

# CS 422 – Data Mining

## Homework 4

25 April 2017

Mayank Bansal

A20392482

# Recitation Problems

## Problem 9.2.1

a)

$$\cosine(A, B) = \frac{(8.2008 + 160000\alpha^2 + 24\beta^2)}{\sqrt{9.3636 + 250000\alpha^2 + 36\beta^2}\sqrt{7.1824 + 102400\alpha^2 + 16\beta^2}}$$

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

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

b)

$$\theta_{A,B} = \cos^{-1}\left(\frac{8.2008 + 160000 + 24}{\sqrt{9.3636 + 250000 + 36}\sqrt{7.1824 + 102400 + 16}}\right) = \mathbf{0.1323}$$

$$\theta_{B,C} = \cos^{-1}\left(\frac{7.8256 + 204800 + 24}{\sqrt{7.1824 + 102400 + 16}\sqrt{8.5264 + 409600 + 36}}\right) = \mathbf{0.2824}$$

$$\theta_{A,C} = \cos^{-1}\left(\frac{8.9352 + 320000 + 36}{\sqrt{9.3636 + 250000 + 36}\sqrt{8.5264 + 409600 + 36}}\right) = \mathbf{0.1748}$$

c)

$$\theta_{A,B} = \cos^{-1}\left(\frac{8.2008 + 160000(0.01)^2 + 24(0.5)^2}{\sqrt{9.3636 + 250000(0.01)^2 + 36(0.5)^2}\sqrt{7.1824 + 102400(0.01)^2 + 16(0.5)^2}}\right) \\ = \mathbf{7.7433}$$

$$\theta_{B,C} = \cos^{-1}\left(\frac{7.8256 + 204800(0.01)^2 + 24(0.5)^2}{\sqrt{7.1824 + 102400(0.01)^2 + 16(0.5)^2}\sqrt{8.5264 + 409600(0.01)^2 + 36(0.5)^2}}\right) \\ = \mathbf{14.2623}$$

$$\theta_{A,C} = \cos^{-1}\left(\frac{8.9352 + 320000(0.01)^2 + 36(0.5)^2}{\sqrt{9.3636 + 250000(0.01)^2 + 36(0.5)^2}\sqrt{8.5264 + 409600(0.01)^2 + 36(0.5)^2}}\right) \\ = \mathbf{7.4516}$$

d)

$$\text{Average Disk Size} = \frac{500 + 320 + 640}{3} = \frac{1460}{3}$$

$$\rightarrow \alpha = \frac{3}{1460} = 0.002$$

$$\text{Average Main - Memory Size} = \frac{6 + 4 + 6}{3} = \frac{16}{3}$$

$$\rightarrow \beta = \frac{3}{16} = 0.1875$$

$$\begin{aligned}
& \theta_{A,B} \\
&= \cos^{-1} \left( \frac{8.2008 + 160000(0.002)^2 + 24(0.1875)^2}{\sqrt{9.3636 + 250000(0.002)^2 + 36(0.1875)^2} \sqrt{7.1824 + 102400(0.002)^2 + 16(0.1875)^2}} \right) \\
&= \mathbf{6.011} \\
& \theta_{B,C} \\
&= \cos^{-1} \left( \frac{7.8256 + 204800(0.002)^2 + 24(0.5)^2}{\sqrt{7.1824 + 102400(0.002)^2 + 16(0.1875)^2} \sqrt{8.5264 + 409600(0.002)^2 + 36(0.1875)^2}} \right) \\
&= \mathbf{10.6385} \\
& \theta_{A,C} \\
&= \cos^{-1} \left( \frac{8.9352 + 320000(0.002)^2 + 36(0.5)^2}{\sqrt{9.3636 + 250000(0.002)^2 + 36(0.1875)^2} \sqrt{8.5264 + 409600(0.002)^2 + 36(0.1875)^2}} \right) \\
&= \mathbf{5.2611}
\end{aligned}$$

### Problem 9.2.3

a)

$$\text{Average Rating} = \frac{4 + 2 + 5}{3} = \frac{11}{3}$$

$$\text{User A Rating: } 4 - \frac{11}{3} = \frac{1}{3}$$

$$\text{User B Rating: } 2 - \frac{11}{3} = -\frac{5}{3}$$

$$\text{User C Rating: } 5 - \frac{11}{3} = \frac{4}{3}$$

b)

$$\text{Processor Speed: } 3.06 \times \frac{1}{3} - 2.68 \times \frac{5}{3} + 2.92 \times \frac{4}{3} = \mathbf{0.4467}$$

$$\text{Disk Size: } 500 \times \frac{1}{3} - 320 \times \frac{5}{3} + 640 \times \frac{4}{3} = \mathbf{486.6667}$$

$$\text{Main Memory Size: } 6 \times \frac{1}{3} - 4 \times \frac{5}{3} + 6 \times \frac{4}{3} = \mathbf{3.3333}$$

### Problem 9.3.1

Utility Matrix

	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

Rounded off Utility Matrix

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

a)

$$\text{Jaccard}(A,B) = \frac{4}{8} = \frac{1}{2}$$

$$Jaccard(B, C) = \frac{4}{8} = \frac{1}{2}$$

$$Jaccard(A, C) = \frac{4}{8} = \frac{1}{2}$$

b)

$$cosDist(A, B) = 1 - \frac{1 + 1 + 1 + 1}{\sqrt{6}\sqrt{6}} = \frac{1}{3}$$

$$cosDist(B, C) = 1 - \frac{1 + 1 + 1 + 1}{\sqrt{6}\sqrt{6}} = \frac{1}{3}$$

$$cosDist(A, C) = 1 - \frac{1 + 1 + 1 + 1}{\sqrt{6}\sqrt{6}} = \frac{1}{3}$$

c)

$$Jaccard(A, B) = \frac{3}{5}$$

$$Jaccard(B, C) = \frac{5}{6}$$

$$Jaccard(A, C) = \frac{4}{6}$$

d)

$$cosDist(A, B) = 1 - \frac{2}{2\sqrt{3}} = 0.422$$

$$cos(B, C) = 1 - \frac{1}{2\sqrt{3}} = 0.711$$

$$cos(A, C) = 1 - \frac{2}{2\sqrt{3}} = 0.422$$

Normalized Utility Matrix

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	
A	0.66	1.66		1.66	-2.33		-0.33	-1.33	<i>avg(A) = 20/6</i>
B		0.66	1.66	0.66	-1.33	-0.33	-1.33		<i>avg(B) = 14/6</i>
C	-1		-2	0		1	2	0	<i>avg(C) = 18/6</i>

e)

f)

$$cosDist(A, B) = 1 - 0.58 = 0.42$$

$$cosDist(B, C) = 1 - 0.039 = 0.961$$

$$cosDist(C, A) = 1 - 0 = 1$$

### Problem 9.4.1

$$M = \begin{bmatrix} 5 & 2 & 4 & \mathbf{4} & 3 \\ 3 & 1 & 2 & \mathbf{4} & 1 \\ \mathbf{2} & & \mathbf{3} & \mathbf{1} & \mathbf{4} \\ 2 & 5 & 4 & \mathbf{3} & 5 \\ 4 & 4 & 5 & \mathbf{4} & \end{bmatrix}$$

a)  $u_{32}$

From figure 9.10,

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

Sum of squares from the 3<sup>rd</sup> row:

$$f(x) = (2 - (x + 1))^2 + 0 + (3 - (x + 1))^2 + (1 - (x + 1))^2 + (4 - (x - 3))^2$$

To find minimum value of  $f(x)$ :

$$\frac{d(f(x))}{d(x)} = 0$$

$$\frac{d(f(x))}{d(x)} = -2(x - 1) - 2(x - 2) - 2(x) - 2(x - 3) = 0$$

Solving for  $x$ ,

$$x = \frac{3}{2}$$

Substituting  $x$ ,

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

b)  $v_{14}$

From figure 9.10,

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

Sum of squares from the 4<sup>th</sup> column:

$$f(y) = (4 - (y + 1))^2 + (4 - (y + 1))^2 + (1 - (y + 1))^2 + (3 - (y + 1))^2 + (4 - (y + 1))^2$$

To find minimum value of  $f(y)$ :

$$\frac{d(f(y))}{d(y)} = 0$$

$$\frac{d(f(y))}{d(y)} = -2(y - 3) - 2(y - 3) - 2y - 2(y - 2) - 2(y - 3) = 0$$

Solving for  $y$ ,

$$y = \frac{11}{5}$$

Substituting  $y$ ,

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

## Practicum Problems

### Problem 2.1

Cosine Similarity (user-200, movie-95) = 0.3875

Cosine Distance (user-200, movie-95) = 0.6125

Cosine Similarity (user-15, movie-95) = 0.2151

Cosine Distance (user-15, movie-95) = 0.7849

The system is more likely to recommend movie 95 to user 200.

### Problem 2.2

Expected rating for the item for user 1 is 4.0