AI506 Homework 6

20194331 김성년

1. Association Rules

1-1.

100 이하이면서 i의 배수가 5개 이상인 i: 1~20

1-2.

support{5,7} = 2, support{5,7,2} = 1. ∴ confidence = 0.5

1-3.

support{2,3,4} = 8, support{2,3,4,5} = 1. ∴ confidence = 1/8

2. PCY Algorithm

2-1.

bucket	0	1	2	3	4	5	6	7	8	9	10	total
pairs		{3,4} {2,6}	{1,2} {4,6}	{1,3}	{1,4} {3,5}	{1,5}	{2,3}	{3,6}	{2,4} {5,6}	{4,5}	{2,5}	
count	0	5	5	3	6	1	3	2	6	3	2	36

2-2.

From the table above, bucket 1, 2, 4, 8 are frequent. (count>=4)

2-3.

{3,4}, {2,6}, {1,2}, {4,6}, {1,4}, {3,5}, {2,4}, {5,6}

3. Content-based Recommendation

3-1.

A: $[3.06,\ 500\alpha, 6\beta]$

B: $[2.68,\ 320\alpha,\ 4\beta]$

C: $[2.92,\ 640\alpha, 6\beta]$

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

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

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

3-2.

$\cos(A, B) = 0.999997,\ \theta(A, B) = 0.0024\ \text{rad} = 0.14°$

$\cos(B, C) = 0.999988,\ \theta(B, C) = 0.0049\ \text{rad} = 0.28°$

$\cos(C, A) = 0.999995,\ \theta(C, A) = 0.0032\ \text{rad} = 0.18°$

3-3.

$\cos(A, B) = 0.990882,\ \theta(A, B) = 0.1351\ \text{rad} = 7.74°$

$\cos(B, C) = 0.969178,\ \theta(B, C) = 0.2489\ \text{rad} = 14.26°$

$\cos(C, A) = 0.991555,\ \theta(C, A) = 0.1301\ \text{rad} = 7.45°$

3-4.

Feature	Average value	Scale
Processor Speed	2.89	1
Disk Size	486.67	$\frac{2.89}{486.67} = 0.006$
Main-Memory Size	5.33	$\frac{2.89}{5.33} = 0.54$

$\therefore \alpha = 0.006,\ \beta = 0.54$

3-5.

A: [3.06, 3, 3.24]

B: [2.68, 1.92, 2.16]

C: [2.92, 3.84, 3.24]

From the same calculation as 3-1~3,

$\theta(\mathrm{A}, \mathrm{B}) = 8.17°$

$\theta(\mathrm{B}, \mathrm{C}) = 14.27°$

$\theta(\mathrm{C}, \mathrm{A}) = 7.49°$

4. UV Decomposition

4-1.

Let u be the value in U and V, and k is for the entries of a given matrix without missing values.

Minimizing RMSE is equivalent to

$$\begin{aligned} &\min_u \sum_k (2u^2 - k)^2 \\ = &\min_u \sum_k 4u^4 - 4u^2 k + k^2 \end{aligned}$$

Then, the loss function becomes

$$\ell = \sum_k 4u^4 - 4u^2 k + k^2$$

$$\frac{\partial \ell}{\partial u} = 23 \times 16u^3 - 8u \sum k = 0$$

$\sum k$ is 75. Therefore, from the above equation, $u = 1.277$.

4-2.

$$\begin{aligned} RMSE &= \sqrt{\frac{1}{23} \sum_k (2u^2 - k)^2} \\ &= \sqrt{\frac{1}{23} \sum_k (3.26^2 - 6.52k + k^2)} \\ &= 1.2926 \end{aligned}$$

4-3.

We only consider the first row because others do not depend on u_{11}.

$$\min_{u_{11}} \sum_{k \in \text{first row}} (1.277u_{11} + 1.277^2 - k)^2$$

By the same method as 4-1,

$$\begin{aligned}\frac{\partial \ell}{\partial u_{11}} &= \sum_k 2.554(1.277u_{11} + 1.277^2 - k) \\ &= 5 \times 2.554 \times 1.277 \times u_{11} + 20.82 - 2.554 \times 18 = 0\end{aligned}$$

Therefore, $u_{11} = 1.542$.

5. Singular Value Decomposition

5-1.

$$M^T M = \begin{bmatrix} 1 & 3 & 5 & 0 & 1 \\ 2 & 4 & 4 & 2 & 3 \\ 3 & 5 & 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 5 & 4 & 3 \\ 0 & 2 & 4 \\ 1 & 3 & 5 \end{bmatrix} = \begin{bmatrix} 36 & 37 & 38 \\ 37 & 49 & 61 \\ 38 & 61 & 84 \end{bmatrix}$$

$$M M^T = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 5 & 4 & 3 \\ 0 & 2 & 4 \\ 1 & 3 & 5 \end{bmatrix} \begin{bmatrix} 1 & 3 & 5 & 0 & 1 \\ 2 & 4 & 4 & 2 & 3 \\ 3 & 5 & 3 & 4 & 5 \end{bmatrix} = \begin{bmatrix} 14 & 26 & 22 & 16 & 22 \\ 26 & 50 & 46 & 28 & 40 \\ 22 & 46 & 50 & 20 & 32 \\ 16 & 28 & 20 & 20 & 26 \\ 22 & 40 & 32 & 26 & 35 \end{bmatrix}$$

5-2.

I used computer for numpy.linalg.eig(matrix).

Eigenvalues for $M^T M$ are

$\lambda_1 = 153.57, \lambda_2 = 15.43, \lambda_3 = 0$

Eigenvalues for MM^T are

$\lambda_1 = 153.57, \lambda_2 = 15.43, \lambda_3 = 0, \lambda_4 = 0, \lambda_5 = 0$

5-3.

Corresponding eigenvectors for $M^T M$:

$$v_1 = \begin{bmatrix} 0.4093 \\ 0.5635 \\ 0.7176 \end{bmatrix}, v_2 = \begin{bmatrix} -0.8160 \\ -0.1259 \\ 0.5642 \end{bmatrix}, v_3 = \begin{bmatrix} 0.4082 \\ -0.8165 \\ 0.4082 \end{bmatrix}$$

Eigenvectors for MM^T:

$$u_1 = \begin{bmatrix} 0.2977 \\ 0.5705 \\ 0.5207 \\ 0.3226 \\ 0.4590 \end{bmatrix}, u_2 = \begin{bmatrix} 0.1591 \\ -0.0332 \\ -0.7359 \\ 0.5104 \\ 0.4143 \end{bmatrix}, u_3 = \begin{bmatrix} 0.9413 \\ -0.1748 \\ -0.0403 \\ -0.1883 \\ -0.2152 \end{bmatrix}, u_4 = \begin{bmatrix} 0.7115 \\ -0.1761 \\ 0.0637 \\ 0.2658 \\ -0.5016 \end{bmatrix}, u_5 = \begin{bmatrix} 0.7115 \\ -0.1761 \\ 0.0637 \\ 0.2658 \\ -0.5016 \end{bmatrix}$$

5-4.

singular values are $\sqrt{153.57} = 12.39$ and $\sqrt{15.43} = 3.93$.

$$\begin{aligned} \mathrm{M} &= [u_1 \quad u_2] \begin{bmatrix} 12.39 & 0 \\ 0 & 3.93 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix} \\ &= \begin{bmatrix} 0.2977 & 0.1591 \\ 0.5705 & -0.0332 \\ 0.5207 & -0.7359 \\ 0.3226 & 0.5104 \\ 0.4590 & 0.4143 \end{bmatrix} \begin{bmatrix} 12.39 & 0 \\ 0 & 3.93 \end{bmatrix} \begin{bmatrix} 0.4093 & 0.5635 & 0.7176 \\ -0.8160 & -0.1259 & 0.5642 \end{bmatrix} \end{aligned}$$

5-5.

$$\begin{aligned} \mathrm{M} &\approx [u_1][12.39][v_1^T] \\ &= \begin{bmatrix} 0.2977 \\ 0.5705 \\ 0.5207 \\ 0.3226 \\ 0.4590 \end{bmatrix} [12.39][0.4093 \quad 0.5635 \quad 0.7176] \\ &= \begin{bmatrix} 1.5097 & 2.0785 & 2.6469 \\ 2.8931 & 3.9831 & 5.0724 \\ 2.6406 & 3.6354 & 4.6296 \\ 1.6360 & 2.2523 & 2.8683 \\ 2.3277 & 3.2046 & 4.0810 \end{bmatrix} \end{aligned}$$

5-6.

$$\frac{153.57}{153.57+15.43} \times 100 = 90.87\%$$