DSC 215

HW 10: QUIZ

WINTER 2024

Please fill your answers using an editor in the spaces provided in the PDF template. Then upload the result in PDF form to gradescope.

Name: Mansi Sharma **PID: A59020070**

Questions regarding privacy and fairness

Q1. Explain in words what the probability Prob($Q(D_{I\pm i}) = R$) is on slide 24 of Task.

The probability notation Prob(Q(D_I±i) = R) refers to the likelihood that a query algorithm Q yields a specific result R when it is executed on a dataset D. In this context, D_I signifies the dataset encompassing a group of individuals I, and D_I±i indicates the dataset with either the inclusion or exclusion of an individual i. This probability is a fundamental aspect of the concept of differential privacy.

The principle of differential privacy is designed to ensure that the addition or removal of any individual's data from a dataset does not significantly alter the output of any analysis conducted on that dataset, thereby safeguarding the privacy of the individuals within the dataset.

Q2. Explain why $\Delta F = 2$ on slide 35 of Task.

On slide 35, ΔF=2 is discussed in the context of adding Laplacian noise to achieve differential privacy. The value ΔF=2 represents the global sensitivity of a query function, indicating the maximum change in the output of the function for any two neighboring datasets that differ by only one element. In the context of differential privacy, especially within social network analysis, this measure quantifies the potential impact of a single individual's data on the overall outcome of the query. The value of 2 suggests that the addition or removal of a single data point (e.g., an individual's response) could change the query's result by at most a value of 2. This sensitivity level guides the amount of noise that needs to be added to the query's output to obscure the influence of any single individual's data, thus ensuring the privacy of the participants while still allowing for the utility of the aggregated data.

Q3. In the COMPAS case, explain why defendants care about the false positive rate while judges care about the false negative rate and COMPAS cares about the error rate.

Defendants are concerned about the false positive rate because:

- It measures the rate at which non-reoffenders are incorrectly assessed as high risk.
- High rates can lead to unfair, more severe consequences.

Judges are concerned about the false negative rate because:

- It indicates how often reoffenders are wrongly predicted as low risk.
- Low rates are essential to prevent potential reoffenses due to lenient rulings.

COMPAS developers focus on the overall error rate because:

- It reflects the accuracy of the tool in predicting reoffense risks.
- Lower rates signify a more reliable tool for guiding judicial decisions.

Q4. Explain why in the toy model of lending, one group can be unfairly denied loans.

In the context of a toy model of lending, one group can be unfairly denied loans due to:

- System Bias: If the decision-making system or model is biased, it might allocate loans unfairly or perpetuate existing inequalities.
- Data Skew: Historical data used for training the model might have inherent biases against a certain group, which the system then learns and perpetuates.
- Disparate Impact: Biases in the system can lead to certain groups experiencing higher error rates, like false positives, where individuals are incorrectly deemed high-risk and denied loans