KFUPM
College of Computer Science and Engineering
Computer Engineering Department
COE 426/526: Data Privacy

Fall 2020 (201)

Assignment 2: Due date Saturday 31/10/2019

Tasks

Question1: Differentially Private Histogram (30 pts)
The goal of this task is to understand the concept of differential privacy by implementing a differentially private histogram using Laplace mechanism. The objective is to generate and draw noisy histogram bins.

Input

Your program takes the following as inputs

- Input dataset files (ipums.txt)
- Privacy budget ϵ
- Number and sizes of bins

Dataset description

The dataset used in this task is the IPUMS data extracted from the 2001 US Census. The dataset has 3 attributes as described in Table 1. The size of the dataset is 20,000 tuples (rows). All attributes include numerical values only. For example, Gender attributes can be either 1 or 2, which represents Male and Female, respectively. The Income attribute is the annual income in thousand USD, for example, an income of 20 means 20,000 (20K) annually.

Age	Gender	Income (K)

Table 1: Scheme of Census dataset

Output

The output of your program is a deferentially private histogram for each dimension. In other words, you need to draw a histogram for Age, Gender, and Income.

Question2: Relative Error (10 pts)
For each histogram you created in Task 1, compute the relative error compared to the histogram created using the original database.

(a) Use the Mean Squared Error (MSE) to compute the error, where MSE is given as the following:

$$MSE = \frac{1}{n}\sum_{i-1}^{n}(X_i - \widehat{X_i}) \tag{1}$$

where x_i is the count in the i^th bin from the original table, $\widehat{X_i}$ is the the count in the i^th bin from the private histogram, and n is the number of bins in the histogram.

(b) What is the effect of n on the value of MSE. In other words, does reducing the range of each bin change the MSE?

Question3: Deferentially Private Max Response (10 pts)
Implement the Report Noisy Max to find the the age range with highest population. Assume that the age ranges are uniform and of size 10, i.e., $\{[0-10), [10-20), [20-30), ...\}$

Submission

The due date of this assignment is 11:59PM 31/10/2019. Please upload all files on the assignment page on BlackBoard. You need to submit the following:

1. A brief report including the obtained deferentially private histograms from Question 1, and your answer to Question 2

2. A zip file that contains the PDF brief report, the source files of your code (in any language), and a README file explaining how to compile/run your program