CS422 HW6
SHIQI LIU
1.1 Chapter 9 Problems: 9.2.1,9.2.3,9.3.1,9.4.1

9.2.1
a)

$$\text{CosineDistance (A,B)} = 1 - \frac{(3.06*1)*(2.68*1) + (500*\alpha)*(320*\alpha) + (6*\beta)x(4*\beta)}{(\sqrt{|3.06|^2 + |500\alpha|^2 + |6\beta|^2})*(\sqrt{|2.68|^2 + |320\alpha|^2|} + |4\beta|^2)} = 1 - \frac{(8.2008) + (160000\alpha^2) + (24\beta^2)}{(\sqrt{9.3636 + 250000\alpha^2 + 36\beta^2})*(\sqrt{7.1824 + 102400\alpha^2 + 16\beta^2})}$$

$$\text{CosineDistance (B,C)} = 1 - \frac{(7.8256) + (204800\alpha^2) + (24\beta^2)}{(\sqrt{7.1824 + 102400\alpha^2 + 16\beta^2})*(\sqrt{8.5264 + 409600\alpha^2 + 36\beta^2})}$$

$$\text{CosineDistance (A,C)} = 1 - \frac{(8.2008) + (320000\alpha^2) + (36\beta^2)}{(\sqrt{9.3636 + 250000\alpha^2 + 32\beta^2})*(\sqrt{8.5264 + 409600\alpha^2 + 36\beta^2})}$$

b)

$$\text{CosineDistance (A,B)} = 1 - \frac{(8.2008) + (160000) + (24)}{(\sqrt{9.3636 + 250000 + 36})*(\sqrt{7.1824 + 102400 + 16})} = 0$$

$$\text{CosineDistance (B,C)} = 1 - \frac{(7.8256) + (204800) + (24)}{(\sqrt{7.1824 + 102400 + 16})*(\sqrt{8.5264 + 409600 + 36})} = 0$$

$$\text{CosineDistance (A,C)} = 1 - \frac{(8.2008) + (320000) + (36)}{(\sqrt{9.3636 + 250000 + 36})*(\sqrt{8.5264 + 409600 + 36})} = 0$$

c)

$$\text{CosineDistance (A,B)} = 1 - \frac{(8.2008) + 160000*(0.01)^2 + 24*(0.5)^2)}{(\sqrt{9.3636 + 250000*(0.01)^2 + 36*(0.5)^2})*(\sqrt{7.1824 + 102400*(0.01)^2 + 16*(0.5)^2})} = 1 - 0.99$$

$$\text{CosineDistance (B,C)} = 1 - \frac{(7.8256) + 204800*(0.01)^2 + 24*(0.5)^2}{(\sqrt{7.1824 + 102400*(0.01)^2 + 16*(0.5)^2})*(\sqrt{8.5264 + 409600*0.01^2 + 36^2})} = 1 - 0.97$$

$$\text{CosineDistance (A,C)} = 1 - \frac{(8.2008) + 320000*(0.01)^2 + 36*(0.5)^2}{(\sqrt{9.3636 + 250000*(0.01)^2 + 36*(0.5)^2})*(\sqrt{8.5264 + 409600*(0.01)^2 + 36*(0.5)^2})} = 1 - 0.99$$

d)

Scale factor for processor speed = 1/2.88

Scale factor for disk size = 1/486.6666

Scale factor for memory size = 1/5.333

$\text{CosineDistance (A,B)} = 1 -$
$$\frac{(3.06*1)*(2.68*1) + (500*\frac{1}{486.6666})*(320*\frac{1}{486.6666}) + (6*\frac{1}{5.333})x(4*\frac{1}{5.333})}{(\sqrt{|3.06|^2 + |500\frac{1}{486.6666}|^2 + |6\frac{1}{5.333}|^2})*(\sqrt{|2.68|^2 + |320\frac{1}{486.6666}|^2| + |4\frac{1}{5.333}|^2})}$$
$$\frac{(8.2008) + 160000(\frac{1}{486.6666})^2 + 24(\frac{1}{5.333})^2}{(\sqrt{9.3636 + 250000*(\frac{1}{486.6666})^2 + 36*(\frac{1}{5.333})^2})*(\sqrt{7.1824 + 102400*(\frac{1}{486.6666})^2 + 16*(\frac{1}{5.333})^2})}$$
$= 1 -$

$= 0.99$

Angle between A & B = 6.01

$\text{CosineDistance (B,C)} = 1 -$
$$\frac{(7.8256) + 204800(\frac{1}{486.6666})^2 + 24(\frac{1}{5.333})^2}{(\sqrt{7.1824 + 102400(\frac{1}{486.6666})^2 + 16(\frac{1}{5.333})^2})*(\sqrt{8.5264 + 409600(\frac{1}{486.6666})^2 + 36(\frac{1}{5.333})^2})}$$

$= 1 - 0.9824$

Angle between B & C = 10.76

CosineDistance (A,C) = .93244

Angle between A & C = 21.18

9.2.3

a)

Avg Ratings = (4 + 2 + 5)/3 = 11/3

Normalized Rating:

A: 4-11/3 = 1/3

B: 2-11/3 = -5/3

C: 5-11/3 = 4/3

b)

Value for processor speed = $((3.06-\frac{1}{3}) + (2.68-\frac{-5}{3}) + (2.92-\frac{4}{3}))/3 = 2.8866$

Value for disk size = $((500-\frac{1}{3}) + (320-\frac{-5}{3}) + (640-\frac{4}{3}))/3 = 486.667$

Value for main memory size = $((6-\frac{1}{3}) + (4-\frac{-5}{3}) + (6-\frac{4}{3}))/3 = 5.3333$

The user profile for the user, with components for processor speed, disk size, and main memory size, is [2.8866, 486.667, 5.3333].

9.3.1

a)

A: {1,1,0,1,1,0,1,1}, B: {0,1,1,1,1,1,1,0}, C: {1,0,0,1,1,0,1,1,1}

Jaccard's Coefficient (A,B) = $\frac{4}{4+2+2} = \frac{1}{2}$

Distance = $1 - \frac{1}{2} = \frac{1}{2}$

Jaccard's Coefficient (B,C) = $\frac{4}{4+2+2} = \frac{1}{2}$

Distance = $1 - \frac{1}{2} = \frac{1}{2}$

Jaccard's Coefficient (A,C) = $\frac{4}{4+2+2} = \frac{1}{2}$

Distance = $1 - \frac{1}{2} = \frac{1}{2}$

b)

Cosine Distance (A,B) = $1 - \frac{(1*0) + (1*1) + (0*1) + (1*1) + (1*1) + (0*1) + (1*1) + (1*0)}{(\sqrt{1,\ 1,\ 0,\ 1,\ 1,\ 0,\ 1,\ 1}) * (\sqrt{0,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 0})} = 1 - 0.6667$

Cosine Distance (B,C) = 1 - 0.6667

Cosine Distance (A,C) = 1 - 0.6667

c)

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}

Jaccard's Coefficient (A,B) = $\frac{2}{2+2+1} = \frac{2}{5}$

Distance = 1 - $\frac{2}{5} = \frac{3}{5}$

Jaccard's Coefficient (B,C) = $\frac{1}{1+1+3} = \frac{1}{5}$

Distance = 1 - $\frac{1}{5} = \frac{4}{5}$

Jaccard's Coefficient (A,C) = $\frac{2}{2+1+2} = \frac{2}{5}$

Distance = 1 - $\frac{2}{5} = \frac{3}{5}$

d)

Cosine Distance (A,B) = 1 - $\frac{(1*0)+(1*1)+(0*1)+(1*1)+(0*0)+(0*0)+(1*0)+(0*0)}{(\sqrt{1,\ 1,\ 0,\ 1,\ 0,\ 0,\ 1,\ 0})*(\sqrt{0,\ 1,\ 1,\ 1,\ 0,\ 0,\ 0,\ 0})}$ = 1 - 0.577 = 0.42

Cosine Distance (B,C) = 1 - 0.288 = 0.71

Cosine Distance (A,C) = 1 - 0.5 = 0.5

e)

Avg for A: 4+5+5+1+3+2 = 20/6 = 3.33

Avg for B: 14/6 = 2.33

Avg for C: 18/6 = 3
A: {0.67, 1.67, 1.67, -2.33, -0.33, -1.33}, B: {0.67, 1.67, 0.67, -1.33, -0.33, -1.33}, C: {-1, -2, 0, 1, 2, 0}

Cosine Distance (A,B) = 1 - $\frac{0.67,\ 1.67,\ 1.67,\ -2.33,\ -0.33,\ -1.33\ dot\ product\ 0.67,\ 1.67,\ 1.67,\ -2.33,\ -0.33,\ -1.33}{(\sqrt{0.67,\ 1.67,\ 1.67,\ -2.33,\ 3,\ -0.33,\ -1.33})*(\sqrt{0.67,\ 1.67,\ 0.67,\ -1.33,\ -0.33,\ -1.33})}$ = 1 - 0.9438 = 0.0562

Cosine Distance (B,C) = 1 - -0.7 = 1.7

Cosine Distance (A,C) = 1 - -0.606 = 1.606

9.4.1
a)

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

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

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

Minimum value of this expression

2 * ((x-1) + (x-2) + x + (x-3)) = 0

X = 1.5
Therefore,

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

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

b)

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

Start with the U and V

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

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

2 ((y-3) + (y-3) + y + (y-2) + (y-3)) = 0

y = 2.2

Therefore,

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

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