Date: 27/06/2023

**Summary of Work**

1. **Basic Differential Privacy Example**

Link: <https://towardsdatascience.com/a-differential-privacy-example-for-beginners-ef3c23f69401>

As a simple example, imagine this: you are in a school with a total student body of 300 people. Each of you are asked, “Have you ever cheated on a test?”

See how this can be a sensitive question? Perhaps those who have cheated would be reluctant to respond yes, out of fear for potential repercussions. So… how do we resolve this issue? This is where DP comes in handy.

Each student to flips a coin. If they land on heads, they tell the truth. If they land on tails, they flip another coin; if they land on heads, respond yes; tails, no.

Based on this flip coin it answers the question. The answer is some random. Based on the probability calculation we can get exact answer that how many students cheated on the exam and there is no loss of user data (Who is cheated or not?).

So basically, differential privacy is that you query on the database and from the database you get the data with added noise. Where users’ data privacy is guaranteed.

1. **Learning through online videos**

I started the course in differential privacy by Gautam Kamath (Assistant Professor of Computer Science at the University of Waterloo).

Link: <https://www.youtube.com/playlist?list=PLmd_zeMNzSvRRNpoEWkVo6QY_6rR3SHjp>

I watched the up to Lecture 5B. In which explains Why differential privacy is needed through the attacks on the published dataset like Linkage attack, memorization attack etc. Then explains about the differential privacy definition and some of the mechanisms.

1. **"Practicing Differential Privacy in Health Care: A Review" paper**

Link: <https://www.tdp.cat/issues11/tdp.a129a13.pdf>

In this there is the explanations about the differential privacy and some mechanisms of it. And there is limitations of the differential privacy and challenges and example in health application.

From this I have basic idea about the differential privacy. As an analyst we want to analyze the health data of patients so we run some queries on the database.

**Example:**

An investigator would run a query on such a database at a recruiting site to identify the number of participants which would meet the eligibility criteria. A study would have a certain threshold of patients that must meet the eligibility criteria for the site to be considered suitable for the trial. If the number of eligible patients is larger than the threshold then the sponsor will consider the site for inclusion in the trial. Running these eligibility ʹcountʹ queries would allow a sponsor to quickly identify candidate sites for a trial that have a sufficiently large patient population. However, providing unrestricted access to count queries on a database can reveal personal information. Differential privacy is used to hide the user database.

Differential privacy is used to add the random noise in the data of output queries.

1. **"Differential Privacy for Text Analytics via Natural Text Sanitization" paper.**

In this paper as per my understanding first they sanitize the texts (using differential privacy) and then using this sanitized text pretrain and fine tune the model. Also, they provide the mechanisms to sanitize the text.

1. **"FOX: Fooling with Explanations: Privacy Protection with Adversarial Reactions in Social Media" paper.**

In this paper an adversarial noise (new comment added to the post) to hide the user's sensitive data. A user who wants to protect their profile against attributes inference attacks may employ fake profiles or ask a list of trusted friends to add the reactions recommended by FOX.

1. **Questions**

My query is that the how differential privacy is apply to these comments scenario?

I am confused in this because that the differential privacy hides the user data by adding random noise to protect users’ privacy loss But as in the Fox: Fooling with examples research paper we add the new comment to hide the user’s sensitive information.

We have to use differential privacy to generate new comments or we have to train ML model using differential privacy to generate new comments?