Seoul National University

M1522.001400-001 Introduction to Data Mining

Homework 5: Advertising on the Web & Recommendation (Chapter 8 & 9)

2017-18538 Hwang SunYoung

Question1

1) Let S={A, B, C, H} and T={D, E, F, G}, then Cut(S, T)=2. Vol(S)=6 and Vol(T)=7. Thus, the normalized cut value of the cut {{A, B, C, H}, {D, E, F, G}} is $\frac{2}{6}+\frac{2}{7}=\frac{13}{21}$.

2) Let S={A, B, C} and T={D, E, F, G, H}, then Cut(S, T)=2. Vol(S)=5 and Vol(T)=8. Thus, the normalized cut value of the cut {{A, B, C}, {D, E, F, G, H}} is $\frac{2}{5}+\frac{2}{8}=\frac{13}{20}$.

3) To find a "good" partition of a graph, we need to minimize a given graph cut value. Because $\frac{13}{21}<\frac{13}{20}$, the node set {{A, B, C, H}, {D, E, F, G}} produces a more balanced partition then the node set {{A, B, C}, {D, E, F, G, H}}.

Question2

Itemsets: A={B, C}, B={A, C, D}, C={A, B}, D={B, E, F, G}, E={D, F}, F={D, E, G}, G={D, F}

1) s=1, t=3

We must find itemsets of size 3 that appear in at least 1 baskets. {A, C, D}, {B, E, F}, {B, E, G}, {B, F, G}, {E, F, G}, {D, E, G} are itemsets that support is 1 and size is 3. So answer is 6.

2) s=2, t=2

We must find itemsets of size 2 that appear in at least 2 baskets. {D, F}, {E, G} are itemsets that support is 2 and size is 2. So answer is 2.

Question3

$$A = U\Sigma V^T$$

$$AA^T = U\Sigma V^T V\Sigma^T \mathrm{U}^T = U\Sigma^2 \mathrm{U}^T$$

$$A^T A = V\Sigma^T \mathrm{U}^T U\Sigma V^T = V\Sigma^2 \mathrm{V}^T$$

Therefore, we can solve the problem by eigen decomposing AA^T and $A^T A$.

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -3 & -3 & -3 \end{bmatrix}$$

First we compute the singular values σ_i by finding the eigenvalues of AA^T.

$$AA^T = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 3 & 0 & -9 \\ 1 & 0 & 1 & 0 \\ 0 & -9 & 0 & 27 \end{bmatrix}$$

det($AA^T - \lambda\mathrm{I}$)=det($\begin{bmatrix} 1-\lambda & 0 & 1 & 0 \\ 0 & 3-\lambda & 0 & -9 \\ 1 & 0 & 1-\lambda & 0 \\ 0 & -9 & 0 & 27-\lambda \end{bmatrix}$) = (1- λ)(3- λ)(1- λ)(27- λ)-81(1- λ)(1- λ)+81-(27- λ)(3- λ)

λ is 30, 2, 0.

so the rank is 2 and the singular values are σ_1=$\sqrt{30}$, σ_2=$\sqrt{2}$, Σ=$\begin{bmatrix} \sqrt{30} & 0 \\ 0 & \sqrt{2} \end{bmatrix}$

$$AA^T - 30\mathrm{I} = \begin{bmatrix} -29 & 0 & 1 & 0 \\ 0 & -20 & 0 & -9 \\ 1 & 0 & -29 & 0 \\ 0 & -9 & 0 & -3 \end{bmatrix}$$

The unit eigenvector of AA^T when eigenvalue 30 $u_1 = \begin{bmatrix} 0 \\ \frac{1}{\sqrt{10}} \\ 0 \\ \frac{-3}{\sqrt{10}} \end{bmatrix}$

$$AA^T - 2\mathrm{I} = \begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -9 \\ 1 & 0 & -1 & 0 \\ 0 & -9 & 0 & 25 \end{bmatrix}$$

The unit eigenvector of AA^T when eigenvalue 2 $u_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}$

So U=$\begin{bmatrix} 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{10}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \\ \frac{-3}{\sqrt{10}} & 0 \end{bmatrix}$

$$A^TA = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 10 & 10 & 10 \\ 0 & 10 & 10 & 10 \\ 0 & 10 & 10 & 10 \end{bmatrix}$$

$$A^TA - 30\mathrm{I} = \begin{bmatrix} -28 & 0 & 0 & 0 \\ 0 & -20 & 10 & 10 \\ 0 & 10 & -20 & 10 \\ 0 & 10 & 10 & -20 \end{bmatrix}$$

The unit eigenvector of $A^T A$ when eigenvalue 30 $v_1 = \begin{bmatrix} 0 \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$

$$A^T A - 2\mathrm{I} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 8 & 10 & 10 \\ 0 & 10 & 8 & 10 \\ 0 & 10 & 10 & 8 \end{bmatrix}$$

The unit eigenvector of $A^T A$ when eigenvalue 2 $v_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

So V= $\begin{bmatrix} 0 & \\ \frac{1}{\sqrt{3}} & 1 \\ \frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{3}} & 0 \\ & 0 \end{bmatrix}$

A= $\begin{bmatrix} 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{10}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \\ \frac{-3}{\sqrt{10}} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{30} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}$

Question4

1) (the number of elements)×(4 bytes) = $10^8 \times 10^6 \times$ (4 bytes) = 4× 10^{14}bytes

We need 400TB to store X naively.

2) rank is 10.

Let M be an m×n matrix, and let the rank of M be r.

U is an m×r column-orthonormal matrix.

V is an n×r column-orthonormal matrix.

Λ is a diagonal matrix. Size is r×r. the number of nonzero values is r. (main diagonal). All elements except main diagonal are zero.

In this question, m=10^8, n=10^6, r=10 (rank is 10)

U has $10^8 \times 10$ nonzero elements.

V has $10^6 \times 10$ nonzero elements.

Λ has 10 nonzero elements.

(the number of elements)×(4 bytes) = $(10^8 \times 10 + 10^6 \times 10 + 10)$×(4 bytes) = 4040000040bytes

We need 4040000040bytes to store U, V, Λ.