Homework 5

Kevin Serrano

April 28, 2014

Part 1

If we multiply R by R^t, we obtain a similarity matrix. Row i shows the number of items that are liked by user i and j, where the user j is the j-th column.
Example:
Let $m = 3$ and $n = 5$. Define

$$R = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Then

$$R \cdot R^t = \begin{bmatrix} 3 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 1 & 3 \end{bmatrix}$$

That means user 1 likes 3 items in common with himself (obviously), 1 with user 2 and 2 with user 3.

By the definition of the cosine similarity, we need to divide each (i, j) by the norm of the number of item liked by the user i and j (the norm is the square root here). To applied this to the all matrix, we just need to divide $R \cdot R^t$ by the sqrt of P : $\sqrt{P}$.
Thus

$$S_U = \sqrt{P}^{-1} \cdot R \cdot R^t \cdot \sqrt{P}^{-1}$$

For the example we obtain

$$\frac{R \cdot R^t}{\sqrt{P} \cdot \sqrt{P}} = \sqrt{P}^{-1} \cdot R \cdot R^t \cdot \sqrt{P}^{-1} = \begin{bmatrix} 1.0000 & 0.4082 & 0.6667 \\ 0.4082 & 1.0000 & 0.4082 \\ 0.6667 & 0.4082 & 1.0000 \end{bmatrix}$$

where for example $0.4082 = \frac{1}{\sqrt{3} \cdot \sqrt{2}}$

By the same way, we can compute the cosine similarity for items as follow :
$R^t \cdot R$ gives us the number of user who like item i and j. With the example it is

$$R^t \cdot R = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 2 & 1 & 2 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 2 & 1 & 3 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

That means for example that 2 users like item 2 and 4. Now we can just divide by the sqrt of Q to get the cosine similarity between items

$$S_I = \sqrt{Q}^{-1} \cdot R^t \cdot R \cdot \sqrt{Q}^{-1}$$

For the example we get

$$\sqrt{Q}^{-1} \cdot R^t \cdot R \cdot \sqrt{Q}^{-1} = \begin{bmatrix} 1.0000 & 0 & 0 & 0.5774 & 0 \\ 0 & 1.0000 & 0.7071 & 0.8165 & 0.7071 \\ 0 & 0.7071 & 1.0000 & 0.5774 & 0 \\ 0.5774 & 0.8165 & 0.5774 & 1.0000 & 0.5774 \\ 0 & 0.7071 & 0 & 0.5774 & 1.0000 \end{bmatrix}$$

Part 2

(a) **User-user collaborative filtering**

We have

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

where

$$x_{ij} = \sum_{i=1}^{m} r_{ij} \cdot su_{ij}$$

and $r_{ij} \in R, su_{ij} \in S_U$. Need to be carefull for matrix dimension, thus we get

$$X = S_U \cdot R = (\sqrt{P}^{-1} \cdot R \cdot R^t \cdot \sqrt{P}^{-1}) \cdot R$$

(b) **Item-item collaborative filtering**
For the same reason, we have

$$x_{ij} = \sum_{j=1}^{n} r_{ij} \cdot si_{ij}$$

with $r_{ij} \in R, si_{ij} \in S_I$. Thus we get

$$X = R \cdot S_I = R \cdot (\sqrt{Q}^{-1} \cdot R^t \cdot R \cdot \sqrt{Q}^{-1})$$

Part 3

Let's apply the algorithms for user Bob (200-th row). I used Matlab for the implementation. See the code for more details.

(a) With the user-user collaborative filtering, I obtain these shows with these scores

TV-show	score
'FOX 28 News at 10pm'	882.8699
'Family Guy'	828.5152
'NBC 4 at Eleven'	780.2172
'2009 NCAA Basketball Tournament'	765.7426
'Access Hollywood'	750.3544

(b) With the item-item collaborative filtering, I obtain these shows with these scores

TV-show	score
'FOX 28 News at 10pm'	23.8006
'NBC 4 at Eleven'	22.6289
'Access Hollywood'	22.6210
'Family Guy'	22.6071
'Two and a Half Men'	22.0486

(c) It is easy to check the precision. For both filtering, I get a precision of 1 for $k = 5$. That means Bob have watched all movies that are proposed by the user-user or item-item filtering.