CS590 DSP: Data Security and Privacy **Spring 2022**

Homework #5

Due date & time: 11:59pm on April 19, 2022. Please hand in your homework on Brightspace.

Late Policy: You have three extra days in total for all your homeworks and projects. Any portion of a day used counts as one day; that is, you have to use integer number of late days each time. If you exhaust your three late days, any late homework won't be graded.

Additional Instructions: (1) The submitted homework must be typed. Using Latex is recommended, but not required.

Problem 1 (25 pts) Read the paper "Understanding Hierarchical Methods for Differentially Private Histograms" by Qardaji et al., and answer the following questions.
http://www.vldb.org/pvldb/vol6/p1954-qardaji.pdf

- Explain why using a hierarchy helps reduce the error of range queries in one-dimensional dataset.
- Explain the key factors where one can optimize the accuracy of hierarchical methods, and their impacts.
- The analysis in the paper assumes that we want to optimize absolute (squared) error. What if we want to optimize for relative error, that is, absolute error divided by true answer, (or some constant σ when the true answer is below σ)? What kind of hierarchical method do you think would be the nest?

Problem 2 (25 pts) Read the paper "PriView: practical differentially private release of marginal contingency tables. " by Qardaji et al., and answer the following questions.
https://dl.acm.org/doi/10.1145/2588555.2588575

- Give a summary of how the PriView method work.
- Explain the different methods for reconstructing a k-way marginal, once a number of views have been obtained and made consistent. How well do they compare with each other?
- How to extend the method to deal with datasets that have attributes that not binary, especially numerical attributes?
- On what kinds of datasets and queries would the method in PriView result in high errors?

Problem 3 (50 pts) Study how Census Bureau applies differential privacy. Write a report analyzing how Census Bureau currently applies Differential Privacy, and in what ways can it be improved. The report should be at least 3 pages in single space and 11pt font, i.e., in the formatting of this assignment document. One starting point is
https://www.ncsl.org/research/redistricting/differential-privacy-for-census-data-explained.aspx