/2)^2}] = sqrt[(1/3) \* 1] = 0.57735

Standard\_deviation (y) = sqrt[((1/(4-1))) \* {(0-1/2)^2 + (1-1/2)^2 + (0-1/2)^2 + (1-1/2)^2}] = sqrt[(1/3) \* 1] = 0.57735

Corr(x,y) = (-1/3) / (0.57735 \* 0.57735)

Corr(x,y) = -1

**Euclidean**

d(x, y) = sqrt((0-1)^2 + (1-0)^2 + (0-1)^2 + (1-0)^2)

Euclidean distance = 2

**Jaccard**

J= (numbar of matching presences) / (number of attributes not involved in 00 matches)

J=(f11)/(f01 + f10 + f11)

f01 = 2 the number of attributes where x was 0 and y was 1

f10 = 2 the number of attributes where x was 1 and y was 0

f00 = 0 the number of attributes where x ws 0 and y was 0

f11 = 0 the number of attributes where x was 1 and y was 1

J= (0) / (2 + 2 + 0)

J = 0

1. **Cosine**

x ● y = 0\*1 + (-1)\*0 + 0\*(-1) + 1\*0 = 0

||x|| = sqrt(0\*0 + (-1)\*(-1) + 0\*0 + 1\*1) = sqrt (2)

||y|| = sqrt(1\*1 + 0\*0 + (-1)\*(-1) + 0\*0) = sqrt (2)

cos(x,y) = (x ● y) / (||x||\*||y||) = (0)/ (sqrt (2) \*sqrt (2))

cos(x,y) = 0

**Correlation**

corr(x, y) = [covariance(x,y)] / [standard deviation(x) \* standard deviation(y)]

Mean of x = (0+(-1)+0+1) / 4 = 0

Mean of y = (1+0+(-1)+0) / 4 = 0

Covariance(x,y) = 1/(4-1) \* [(0-0)(1-0) + (-1-0)(0-0) +(0-0)(-1-0) +(1-0)(0-0) ] = (1/3) \* 0 = 0

corr(x,y) = 0

**Eulidean**

d(x, y) = sqrt((0-1)^2 + (-1-0)^2 + (0+1)^2 + (1-0)^2)

Euclidean distance = 2

1. **Cosine**

x ● y = 1\*1 + 1\*1 + 0\*1 + 1\*0 + 0\*0 + 1\*1 = 3

||x|| = sqrt(1\*1 + 1\*1 + 0\*0 + 1\*1 + 0\*0 + 1\*1) = 2

||y|| = sqrt(1\*1 + 1\*1 + 1\*1 + 0\*0 + 0\*0 + 1\*1) = 2

cos(x,y) = (x ● y) / (||x||\*||y||) = (3)/ (2 \* 2)

cos(x,y) = ¾ = 0.75

**Correlation**

corr(x, y) = [covariance(x,y)] / [standard deviation(x) \* standard deviation(y)]

Mean of x = (1+1+0+1+0+1) / 6 = 4/6

Mean of y = (1+1+1+0+0+1) / 6 = 4/6

Covariance(x,y) = 1/(6-1) \* [(1-4/6)(1-4/6) + (1-4/6)(1-4/6) +(0-4/6)(1-4/6) +(1-4/6)(0-4/6) + (0-4/6)(0-4/6) + (1-4/6)(1-4/6) ] = (1/5)(1/3)=1/15

Standard\_deviation (x) = sqrt[((1/(6-1))) \* {(1-4/6)^2 + (1-4/6)^2 + (0-4/6)^2 + (1-4/6)^2 + (0-4/6)^2 + (1-4/6)^2}] = sqrt[(1/5) \* (4/3)] = 0.5164

Standard\_deviation (y) = sqrt[((1/(6-1))) \* {(1-4/6)^2 + (1-4/6)^2 + (1-4/6)^2 + (0-4/6)^2 + (0-4/6)^2 + (1-4/6)^2}] = sqrt[(1/5) \* (4/3)] = 0.5164

Corr(x,y) = (1/15) / (0.5164 \* 0.5164)

Corr(x,y) = 0.25

**Jaccard**

J= (numbar of matching presences) / (number of attributes not involved in 00 matches)

J=(f11)/(f01 + f10 + f11)

f01 = 1 the number of attributes where x was 0 and y was 1

f10 = 1 the number of attributes where x was 1 and y was 0

f00 = 1 the number of attributes where x ws 0 and y was 0

f11 = 3 the number of attributes where x was 1 and y was 1

J= (3) / (1 + 1 + 3)

J = 3/5 = 0.6

1. **Cosine**

x ● y = 2\*(-1) + (-1)\*1 + 0\*(-1) + 2\*0 + 0\*0 + (-3)\*(-1) = 0

||x|| = sqrt(2\*2 + (-1)\*(-1) + 0\*0 + 2\*2 + 0\*0 + (-3)\*(-3)) = sqrt (18)

||y|| = sqrt((-1)\*(-1) + 1\*1 + (-1)\*(-1) + 0\*0 + 0\*0 + (-1)\*(-1)) = 2

cos(x,y) = (x ● y) / (||x||\*||y||) = (0)/ (sqrt (18) \* 2)

cos(x,y) = 0

**Correlation**

corr(x, y) = [covariance(x,y)] / [standard deviation(x) \* standard deviation(y)]

Mean of x = (2+(-1)+0+2+0+(-3)) / 6 = 0

Mean of y = ((-1)+1+(-1)+0+0+(-1)) / 6 = -1/6

Covariance(x,y) = 1/(6-1) \* [(2-0)(-1+1/6) + (-1-0)(1+1/6) +(0-0)(-1+1/6) + (2-0)(0+1/6) + (0-0)(0+1/6) + (-3-0)(-1+1/6) ] = (1/5) \* 0 = 0

Corr(x,y) = 0