CIS 6930: Privacy & Machine Learning (Fall 2019)
Homework 1 — Data Privacy

Name: Kiana Alikhademi

October 2, 2019

This is an individual assignment!

Instructions

Please read the instructions and questions carefully. Write your answers directly in the space provided. Compile the tex document and hand in the resulting PDF.

For this homework, you will solve several data privacy problems. The fourth problem asks you to implement a differential privacy mechanism using Python. Use the code skeleton provided and submit the completed source file(s) alongside with the PDF.[1] *Note: bonus points you get on this homework *do* carry across assignments/homework.*

[1] You are encouraged to use Python3. You may use HiPerGator or your own system.

Problem 1: Syntactic Metrics (20 pts)

Consider the data set depicted in Table 1. Answer the following questions. (Justify your answers as appropriate.)

Age	Zip Code	Sex	Credit Score	Yearly Income	Loan
30-39	32607	M	678	90k	Approved
30-39	32607	M	799	120k	Approved
40-49	32611	F	451	35k	Declined
20-29	32607	F	783	30k	Approved
20-29	32607	F	560	70k	Declined
40-49	32611	M	725	22k	Declined

Table 1: Anonymized Data Set.

1. (4 pts) What are the quasi-identifiers? What are the sensitive attributes?

   quasi-identifiers: Age, Zip code, Sex Sensitive attributes: Credit Score, Yearly Income, Loan

2. (4 pts) What is the largest integer k such that the data set satisfies k-anonymity? What is the largest integer l such that the data set satisfies l-diversity?

   k=1, l=1

3. (6 pts) Modify the data set using generalization and suppression to ensure that it satisfies 3-anonymity and 2-diversity. Here we are looking for a solution that minimally affects the utility of the data. Write the modified data set below.

Age	Zip Code	Sex
*	326**	M
*	326**	M
*	326**	F
*	326**	F
*	326**	F
*	326**	M

Table 2: Data Set after suppression and generalization

4. (6 pts) Your student friend Alice (who is not in the anonymized data set) was recently declined for a loan despite her 30k yearly income. She thinks she may have been discriminated against.

   The bank who declined Alice's loan has published the following transparency report about their loan approval model.

   - If `yearly_income` $\geq$ 50k then: return `APPROVED`
   - If `yearly_income` $\geq$ 25k:
     - If `student`:
       - If `credit_score` $\geq$ 550 then: return `APPROVED`

        * Else: return `DECLINED`
    - Else (not `student`) if `credit_score` $\geq$ 500 then: return `APPROVED`
- If `yearly_income` $\geq$ 20k and `credit_score` $\geq$ 650 then: return `APPROVED`
- return `DECLINED`

What can you infer about Alice assuming that the transparency report accurately reflects the loan approval model? What do you conclude about the possible tension between algorithmic transparency and privacy? (Explain your answer.)

> *The Alice credit score should be less than 550. Transparency in this case led to the possible tension as adversaries can fine the sensitive data as we found what is Alice credit score is.*

Problem 2: Randomized Response & Local Differential Privacy (25 pts)

Social science researchers at the University of Florida want to conduct a study to explore the prevalence of crime among students. Specifically they want to ask questions of the form: *have you ever committed crime X?* (Here X stands for a specific crime or crime category.)

Researchers are ethical so they want to carefully design the study to ensure that participants respond truthfully and that privacy is protected. They reached out to you, a data privacy expert, to evaluate their methods.

Consider a participant that is asked the question have you ever committed crime X? This question admits a yes or no answer. Before answering the participant is instructed to use the following algorithm to compute a "noisy" answer given their true answer and only report the noisy answer to the researchers.

NoisyAnswer(`true_answer`, $p \in (0,1)$):

- If `true_answer` is `NO`, then:
  - With probability p return `NO`
  - With probability $1-p$ return `YES`
- Else (if `true_answer` is `YES`), then: return `YES`

Answer the following questions.

1. (10 pts) Suppose the researchers obtain noisy answers $z_1, z_2, \ldots, z_n$ from the n study participants. You can assume that `YES` is encoded as 1 and `NO` is encoded as 0. Explain how the researchers can estimate the true proportion of `YES` from the noisy answers. Specifically, give formulae for (1) the expected number of `YES` answers and (2) the variance (or error) of the estimate.

		True Value		
		Yes	*No*	*Total*
Reported Value	*Yes*	1	$1-p$	$2-p$
	No	0	p	p
	Total	$1+p$	$1-p$	

To find the number of true yes responses, we should look at following metrics:

- *True Positive (TP): Proportion of the responses which are actually yes and returned as Yes.*
- *False Positive (FP): Proportion of the responses which are actually No but returned as Yes*

By considering these two metrics the proportion of true Yes will be $\frac{TP}{TP+FP}$ *which is* $\frac{1}{2-p}$*. Expected number will be* $\frac{n}{2-p}$*. variance could be computed using formula* $Var(X) = E(X^2) - (E(X))^2 = np(1-p)$ *will be equal to* $\frac{n}{2-p}(1-\frac{1}{2-p})$*.*

2. (5 pts) Consider the following definition of (Local) Differential Privacy.

Definition 1. *A randomized algorithm $\mathcal{F}$ which takes input in some set X satisfies* ε-differential privacy *(for some $\varepsilon > 0$) if for any two input records $x \in X$, $x' \in X$ and any output $z \in$ Range$(\mathcal{F})$:*

$$\Pr\{z = \mathcal{F}(x)\} \leq e^{\varepsilon} \Pr\{z = \mathcal{F}(x')\} \ .$$

Does the noisy answer algorithm satisfy Definition 1? Produce a proof or a counter-example. If it does, also give an expression for ε in terms of p.

counter examples to show this mechanism does not satisfy the ϵ -differential privacy is about following two cases:

- *the probability that reported value is No and True value is No over the probability that reported value is No and true value is Yes. These fraction will be $\frac{p}{0}$.*

3. (5 pts) Now consider the following (more general) variant of the algorithm.

> GeneralizedNoisyAnswer(`true_answer`, $p, p' \in (0, 1)$):
>
> - If `true_answer` is `NO`, then:
>   - With probability p return `NO`
>   - With probability $1 - p$ return `YES`
> - Else (if `true_answer` is `YES`), then:
>   - With probability p' return `NO`
>   - With probability $1 - p'$ return `YES`

Prove that this general variant satisfies ε-(local) differential privacy (Definition 1). Give an expression for ε in terms of p and p'.

To show that what is the ε with respect to the p and p', I computed the two following relations:

(a) $\frac{PrYes|Yes}{PrNo|Yes}$

(b) $\frac{PrYes|No}{PrNo|No}$

Part a and b are equal with $\frac{1-p'}{p'}$ and $\frac{1-p}{p}$ respectively. To satisfy the differential privacy the following relations should be satisfied:

- $\epsilon \geq log(1\text{-}p'_{\overline{p'}})$
- $\epsilon \geq log(1\text{-}p_{\overline{p}})$

4. (5 pts) Suppose we can arbitrarily set p and p'. Explain the trade-off between minimizing ε and minimizing the error between the true answers and the one estimated from noisy answers.

There is a trade off between the epsilon and the difference of true and noisy answers. With high privacy(low epsilon) the difference is high between true and noisy answers. On the other side, the true and noisy answers do not differ that much when the privacy is low(epsilon high)

Problem 3: Privacy & Sampling (30 pts)

Consider the data set shown in Table 3 and the function $f(\mathbf{x}) = \text{mode}(\mathbf{x})$. Here the mode of a dataset is 0 if the proportion of individuals who are HIV negative (-) is higher than 0.5, 1 otherwise.

	HIV	
	+	-
# of individuals	7	23

Table 3: Data set.

We are interested in various ways of designing a differentially private mechanism to compute f on an arbitrary data set of the same form as the one in Table 3.

1. (2 pts) What is the *local* sensitivity of f (with respect to the data set shown in Table 3)?

   If we add one more sample to minority and remove one, the majority is still the negative case and therefore the local sensitivity will be 0.

2. (2 pts) What is the global sensitivity of f? (Justify your answer.)

   We can completely change the dataset so at most the global sensitivity will be 1.

3. (3 pts) Consider the mechanism defined by $\mathcal{F}(\mathbf{x}) = f(\mathbf{x})$. Does this mechanism satisfy ε-differential privacy? Why or why not?

   If we consider that the majority of dataset X is positive and the majority of X' is negative, this lead to $1/0$ which does not satisfy the differential privacy. As it is constant mechanism and not randomized it will not preserve the data privacy.

Now consider your answers to the previous questions. We are interested in using Laplace noise to obtain a ε-differentially private mechanism for f.

4. (3 pts) Explain how you could add Laplace noise to obtain ε-differential privacy for f. Call the resulting mechanism $\mathcal{F}$.

   $F(x) = f(x) + Z$ in which Z is the Laplace noise. Z $Lap(b)$ in which b is equal with global sensitivity divided by epsilon. So $F(x)$ will be as follow:

$$F(X) = Mod(x) + \frac{1}{2b}\exp(-\frac{z}{b}) \tag{1}$$

5. (5 pts) Now consider the following post-processing step (after adding Laplace noise as you explained): return 0 if $\mathcal{F}(\mathbf{x}) < 0.5$ and 1 otherwise. If the data set $\mathbf{x}$ is such that $f(\mathbf{x}) = 1$, what is the probability that (after the post-processing step) the output is 0? What do you conclude about this mechanism?

   $F(x)$ which is $f(x) + Z$ needs to be less than 0.5 to output 0. By replacing the mode value we will find that we need to find the probability that $Z < -0.5$. The Integral of Laplacian PDf function will help us to reach our goal. The integral is as follow:

$$\int_{-\infty}^{-0.5} \frac{1}{2b} e^{\frac{-|Z|}{b}} dz \tag{2}$$

The result of integral will be equal to the following equation:

$$-\frac{Ze^{\frac{-|z|}{b}})}{2|Z|} + C \tag{3}$$

If our ϵ is assumed to be equal to 0.5, the b will be equal to 2. By substituting the value of b, upper bound and lower bounds of integral the final result will be as follow:

$$\frac{e^{\frac{-1}{4}}}{2} \cong 0.39 \tag{4}$$

6. (5 pts) Can you come up with a different ε-differential privacy mechanism for f that adds Laplace noise but provides more accurate outputs?

   We can change the way that outputs are produced in F(x). For instance, we can use Quartiles of data to map into 0 or 1. The mapping function will be as follows:

7. (10 pts) Finally, consider the following mechanism.

   SampleAndComputeMode(data set $\mathbf{x}$, $p \in (0, 1)$):
   - Let $\mathcal{M}$ be a ε-differentially private mechanism to compute the mode of a data set.
   - Let $\mathbf{s}$ be the data set obtained by independently selecting each record of $\mathbf{x}$ with probability p. (For each record, we flip a coin with probability of heads p, if heads then we add this record to $\mathbf{s}$, otherwise we do not.)
   - return $\mathcal{M}(\mathbf{s})$.

   Prove that SampleAndComputeMode() satisfies ε'-differential privacy and give an expression for ε' in terms of p and ε.

   According to naive composition, the ϵ' should be equal with len(s)ϵ.*

Problem 4: Implementing DP Mechanisms (25 pts)

For this problem you will implement several differential privacy mechanisms we talked about in class. Please use the comments in the Python files provided to guide you in the implementation.

For this question, we will use the dataset `data/ds.csv`. It contains pairs of age and yearly income for several individuals. For the purpose of calculating sensitivity, assume that the age range for any individual is [16, 100] and the yearly income range is [0, 1000000].

1. (5 pts) Fill in the implementation of `laplace_mech()`, `gaussian_mech()`. Also fill in the (global) sensitivity in the `mean_age_query()` function.

   You can test your implementation by running: '`python3 hw1.py problem4.1`'.

   How close are the noisy answers to the true answer?

   Problem 4.1: true mean age 38.10, laplace noisy answer: 39.09, gaussian noisy answer: 38.10 [epsilon = 1.000, log2 delta = -20.0, sensitivity = 2.000]

2. (5 pts) Complete the implementation of the `dp_accuracy_plot()` and run it for $\varepsilon = 0.1, 0.5, 1.0, 5.0$ on `mean_age_query()`. Paste the plots below.

   To run the code: '`python3 hw1.py problem4.2 <epsilon>`'. By default, figures are saved in `./plots` and named based on the value of ε.

   What do you conclude?

   By increasing the ϵ accuracy decreases. Lower value of ϵ shows higher privacy and ultimately higher accuracy.

3. (5 pts) Implement the function called `budget_plot()`. Use it to produce a plot of the budget of naive composition and advanced composition (refer to the course materials for details) when using `gaussian_mech()` to perform `mean_age_query()` $m > 1$ times. Plot the naive composition and advanced composition budgets (i.e., total ε) for varying m from 1 to 100 keeping $\delta \leq 2^{-30}$. Paste the plot below. For what values of n is naive composition better than advanced composition? (Justify your answer.)

   The answer depends on the value of ϵ as it is shown in 2. For ϵ with value of 0.1 we will have better navies for greater than 40. In the advanced compositions formula there is not a linear relation between m and the ϵ' however it is heavilly depend on the ϵ.

4. (10 pts) Finally, suppose we want to compute the average ratio of yearly income and age in the dataset, i.e., how many extra dollars does one earn for an increase of one year of age (on average). Consider two ways of performing this query with differential privacy:

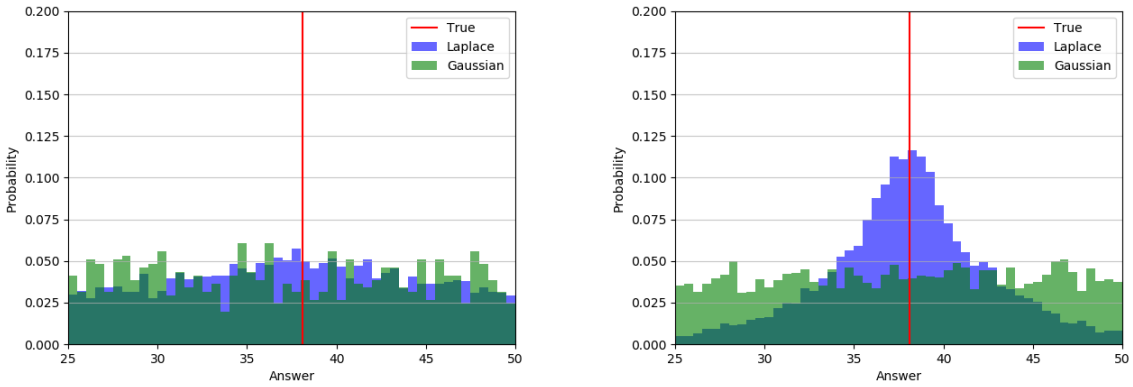
   (a) Compute the (global) sensitivity of this query (`income_per_age_query()`) and use the Laplace mechanism.

   (b) Use the Laplace mechanism to compute the mean yearly income. Use the Laplace mechanism to compute the mean age. Divide the two (noisy) results to obtain the ratio.

   Implement this functionality in `income_per_age_comp()`. Feel free to modify the signature of `income_per_age_comp()` and the corresponding code in `main()`. Set $\varepsilon = 1$. Paste the comparison plot below.

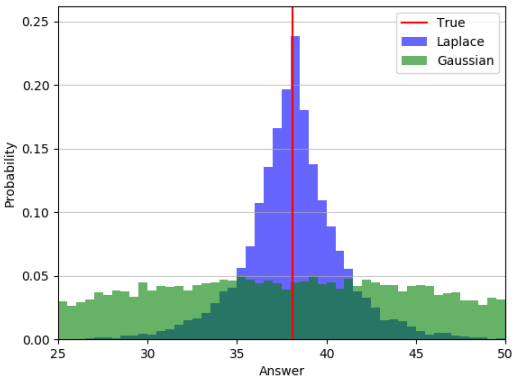
   What do you conclude?

   We can conclude that the two different types of query functions we tested yield almost the same results. However the mean age over mean income query seems more centerred.

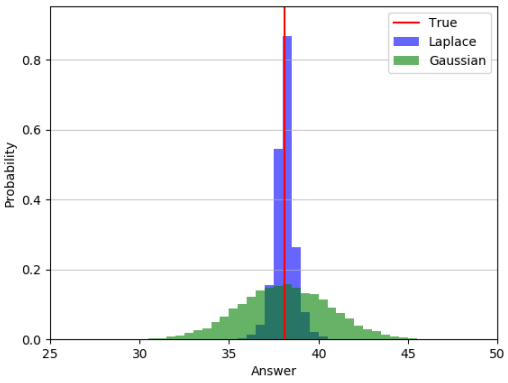


(a) $\epsilon = 0.1$

(b) $\epsilon = 0.5$



(c) $\epsilon = 1.0$



(d) $\epsilon = 5.0$

Figure 1: Accuracy plots

[Bonus] Problem 5: Privacy with Binomial Noise (20 pts)

Suppose we are interested in non-negative count functions f. For example f is the number of records in the dataset which satisfy some property P. Consider the mechanism $\mathcal{F}(\mathbf{x}) = f(\mathbf{x}) + B$, where $B \sim \text{Binom}(n, p) - \mathbb{E}[\text{Binom}(n, p)]$. Here $n = |\mathbf{x}| > 0$ is the size of the dataset and $p \in (0, 1)$ is a parameter. In other words $\mathcal{F}$ adds noise from the binomial distribution but centered at 0.

1. (2 pts) How would you set the value of p? (Explain your answer.)

   Your answer here.

2. (3 pts) Under what condition(s) does $\mathcal{F}$ satisfy ε-differential privacy? (Justify your answer.)

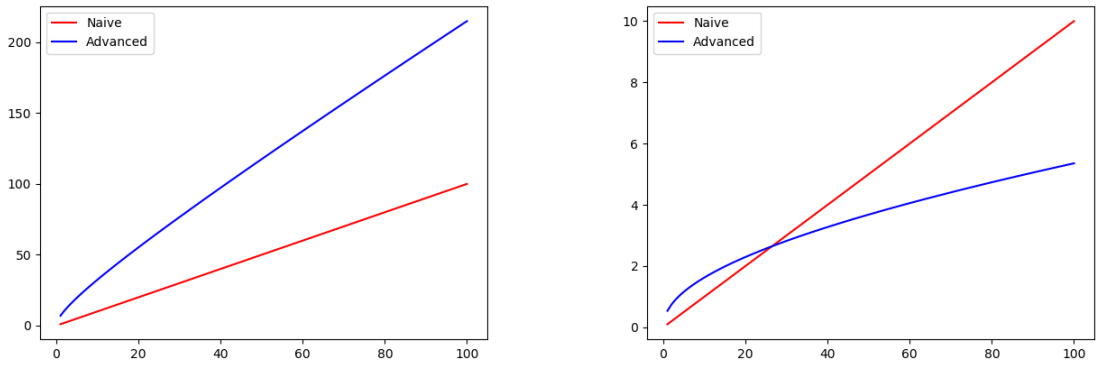
   Your answer here.

3. (10 pts) Prove that $\mathcal{F}$ satisfies (ε, δ)-differential privacy as long as $p \neq 0, 1$.

   Your answer here.

4. (5 pts) Characterize the trade-off between ε and δ.

   Your answer here.



(a) $\epsilon = 1.0$ (b) $\epsilon = 0.1$

Figure 2: Budget plots

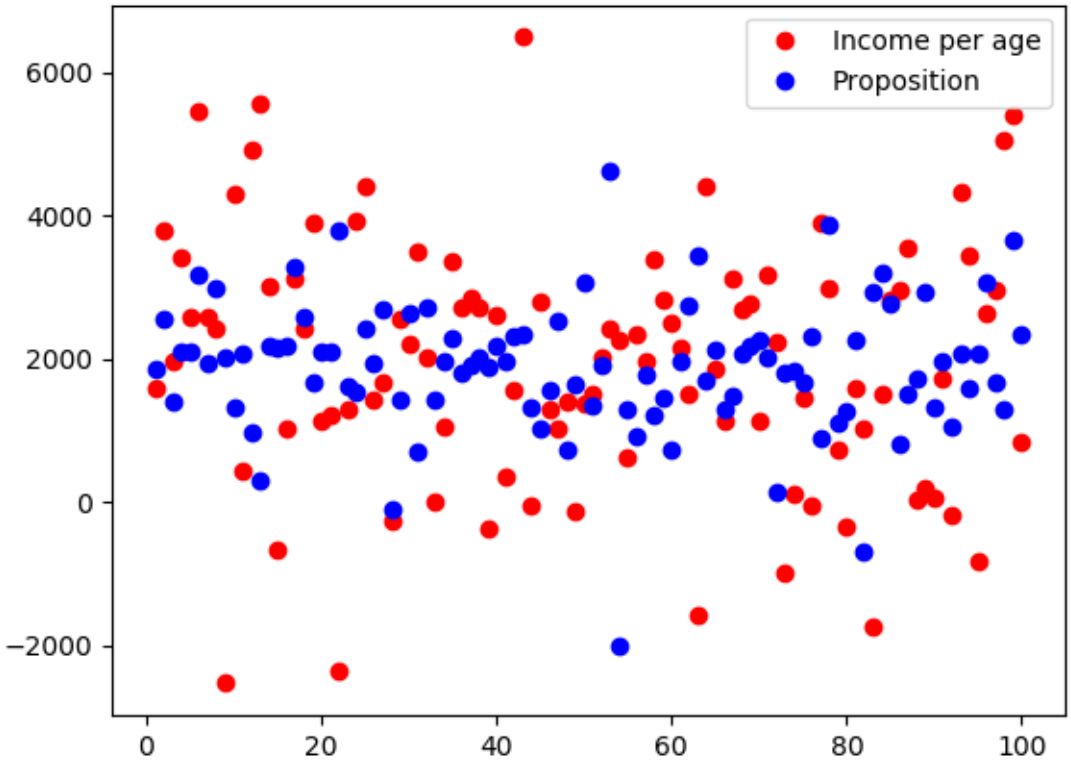


Figure 3: comparison between income per age query and mean age and income query