Assignment 3

Start Assignment

Due Friday by 11:59pm   **Points** 100   **Submitting** a file upload   **File Types** zip
Available Mar 25 at 12am - Apr 15 at 11:59pm 22 days

1. (10 points) Prove that Randomized Response is $\ln 3$-differentially private. Consider two neighboring databases $\mathcal{D}_1 = \mathcal{D} \cup \{x_1\}$, $\mathcal{D}_2 = \mathcal{D} \cup \{x_2\}$ where $x_1 \neq x_2$. Let $\mathbf{y}_1 = (y_{11}, \ldots, y_{1n})$ be the response of $\mathcal{D}_1$, where y_{1i} is the response of the i-th respondent of the survey. Similarly we have $\mathbf{y}_2$ for $\mathcal{D}_2$. Bound the ratio $\frac{\Pr(\mathbf{y}_1 = \mathbf{y}|\mathcal{D}_1)}{\Pr(\mathbf{y}_2 = \mathbf{y}|\mathcal{D}_2)}$.

2. (20 points) Find a singular value decomposition of the matrix $A = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$.

3. (10 points) Define the graph G_n to have the $2n$ nodes

$$a_0, a_1, \ldots, a_{n-1}, b_0, b_1, \ldots, b_{n-1}$$

and the following edges. Each node a_i, for $i = 0, 1, \ldots, n-1$, is connected to the nodes b_j and b_k, where

$$j = 2i \bmod n \text{ and } k = (2i+1) \bmod n$$

For instance, the graph G_4 has the following edges: $(a_0, b_0), (a_0, b_1), (a_1, b_2), (a_1, b_3), (a_2, b_0), (a_2, b_1), (a_3, b_2)$, and (a_3, b_3).

(a) Find a perfect matching for G_4. (5 points)

(b) Find a perfect matching for G_5. (5 points)

4. (60 points) Download the files "training.txt", "testing.txt" and "item_tag.txt". In the file "training.txt", each line is in the form

$$u, i, r_{ui}$$

which means the rating of user u on the item i is r_{ui}. In the file "testing.txt", each line is represented by

$$u, i, ?$$

which means you are required to predict the rating of user u on the item i. Use the training dataset "training.txt" to build a recommender system and make predictions for the testing dataset "testing.txt" by replacing all the "?" with your predicted ratings. All the ratings are within the range $[0, 5]$.

You may also want to use the file "item_tag.txt", where each line

$$i{:}t_1, t_2, \ldots, t_k$$

indicates that the item i has tags $t_1, t_2, \ldots, t_k$. Note that some items may not have any tags so it is normal if you cannot find some items in the file "item_tag.txt".

(a) Describe your method of learning the recommender system. (20 points)

(b) As you only have the training data, to verify how good is your recommender system without the testing data, please adopt 10-fold cross-validation (google what is 10-fold cross-validation if you do not know) in the training data and report the average RMSE value on the validation datasets. (15 points)

(c) You need to submit your code as well as a readme file describing how to run your code. (10 points)

(d) After you replace all the "?" with your predicted ratings, save the results into a file called "prediction.txt" and submit it. The TA will compare your prediction with the ground truth, which is not accessible to you, and then grade your submission based on your prediction's performance w.r.t. RMSE. (15 points)