CS 422

Data Mining

Homework Assignment 4

Ronit Rudra
A20379221
rrudra@hawk.iit.edu

Course Instructor
Prof. Jawahar Panchal

1 Textbook Questions

1.1 Question 9.2.1

We are given:

$$A = [3.06\ , 500\alpha\ , 6\beta]$$
$$B = [2.68\ , 320\alpha\ , 4\beta]$$
$$C = [2.92\ , 640\alpha\ , 6\beta]$$

The cosine distance between two vectors is:

$$Cosine(XY) = \frac{X \cdot Y}{|X| \cdot |Y|} \tag{1}$$

0.696874047 (a) Plugging in the values we get:

$$\begin{aligned}
Coisne(AB) &= \frac{(3.06)(2.68) + (500)(320\alpha^2) + (6)(4\beta^2)}{\sqrt{(3.06)^2 + (500\alpha)^2 + (6\beta)^2} \cdot \sqrt{(2.68)^2 + (320\alpha)^2 + (4\beta)^2}} \\
&= \frac{(8.2008) + (1.6 \cdot 10^5\alpha^2) + (24\beta^2)}{\sqrt{2.56 \cdot 10^{10}\alpha^4 + 7.68 \cdot 10^6\alpha^2\beta^2 + 2.75 \cdot 10^6\alpha^2 + 576\beta^4 + 408.384\beta^2 + 67.2531}} \\
Coisne(AC) &= \frac{(3.06)(2.92) + (500)(640\alpha^2) + (6)(6\beta^2)}{\sqrt{(3.06)^2 + (500\alpha)^2 + (6\beta)^2} \cdot \sqrt{(2.92)^2 + (640\alpha)^2 + (6\beta)^2}} \\
&= \frac{(8.9352) + (32 \cdot 10^4\alpha^2) + (26\beta^2)}{\sqrt{1.024 \cdot 10^{11}\alpha^4 + 2.375 \cdot 10^6\alpha^2\beta^2 + 5.96 \cdot 10^6\alpha^2 + 1296\beta^4 + 644.04\beta^2 + 79.8378}} \\
Coisne(BC) &= \frac{(2.68)(2.92) + (320)(640\alpha^2) + (6)(6\beta^2)}{\sqrt{(2.68)^2 + (320\alpha)^2 + (4\beta)^2} \cdot \sqrt{(2.92)^2 + (640\alpha)^2 + (6\beta)^2}} \\
&= \frac{(7.8256) + (204800\alpha^2) + (24\beta^2)}{\sqrt{4.19 \cdot 10^{10}\alpha^4 + 9.4 \cdot 10^6\alpha^2\beta^2 + 3.81 \cdot 10^6\alpha^2 + 504\beta^4 + 377.936\beta^2 + 61.24}}
\end{aligned}$$

(b) $\alpha = \beta = 1$

$$\begin{aligned}
Coisne(AB) &= \frac{(3.06)(2.68) + (500)(320) + (6)(4)}{\sqrt{(3.06)^2 + (500)^2 + (6)^2} \cdot \sqrt{(2.68)^2 + (320)^2 + (4)^2}} \\
&= \frac{(8.2008) + (1.6 \cdot 10^5) + (24)}{\sqrt{2.56 \cdot 10^{10} + 7.68 \cdot 10^6 + 2.75 \cdot 10^6 + 576 + 408.384 + 67.2531}} \\
&= 0.999 = 2.6^\circ \\
Coisne(AC) &= \frac{(3.06)(2.92) + (500)(640) + (6)(6)}{\sqrt{(3.06)^2 + (500)^2 + (6)^2} \cdot \sqrt{(2.92)^2 + (640)^2 + (6)^2}} \\
&= \frac{(8.9352) + (32 \cdot 10^4) + (26)}{\sqrt{1.024 \cdot 10^{11} + 2.375 \cdot 10^6 + 5.96 \cdot 10^6 + 1296 + 644.04 + 79.8378}} \\
&= 0.999 = 2.6^\circ \\
Coisne(BC) &= \frac{(2.68)(2.92) + (320)(640) + (6)(6)}{\sqrt{(2.68)^2 + (320)^2 + (4)^2} \cdot \sqrt{(2.92)^2 + (640)^2 + (6)^2}}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{(7.8256) + (204800) + (24)}{\sqrt{4.19 \cdot 10^{10} + 9.4 \cdot 10^{6} + 3.81 \cdot 10^{6} + 504 + 377.936 + 61.24}} \\
&= 0.99 = 8.1^{\circ}
\end{aligned}
$$

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

$$
\begin{aligned}
Coisne(AB) &= \frac{(8.2008) + (16) + (6)}{\sqrt{256 + 192 + 275 + 36 + 102.87 + 67.2531}} \\
&= 0.9881 = 3.5^{\circ} \\
Coisne(AC) &= \frac{(8.9352) + (32) + (9)}{\sqrt{1024 + 59.29 + 596 + 81 + 161.01 + 79.8378}} \\
&= 0.9915 = 7.5^{\circ} \\
Coisne(BC) &= \frac{(7.8256) + (20.48) + (6)}{\sqrt{4194.304 + 235.52 + 381 + 31.5 + 94.484 + 61.24}} \\
&= 0.9691 = 14.3^{\circ}
\end{aligned}
$$

(d) Assuming constant of proportionality k_1 and k_2:

$$
\alpha \propto \frac{1}{\frac{500+640+320}{3}} = \frac{3}{1460}.k_1 \tag{2}
$$

$$
\beta \propto \frac{1}{\frac{6+4+6}{3}} = \frac{3}{16}.k_2 \tag{3}
$$

For ease of calculations let us assume k_1 and k_2 as 1. Therefore, the vectors are:

$$
\begin{aligned}
A &= [3.06\ , 1.027\ , 1.125] \\
B &= [2.68\ , 0.657\ , 0.75] \\
C &= [2.92\ , 1.315\ , 1.125]
\end{aligned}
$$

The Cosines between the vectors are:

$$
\begin{aligned}
Cosine(A, B) &= 0.9914 = 7.5^{\circ} \\
Cosine(A, C) &= 0.9957 = 5.3^{\circ} \\
Cosine(B, C) &= 0.9793 = 11.7^{\circ}
\end{aligned}
$$

$$
\begin{aligned}
\alpha &= \frac{3}{1460} \\
\beta &= \frac{3}{16}
\end{aligned}
$$

1.2 Question 9.2.3

(a) The average rating is $\frac{4+2+5}{3} = \frac{11}{3} = 3.67$.

Normalized ratings for A,B,C are:

$$A = 4 - 3.67 = 0.33$$
$$B = 2 - 3.67 = -1.67$$
$$C = 5 - 3.67 = 1.33$$

(b) Creating User Profiles based on ratings:

$$Processor\ Speed = (3.06)(0.33) + (2.68)(-1.67) + (2.92)(1.33) = 0.4467$$
$$Disk\ Size = (500)(0.33) + (320)(-1.67) + (640)(1.33) = 486.667$$
$$Memory\ Size = (6)(0.33) + (4)(-1.67) + (6)(1.33) = 3.33$$

1.3 Question 9.3.1

The Jaccard Distance and Cosine Distance are:

$$J = \frac{f_{11}}{f_{10} + f_{01} + f_{11}} \tag{4}$$
$$C(XY) = 1 - \frac{X \cdot Y}{|X| \cdot |Y|} \tag{5}$$
$$\tag{6}$$

(a) Treating Utility Matrix as boolean we get:
A = [11011011]
B = [01111110]
C = [10110111]

$$J(AB) = \frac{4}{8} = 0.5$$
$$J(BC) = \frac{4}{8} = 0.5$$
$$J(AC) = \frac{4}{8} = 0.5$$

(b)

$$C(AB) = 1 - \frac{4}{6} = 1 - 0.67 = 0.33$$
$$C(BC) = 1 - \frac{4}{6} = 1 - 0.67 = 0.33$$
$$C(AC) = 1 - \frac{4}{6} = 1 - 0.67 = 0.33$$

(c) The Vectors are:
A = [11010010]
B = [01110000]
C = [00010111]

$$J(AB) = \frac{2}{5} = 0.4$$
$$J(BC) = \frac{1}{6} = 0.167$$
$$J(AC) = \frac{2}{6} = 0.33$$

(d)

$$C(AB) = 1 - \frac{2}{\sqrt{4} \cdot \sqrt{3}} = 1 - 0.577 = 0.423$$
$$C(BC) = 1 - \frac{1}{\sqrt{3} \cdot \sqrt{4}} = 1 - 0.288 = 0.712$$
$$C(AC) = 1 - \frac{2}{\sqrt{4} \cdot \sqrt{4}} = 1 - 0.5 = 0.5$$

(e) Average ratings are:

$$A = \frac{4+5+5+1+3+2}{6} = \frac{10}{3}$$
$$B = \frac{3+4+3+1+2+1}{6} = \frac{7}{3}$$
$$C = \frac{2+1+3+4+5+3}{6} = \frac{17}{6}$$

The new vectors are:
A = [1.33, 1.67, 0, 1.67, -2.33, 0, -0.33, -1.33]
B = [0, 0.67, 1.67, 0.67, -1.33, -0.33, -1.33, 0]
C = [-0.833, 0, -1.83, 1.67, 0, 1.167, 2.167, 1.167]

(f) Using the above vectors to calculate the Cosine Distance:

$$C(AB) = 0.4428$$
$$C(BC) = 1.509$$
$$C(AC) = 1.04$$

1.4 Question 9.4.1

We are given:

$$U = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$V = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

(a) Optimize with u_{32}

$$U = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & x \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

The UV Matrix is:

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

$$Sum\ of\ Squares = (2-1-x)^2 + (3-1-x)^2 + (1-1-x)^2 + (4-1-x)^2 = (1-x)^2 + (2-x)^2 + (-x)^2 + (3-x)^2$$

$$To\ Minimize\ we; find\ \frac{\partial f(x)}{\partial x} = 0$$

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

$$= x = 1.5$$

Therefore U is:

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1.5 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

UV is:

$$\begin{bmatrix} 2 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 \\ 2.5 & 2.5 & 2.5 & 2.5 & 2.5 \\ 2 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 \end{bmatrix}$$

(b) Optimize with v_{41}.

$$V = \begin{bmatrix} 1 & 1 & 1 & x & 1 \\ 1 & 1 & 1 & x & 1 \end{bmatrix}$$

The UV Matrix is:

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

$$Sum\ of\ Squares = (4-1-x)^2 + (4-1-x)^2 + (1-1-x)^2 + (3-1-x)^2 + (4-1-x)^2$$
$$= 3 \cdot (3-x)^2 + (-x)^2 + (2-x)^2$$
$$To\ Minimize\ we; find\ \frac{\partial f(x)}{\partial x} = 0$$
$$= -6(3-x) + 2x - 2(2-x) = 0$$
$$= x = 2.2$$

Therefore V is:

$$\begin{bmatrix} 1 & 1 & 1 & 2.2 & 1 \\ 1 & 1 & 1 & 2.2 & 1 \end{bmatrix}$$

The UV Matrix is:

$$\begin{bmatrix} 2 & 2 & 2 & 3.2 & 2 \\ 2 & 2 & 2 & 3.2 & 2 \\ 2 & 2 & 2 & 3.2 & 2 \\ 2 & 2 & 2 & 3.2 & 2 \\ 2 & 2 & 2 & 3.2 & 2 \end{bmatrix}$$