9.2.1

Feature	A	B	C
Processor Speed	3.06	2.68	2.92
Disk Size	500	320	640
Main-Memory Size	6	4	6

a) For two vectors, x & y Cosine of the angle between them is dot product of x.y divided by L_2 - norms of x and y (i.e., their Euclidean distances from the origin).

$$\therefore \cos(A,B) = \frac{3.06 \times 2.68 + 500\alpha * 320\alpha + 6\beta * 4\beta}{\sqrt{(3.06)^2 + (500\alpha)^2 + (6\beta)^2}\ \sqrt{(2.68)^2 + (320\alpha)^2 + (4\beta)^2}}$$

$$= \frac{8.2008 + 160000\alpha^2 + 24\beta^2}{\sqrt{9.3636 + 250000\alpha^2 + 36\beta^2}\ \sqrt{7.1824 + 102400\alpha^2 + 16\beta^2}}$$

Similarly $\cos(B,C) = \frac{7.8256 + 204800\alpha^2 + 24\beta^2}{\sqrt{7.1824 + 102400\alpha^2 + 16\beta^2}\ \sqrt{8.5264 + 409600\alpha^2 + 36\beta^2}}$

Similarly

$$\cos(A,C) = \frac{8.9352 + 320000\alpha^2 + 36\beta^2}{\sqrt{9.3636 + 250000\alpha^2 + 36\beta^2}\ \sqrt{8.5264 + 409600\alpha^2 + 36\beta^2}}$$

Scanned by CamScanner

b) When $\alpha = B = 1$

$$\cos(A,B) = \frac{8.2008 + 160000 + 24}{\sqrt{250045.36}\ \sqrt{102423.18}}$$

$$= \frac{160032.2001}{500.04 \times 320.036} \simeq 1$$

if the value of $\cos(A,B) \simeq 1$
then the angle is 0°

Similarly, $\cos(B,C) = \frac{2.04831.826}{\sqrt{102423.18} \times \sqrt{409644.526}}$

$$= \frac{2.04831.826}{320.036 \times 640.03}$$

$$\simeq 1$$

$\therefore$ the angle is 0°

Similarly $\cos(A,C) = \frac{320044.935}{500.04 \times 640.03} \simeq 1$

$\therefore$ the angle is 0°

Scanned by CamScanner

c) $\alpha = 0.01$ & $\beta = 0.5$

$$\text{Cos}(A,B) = \frac{8.2008 + 160000 \times (0.01)^2 + 24 \times (0.5)^2}{\sqrt{9.3636 + 250000(0.01)^2 + 36 \times (0.5)^2}\ \sqrt{7.1824 + 102400(0.01)^2 + 16 \times (0.5)^2}}$$

$$= \frac{30.2008}{\sqrt{43.3636} \times \sqrt{21.4224}} = \frac{30.20}{30.47}$$

$$\text{Cos}(A,B) = 0.9911$$

The angle between Cos(A,B) will be = 7.64°

Similarly $$\text{Cos}(B,C) = \frac{7.825 + 204800 \times (0.01)^2 + 24 \times (0.5)^2}{\sqrt{7.1824 + 102400(0.01)^2 + 16 \times (0.5)^2} \times \sqrt{8.5264 + 409600(0.0001) + 36 \times 0.25}}$$

$$= \frac{34.305}{35.20} = 0.974$$

The angle between B, C will be $\approx 12^\circ$

Similarly $$\text{Cos}(A,C) = \frac{8.9352 + 320000 \times 0.0001 + 36 \times 0.25}{\sqrt{9.8.5264 + 409600(0.0001) + 36 \times 0.25}\ \sqrt{9.3636 + 250000(0.01)^2 + 36 \times 0.5^2}}$$

$$= \frac{49.9352}{50.3554}$$

$$\simeq 0.9916$$

The angle between A & C will be 7.43°

Scanned by CamScanner

a)

$\alpha = 0.009$

$\beta = 0.004$

Now $\cos(A, B) = 0.988$

$\cos(B, C) = 0.96$

$\cos(A, C) = 0.99$

Scanned by CamScanner

9.2.3

a) Average = (4+2+5)/3 = 11/3

$A = 4 - \frac{11}{3} = \frac{1}{3}$

$B = 2 - \frac{11}{3} = \frac{-5}{3}$

$C = 5 - \frac{11}{3} = \frac{4}{3}$

b) Processor Speed = $3.06 \times \frac{1}{3} - 2.68 \times \frac{5}{3} + 2.92 \times \frac{4}{3}$

$= 0.4467$

Disk size = $500 \times \frac{1}{3} - 320 \times \frac{5}{3} + 640 \times \frac{4}{3}$

$= 486.6667$

Main Memory = $6 \times \frac{1}{3} - \frac{4 \times 5}{3} + 6 \times \frac{4}{3}$

$= 3.333$

Scanned by CamScanner

9.3.1

	a	b	c	d	e	f	g	h
A	4	5		5	1		3	2
B		3	4	3	1	2	1	
C	2		1	3		4	5	3

a)

	a	b	c	d	e	f	g	h
A	1	1	0	1	1	0	1	1
B	0	1	1	1	1	1	1	0
C	1	0	1	1	0	1	1	1

Boolean Utility Matrix

Jaccard distance (A,B) = $\frac{1+1+1+1}{8} = \frac{4}{8} = 0.5$

Jaccard distance (B,C) = $\frac{1+1+1+1}{8} = 0.5$

Jaccard distance (A,C) = $\frac{1+1+1+1}{8} = 0.5$

b)

Cosine Distance (A,B) = $\frac{(1*0+1*1+0*1+1*1+1*1+0*1+1*1+1*0)}{\sqrt{1^2+1^2+0^2+1^2+0^2+1^2+1^2}\ \sqrt{0^2+1^2+1^2+1^2+1^2+1^2+1^2+0^2}}$

$= \frac{4}{\sqrt{6} \times \sqrt{6}} = \frac{4}{6} = 0.6$

Cosine (B,C) = $\frac{11}{8\sqrt{10}} = 0.435$

Cos (A,C) = $\frac{11}{8\sqrt{5}} = 0.615$

Scanned by CamScanner

Date __/__/__

c)

	a	b	c	d	e	f	g	h
A	1	1	0	1	0	0	0	0
B	0	1	1	1	0	0	0	0
C	0	0	0	1	0	1	1	1

Jaccard Similarity (A,B) = 2/5
Jaccard Distance (A,B) = 1 − 2/5 = 3/5

Jaccard Similarity (A,C) = 2/6
Jaccard Distance (A,C) = 1 − 2/6 = 2/3

Jaccard Similarity (B,C) = 1/6
Jaccard Distance (B,C) = 1 − 1/6 = 5/6

d) Cosine Distance (A,B) = 0.423
Cos distance (B,C) = 0.712
Cosine Distance (A,C) = 0.5

e) ~~Avg value of cosine~~ =
Matrix after normalization

	a	b	c	d	e	f	g	h
A	0.67	1.67	0	1.67	-1.33	0	-033	-1.33
B	0	0.67	1.67	0.67	-1.33	-0.33	-1.33	0
C	-1	0	-2	0	0	1	2	0

f) cos (A,B) = 0.584
cos (A,C) = -0.1154
cos (B,C) = -0.7395

Scanned by CamScanner

9.4.1 a) U & V

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & x \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 \\ 1+x & 1+x & 1+x & 1+x & 1+x \\ 2 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 \end{bmatrix}$$

The contribution to the sum of squares from the third row is

$$(x-1)^2 + (x-2)^2 + x^2 + (x-3)^2$$

We find the min value of this expression by differentiating & equating to 0 as:

$$2 \times ((x-1) + (x-2) + x + (x-3)) = 0$$

$$\therefore \quad x = 1.5$$

$$\therefore \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1.5 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 \\ 2.5 & 2.5 & 2.5 & 2.5 & 2.5 \\ 2 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 \end{bmatrix}$$

Scanned by CamScanner

b)
$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & y & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 2 & y+1 & 2 \\ 2 & 2 & 2 & y+1 & 2 \\ 2 & 2 & 2 & y+1 & 2 \\ 2 & 2 & 2 & y+1 & 2 \\ 2 & 2 & 2 & y+1 & 2 \end{bmatrix}$$

The contribution to the sum of squares from the fourth column is

$$(y-3)^2 + (y-3)^2 + y^2 + (y-2)^2 + (y-3)^2 .$$

The min value of this expression by differentiating & equating to 0 as:

$$2 \times ((y-3) + (y-3) + y + (y-2) + (y-3)) = 0$$

$$y = 2.2$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 1 & 1 & 2.2 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 2 & 3.2 & 2 \\ 2 & 2 & 2 & 3.2 & 2 \\ 2 & 2 & 2 & 3.2 & 2 \\ 2 & 2 & 2 & 3.2 & 2 \\ 2 & 2 & 2 & 3.2 & 2 \end{bmatrix}$$

Scanned by CamScanner

Practicum Problem 1.

Load the Movie lens 100k dataset (ml-100k.zip) into Python using Pandas data frames. Build a user profile on unscaled data for both users 200 and 15, and calculate the cosine similarity and distance between the user's preferences and the item/movie 95. Which user would a recommender system suggest this movie to?

Answer.

The first goal of the problem is to make a user profile on unscaled data for both the users 200 and 15. There is need of similarity in the structure so that a similarity measure could be implemented to find similarity between users and movie.

A dummy user profile is created, it is created in which each feature has a specific rating and the mean is calculated for that. The mean is subtracted from the individual feature, so that the features with less values can be eliminated.

For creating the dummy matrix there can be three cases which are:

- Both the item profile and utility matrix as binary
- Item profile as Boolean and utility matrix has rating
- When both have the ratings.

Using the above three cases a utility matrix is created. After successfully creating the user profile now we have to see whom to recommend the movie id_95 out of the two. For that we will calculate the cosine distance and similarity for both the users.

User 200
cosine distance: 1.18
cosine similarity: -.182

User 15
cosine distance: .65
cosine similarity: .34

Now as we can see that the user 200 has more of cosine distance then that of user 15 so it will be prefer. Also, cosine similarity is more of the negative value that is it is less than that of user 15 so it would be suggested by the recommender system.

Practicum Problem 2.

Load the Movie lens 100k dataset (ml-100k.zip) into Python using Pandas data frames. Convert the ratings data into a utility matrix representation, and find the 10 most similar users for user 1 based on cosine similarity of the user ratings data. Based on the average of the ratings for item 508 from the similar users, what is the expected rating for this item for user 1?

Answer.

The Movie lens 100k dataset is loaded into the python using the Pandas data frames. The rating are converted into a utility matrix representation and the 10 most similar users for user 1 based on the cosine similarity of the user rating data is calculated. The expected rating for this item for user 1 should be 4.5 as out of the 10 most similar users. Two of the user got the rating of five and two users got the rating of 4, others are zero so the average will come around 4.5.

```
('Item: ', 738)
('Cosine Similarity ', array([[ 0.29148679]]))
('Item: ', 592)
('Cosine Similarity ', array([[ 0.27840172]]))
('Item: ', 276)
('Cosine Similarity ', array([[ 0.26815054]]))
('Item: ', 267)
('Cosine Similarity ', array([[ 0.26476147]]))
('Item: ', 643)
('Cosine Similarity ', array([[ 0.2640026]]))
('Item: ', 757)
('Cosine Similarity ', array([[ 0.26236785]]))
('Item: ', 457)
('Cosine Similarity ', array([[ 0.26233704]]))
('Item: ', 606)
('Cosine Similarity ', array([[ 0.26084701]]))
('Item: ', 916)
('Cosine Similarity ', array([[ 0.25562438]]))
('Item: ', 44)
('Cosine Similarity ', array([[ 0.2529544]]))
```

The expected rating of user 1 for item 508 would be 4.5.

```
print(np_array_similar_users.sum()/1)

Expected rating of user 1 for item 508:
4.5
```