Source folder bookmarked

Slide 2:

Differential Privacy attempts to solve the problem of protecting individuals privacy, while still being able to use and analyze their data to help solve real problems. An example of this could be the use of medical records, which hold confidential personal information to identify trends in certain regions. This data can be used to figure out the cause of a certain medical problem in a region.

Slide 3:

A common suggestion to protect people’s data while still using it is by simply making it anonymous, or de-identifying it. This can be done a various of ways, for example, releasing a dataset that is stripped of any identification information, such as names, usernames, etc. However, this has been proven time and time again that it does not work. In the 2000’s, Netflix ran a competition now dubbed, “The Netflix Prize,” where they challenged computer scientists to create a better film recommendation algorithm. To do this, they released a dataset of viewers which had been ‘de-identified’. However, researchers at the University of Texas at Austin were able to identify users by cross referencing the ratings data to iMDB. [Netflix Prize Source]

Slide 4:

Further proof that de-identifying data does not work can be backed up by research completed on the 1990 census by Latanya Sweeney. In her research, she found that 87 percent of people were identifiable based on the following information: Their 5 digit zip code, their gender, and their date of birth. Additionally, she found that 53 percent of people were identifiable by a general place, gender, and their date of birth. The general place can mean the city, town, or municipality where the subject resides.

Slide 5:

Sweeney was able to identify the governor of Massachusetts William Weld’s medical records that were apart of a “de-identified” dataset. An anonymized dataset of the health records of state employees of Massachusetts was released. Sweeney was able to use Weld’s zip code, date of birth, and gender to narrow down the dataset to one record.

Slide 6:

The problem of analyzing data without sacrificing privacy has existed for decades, and it was not until 2003 when researchers Kobbi Nissim and Irit Dinur discovered it was impossible to avoid revealing some private information when working with private databases. [Link article] They discovered what is known as the Fundamental Law of Information Recovery. [Link article] This was the first step to the breakthrough of differential privacy, as they found that in order to protect privacy, some sort of statistical noise must be injected into the database at question.

Slide 7:

The work by Nissim and Dinur paved the way for Dwork’s team of researchers to breakthrough and discover a method to protect privacy and enable the use of that data simultaneously. Dwork and her team found the amount of noise that would need to be added to a dataset and a mechanism for doing so. These developments would go on to win the prestigious Gödel Prize in 2017, a prize for papers in theoretical computer science.

Slide 8:

A very simplified example of how differential privacy works is by using a hypothetical survey of one yes or no question. Once the subject’s answer is acquired, a coin is flipped. If the result is heads, then the answer is kept the same. If the answer is tails, then another coin flip will take place. On this second coin flip, if the result is heads, then the original answer is kept. If the result is tails again, then the answer, for example ‘yes’, will be changed to ‘no’. The result of this is that in a large enough dataset, the result will still be very similar to if the original answers had been used. However, each individual answer cannot be trusted as being authentic, as there is a 25% chance that it is opposite of what the true answer is.

Slide 9:

Dwork’s team essentially figured out the amount of noise that would need to be injected into a dataset to achieve differential privacy. D1 and D2 are the datasets that differ by a single element. [explain e^epsilon ratio, randomized algorithm, etc.]