CSC6223 Privacy - Assignment 2
Deadline: Nov 30, 2018

1. k-anonymity and l-diversity

K-anonymity describes that for each person within the data, their information cannot be distinguished from at least k − 1 other individuals whose information also appears in the data. Note that "distinguishing" is defined over quasi-identifiers. Assume that for this exercise, the attributes age and gender are the quasi-identifiers.

$Dataset_1$

ID	Age	Gender	Fav.Show
1	12-15	female	Friends!
2	19-25	male	Friends!
3	19-25	male	Friends!
4	12-15	female	Friends!
5	19-25	male	G.o.T.
6	19-25	male	G.o.T.
7	19-25	male	G.o.T.

$Dataset_2$

ID	Age	Gender	Fav.Show
1	19-25	female	Grey's A.
2	19-25	female	Simpsons
3	19-25	female	Futurama
4	19-25	female	Friends!
5	19-25	male	G.o.T.
6	19-25	male	C.Minds
7	19-25	male	Br.Ba.

Table 1

a) Does Dataset1 satisfy k-anonymity? If so: what is the maximal k for which it satisfies k-anonymity? Explain your answer by giving the anonymity sets consisting of the equivalent identities.

b) Does Dataset2 satisfy k-anonymity? If so: what is the maximal k for which it satisfies k-anonymity? Explain your answer as above.

c) Does Dataset 1 and 2 satisfy l-diversity? If so: what is the maximal l for which it satisfies l-diversity? Explain your answer as above.

2. t-closeness KL Distance

Let P and Q be the distributions shown in the table. Calculate the KL distance between P and Q.

	0	1	2	3
Distribution P	0.1	0.2	0.3	0.4
Distribution Q	0.4	0.3	0.3	0.1

Table 2

3. t-closeness EMD

Table 3 is the original table, and Table 4 shows an anonymized version satisfying distinct and entropy 3- diversity. Q = {3k, 4k, 5k, 6k, 7k, 8k, 9k, 10k, 11k} and P3={7k, 9k, 10k}.

Calculate the D[P3, D] using EMD.

	ZIP Code	Age	Salary	Disease
1	47677	29	3K	gastric ulcer
2	47602	22	4K	gastritis
3	47678	27	5K	stomach cancer
4	47905	43	6K	gastritis
5	47909	52	11K	flu
6	47906	47	8K	bronchitis
7	47605	30	7K	bronchitis
8	47673	36	9K	pneumonia
9	47607	32	10K	stomach cancer

Table 3. Original Salary/Disease Table

	ZIP Code	Age	Salary	Disease
1	476**	2*	3K	gastric ulcer
2	476**	2*	4K	gastritis
3	476**	2*	5K	stomach cancer
4	4790*	≥ 40	6K	gastritis
5	4790*	≥ 40	11K	flu
6	4790*	≥ 40	8K	bronchitis
7	476**	3*	7K	bronchitis
8	476**	3*	9K	pneumonia
9	476**	3*	10K	stomach cancer

Table 4. A 3-diverse version of Table 3

4. Local Sensitivity

There is a dataset D = {1,1,2,2,4,6,7,8,8,8,9}. Please calculate the Count query local sensitivity and Median query local sensitivity.

5. Differential Privacy Definition

To protect Bob from opting in D or out of D', i.e., Bob's health status cannot be inferred confidently, we set Pr(Bob$\in D$)=0.3 and Pr(Bob$\in D'$)=0.6. Please calculate the differential privacy protection level.

6. Differential Privacy -- Random Response:

Study participants are told to report whether or not they have property P as follows:

1) Flip a coin.
2) If tails, then respond truthfully.
3) If heads, then flip a second coin and respond "Yes" if heads and "No" if tails.

Please calculate the differential privacy protection level.

7. Game Theory Concept

Draw the payoff table of rock-paper-scissors game. Does this game has a Nash Equilibrium? Explain the reason of your answer.