1. Section 1

A transaction database is given below. Using the A-priori algorithm, determine all frequent item-sets with minimum support of 30%. Show results at each step of the algorithm.

Minimum support of 30% = 3

$TID\#$	$ItemsBought$
1	A, B, D, E
2	B, C, D
3	A, B, D, E
4	A, C, D, E
5	B, C, D, E
6	B, D, E
7	C, D
8	A, B, C
9	A, D, E
10	B, D

$Item$	$Support$
A	5
B	7
C	5
D	9
E	6

$Item$	$Support$	
A, B	3	
A, C	2	$< -Drop$
A, D	4	
A, E	4	
B, C	3	
B, D	6	
B, E	4	
C, D	4	
C, E	2	$< -Drop$
D, E	6	

Droping {A, C}, {C, E} Because they are bellow the minimum of 3

$Item$	$Support$	
A, B, D	2	$< -Drop$
A, B, E	2	$< -Drop$
A, D, E	4	
B, C, D	2	$< -Drop$
B, D, E	4	

Droping {A, B, D}, {A, B, E}, {B, C, D} Because they are bellow the minimum of 3 There are no other permutaions left to process

2. Section 2

Consider the following simple IR situation. We have two keywords K1 and K2, and three documents D1, D2 and D3. The weights of keywords are given by the following term-document matrix F.

$$\begin{array}{cc} & \begin{array}{ccc} D1 & D2 & D3 \end{array} \\ \begin{array}{c} K1 \\ K2 \end{array} & \begin{bmatrix} 0.25 & 0.53 & 0.75 \\ 0.73 & 0.50 & 0.23 \end{bmatrix} \end{array}$$

2.1 Obtain the singular value decomposition of F.

$$d <- \begin{bmatrix} 1.2292184 & 0 \\ 0 & 0.5102176 \end{bmatrix}$$

$$u <- \begin{bmatrix} -0.7184647 & -0.6955634 \\ -0.6955634 & 0.7184647 \end{bmatrix}$$

$$v^t <- \begin{bmatrix} -0.570516 & -0.59270833 & -0.5685142 \\ 0.715298 & -0.01845542 & -0.6985758 \end{bmatrix}$$

$$F <- \begin{bmatrix} -0.7184647 & -0.6955634 \\ -0.6955634 & 0.7184647 \end{bmatrix} * \begin{bmatrix} 1.2292184 & 0 \\ 0 & 0.5102176 \end{bmatrix} * \begin{bmatrix} -0.570516 & -0.59270833 & -0.5685142 \\ 0.715298 & -0.01845542 & -0.6985758 \end{bmatrix}$$

2.2 Reconstruct F ignoring the smaller of the two singular values.

$$d <- \begin{bmatrix} 1.2292184 & 0 \\ 0 & 0 \end{bmatrix}$$

$$u <- \begin{bmatrix} -0.7184647 & -0.6955634 \\ -0.6955634 & 0.7184647 \end{bmatrix}$$

$$v^t <- \begin{bmatrix} -0.570516 & -0.59270833 & -0.5685142 \\ 0.715298 & -0.01845542 & -0.6985758 \end{bmatrix}$$

$$F <- \begin{bmatrix} -0.7184647 & -0.6955634 \\ -0.6955634 & 0.7184647 \end{bmatrix} * \begin{bmatrix} 1.2292184 & 0 \\ 0 & 0 \end{bmatrix} * \begin{bmatrix} -0.570516 & -0.59270833 & -0.5685142 \\ 0.715298 & -0.01845542 & -0.6985758 \end{bmatrix} = \begin{bmatrix} 0.5038512 & 0.5234504 \\ 0.4877908 & 0.5067653 \end{bmatrix}$$

$$\begin{array}{cc} & \begin{array}{ccc} D1 & D2 & D3 \end{array} \\ \begin{array}{c} K1 \\ K2 \end{array} & \begin{bmatrix} 0.5038512 & 0.5234504 & 0.5020833 \\ 0.4877908 & 0.5067653 & 0.4860793 \end{bmatrix} \end{array}$$

2.3 Assume a query with keyword K1. Show its representation in the LSI space. Using the cosine similarity measure in the LSI space, rank the three documents D1, D2 and D3.

$$\begin{array}{cc} & \begin{array}{ccc} D1 & D2 & D3 \end{array} \\ \begin{array}{c} K1 \\ K2 \end{array} & \begin{bmatrix} -0.7012887 & -0.728568 & -0.6988281 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

3. Section 3

Select ten words from different fields, e.g. politics, sports, geography, finance etc. Compute Google distance between pairs of these words and apply the single-linkage clustering algorithm on the resulting distance matrix. Comment on the nature of results.

4. Section 4

This exercise is based on Section 9.4 of Recommender System chapter. In this section, the issue of sparse entries in the user-item matrix is discussed with the SVD approach. Carefully read the chapter and do Exercise 9.4.1 given on page 316. Provide details of calculations of each step of your work.

Exercise 9.4.1 : Starting with the decomposition of Fig. 9.10, we may choose any of the 20 entries in U or V to optimize first. Perform this first optimization step assuming we choose: (a) u32 (b) v41.

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

4.1 u_{32}

First optimizaiton step Make u_{32} a variable

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

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

$$(2-(x+1))^2 + (3-(x+1))^2 + (1-(x+1))^2 + (4-(x+1))^2 \tag{4.1}$$

$$(1-x)^2 + (2-x)^2 + (0-x)^2 + (3-x)^2 \tag{4.2}$$

$$-2 * ((1-x) + (2-x) + (0-x) + (3-x)) = 0 \tag{4.3}$$

$$-2 * (6-4x) = 0 \tag{4.4}$$

$$x = 1.5 \tag{4.5}$$

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

4.2 v_{41}

First optimizaiton step Make v_{41} a variable

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

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

$$(4-(y+1))^2 + (4-(y+1))^2 + (1-(y+1))^2 + (3-(y+1))^2 + (4-(y+1))^2 \tag{4.6}$$

$$(3-y)^2 + (3-y)^2 + (0-y)^2 + (2-y)^2 + (3-y)^2 \tag{4.7}$$

$$-2 * ((3-y) + (3-y) + (0-y) + (2-y) + (3-y)) = 0 \tag{4.8}$$

$$-2 * (11 - 5y) = 0 \tag{4.9}$$

$$y = 2.2 \tag{4.10}$$

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

5. Section 5

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Remove Dead Ends 1, 3 and 6

$$\begin{bmatrix} 0 & 1/2 & 0 \\ 0 & 0 & 1 \\ 1 & 1/2 & 0 \end{bmatrix}$$

Start with equal vector equal to 1/3

$$\begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} \begin{bmatrix} 1/6 \\ 2/6 \\ 3/6 \end{bmatrix} \begin{bmatrix} 1/6 \\ 3/6 \\ 2/6 \end{bmatrix} \begin{bmatrix} 0.2500000 \\ 0.3333333 \\ 0.4166667 \end{bmatrix} \dots \begin{bmatrix} 0.2 \\ 0.4 \\ 0.4 \end{bmatrix}$$