Quiz 2

Name:

1. Suppose we have a user-item rating table as follows:

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	2		4		4	1
User 2	3	1	3		1	2
User 3	4	2	3	1		1
User 4	3	3	2	1	3	1
User 5		3		1	2	
User 6	4	3		3	3	2
User 7		5		1	5	1

(1) Give an example of matrix factorization (i.e., P and Q) assuming the following model

$$r_{u,i} = q_i^T \cdot p_u$$

where

- $r_{u,i}$ = the rating of item i by user u
- q_i = 1-dimensional factor vector of item i
- p_u = 1-dimensional factor vector of user u

The factor matrix P and Q ***need not to be optimal***.

(sol)

$$P = \begin{bmatrix} 2.75 \\ 2 \\ 2.2 \\ 2 \\ 2 \\ 3 \\ 3 \end{bmatrix}$$

$$Q = [1\ 1\ 1\ 1\ 1\ 1]$$

(2) Compute the mean of squared errors (MSE; cost function) for the known ratings by your matrix factorization.

$$\mathrm{P} \cdot \mathrm{Q} = \begin{bmatrix} 2.75 & 2.75 & 2.75 & 2.75 & 2.75 & 2.75 \\ 2 & 2 & 2 & 2 & 2 & 2 \\ 2.2 & 2.2 & 2.2 & 2.2 & 2.2 & 2.2 \\ 2 & 2 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 & 3 & 3 \\ 3 & 3 & 3 & 3 & 3 & 3 \end{bmatrix}$$

$$MSE = \frac{1}{|\mathrm{T}|} \sum_{(u,i) \in \mathrm{T}} \left(\hat{r}_{u,i} - r_{u,i}\right)^2$$

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	2		4		4	1
User 2	3	1	3		1	2
User 3	4	2	3	1		1
User 4	3	3	2	1	3	1
User 5		3		1	2	
User 6	4	3		3	3	2
User 7		5		1	5	1

$\mathrm{MSE} = (2-2.75)^2 + (4-2.75)^2 + (4-2.75)^2 + (1-2.75)^2$

$+(3-2)^2 + (1-2)^2 + (3-2)^2 + (1-2)^2 + (2-2)^2$

$+(4-2.2)^2 + (2-2.2)^2 + (3-2.2)^2 + (1-2.2)^2 + (1-2.2)^2$

$+(3-2)^2 + (3-2)^2 + (2-2)^2 + (1-2)^2 + (3-2)^2 + (1-2)^2$

$+(3-2)^2 + (1-2)^2 + (2-2)^2 +$

$+(4-3)^2 + (3-3)^2 + (3-3)^2 + (3-3)^2 + (2-3)^2$

$+(5-3)^2 + (1-3)^2 + (5-3)^2 + (1-3)^2$

$= 42.55$

(3) Which book are you going to recommend for User1 between Item2 and Item4? Answer based on the matrix factorization suggested in the above.

(sol) For user1, item2 and item4 are both 2.75. Therefore, both items are recommended.

2. The matrix factorization that approximates user's rating on i based on the model $r_{u,i} = q_i^T \cdot p_u$ originally didn't consider the bias by users and items. By introducing user b_u and item biases b_i, the rating can be modeled as

$$r_{u,i} = q_i^T \cdot p_u + b_i + b_u + \mu$$

(1) Let us say that a user Max is very generous for rating. What kind of b_i (or b_u) will we obtain by matrix factorization?

(sol) For that user, the bias b_u can be high.

(2) Let us say that a movie Avatar was very popular, and everybody enjoyed it. What kind of b_i (or b_u) will we obtain by matrix factorization?

(sol) For a movie Avatar, bias b_i can be high.

3. Suppose that we utilize the model

$$r_{u,i} = q_i^T \cdot p_u + b_i + b_u + \mu$$

for approximating a user-item rating matrix. Moreover, we have a confidence value for each user u, which is denoted by c_u. If the confidence score on a user u is high, we can assume that the ratings on all items by u are very close to the true ratings. Define a cost function using the confidence.

(sol)

$$\sum_{(u,i)\in \mathrm{D}} c_u\left(r_{u,i} - q_i^T \cdot p_u - b_i - b_u - \mu\right)^2 + \lambda \sum_{u\in U} (\|p_u\|^2 + b_u^2) + \lambda \sum_{i\in I} \left(\|q_i\|^2 + b_i^2\right)$$