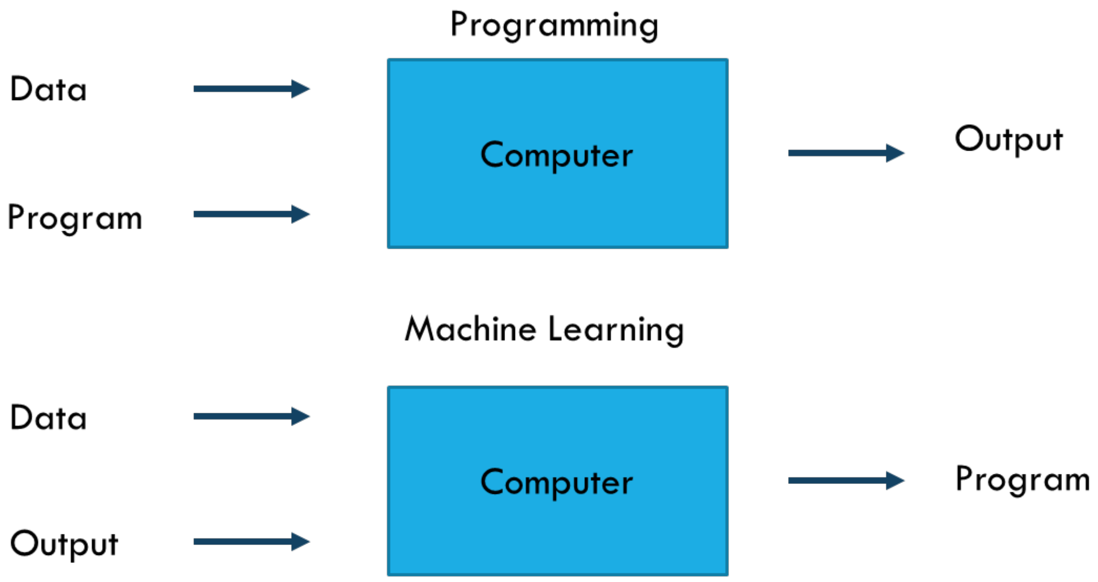
PRIVACY AND ML

Manolis Terrovitis
Research Director
Athena Research Center
mter@athenarc.gr

MACHINE LEARNING



ML is used when:

- Human expertise does not exist (navigating on Mars)
- Humans can't explain their expertise (speech recognition)
- Models must be customized (personalized medicine)
- Models are based on huge amounts of data (genomics)

Learning isn't always useful:

- There is no need to "learn" to calculate payroll

FEDERATED LEARNING

Centralized

- There is a centralized server to coordinate learning

Decentralized

- No centralized server

Heterogeneous

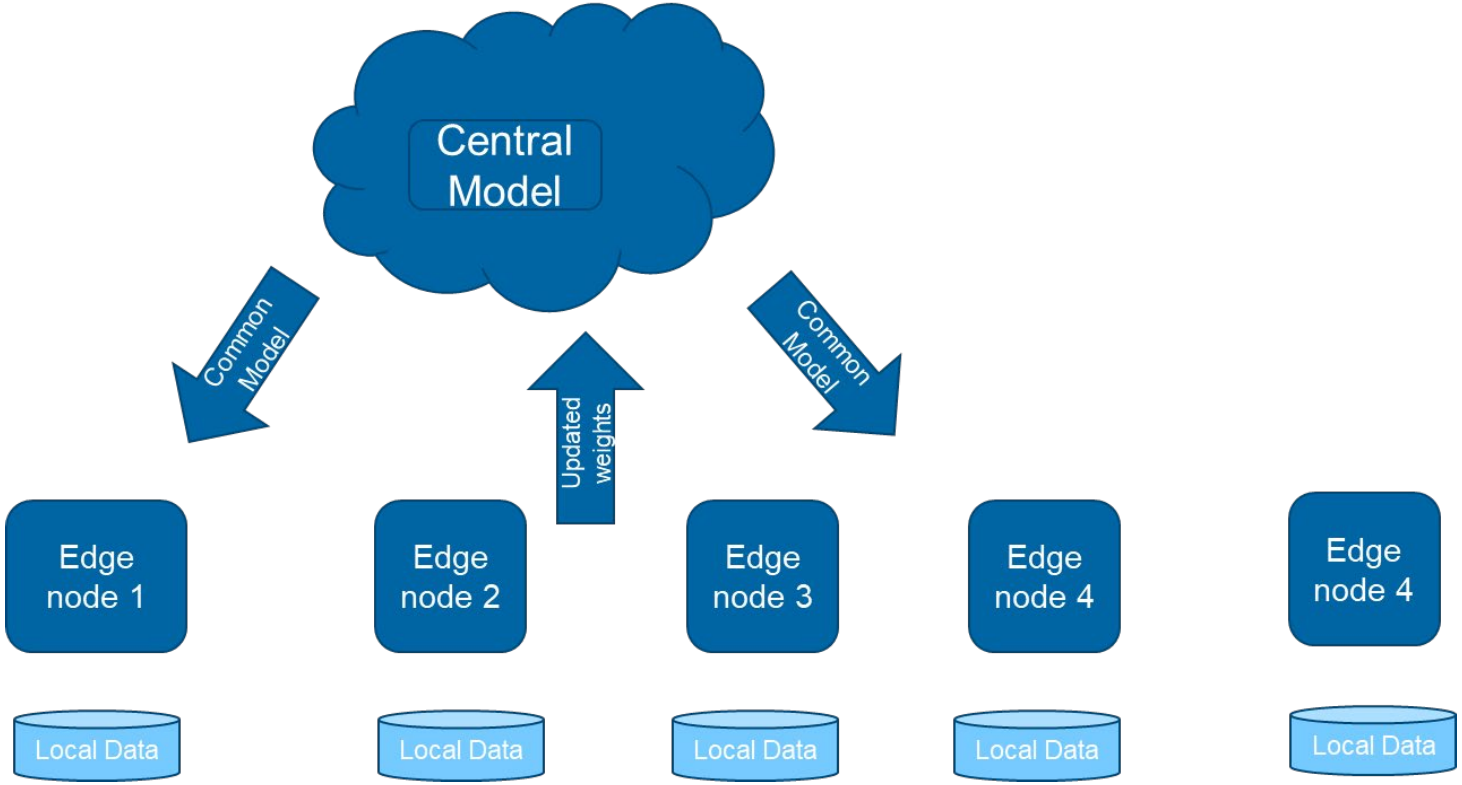
- Nodes with different capabilities, i.e., mobile phones and IoT

Systems Heterogeneity

Statistical Heterogeneity

Privacy issues

- *DP*
- *MPC*



ATTACKS ON DATA

Membership inference attack

- Adversarial goal: determine whether or not an individual data instance x^* is part of the training dataset $\mathcal{D}$ for a model
- The attack typically assumes black-box query access to the model
- Cause: overfitting

Reconstruction attacks

- Given output labels and partial knowledge of some features, try to recover sensitive features or the full data sample

Property inference

- The ability to extract dataset properties which were not explicitly encoded as features or were not correlated to the learning task, is called property inference. An example of property inference is the extraction of information about the ratio of women and men in a patient dataset when this information was not an encoded attribute or a label of the dataset

PROTECTION STRATEGIES

- **Differentially Private Stochastic Gradient Descent (DP-SGD)**
  - Add **noise** to the gradients during training.
  - Apply **clipping** to bound each individual's contribution before adding noise.
  - **Tools & Examples:**
    - **TensorFlow Privacy** and **PyTorch Opacus** implement DP-SGD.
    - Used in **Google's GBoard keyboard** to train next-word prediction models under local DP.
- **Private Aggregation of Teacher Ensembles (PATE)**
  - Train multiple teacher models on disjoint subsets of sensitive data.
  - Aggregate their predictions (with added noise) to label public/unlabeled data.
  - Train a student model on the noisy labels → inherits DP guarantees.
  - Examples:
    - OpenAI and Google have explored PATE for training private classifiers.
    - Works well when unlabeled public data is available.

  - **Anonymize training data**
    - Model inherits the guarantee provided for the data

Personal data which have undergone **pseudonymisation,** which could be attributed to a natural person by the use of additional information should be considered to be **information on an identifiable natural person.** [3]

To determine whether a natural person is identifiable, account should be taken of all the**means reasonably likely to be used**, such as **singling out,** either by the controller or by another person to identify the natural person directly or indirectly.

[4]To ascertain whether means are reasonably likely to be used to identify the natural person, **account should be taken of all objective factors,** such as the **costs** of and the amount of **time** required for identification, taking into consideration the available **technology** at the time of the processing and technological developments. [5]

The principles of data protection should therefore not apply to anonymous information , namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. [6]This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.

ANONYMOUS PERSONAL DATA – RECITAL 26

PSEUDONYMIZATION VS ANONYMIZATION

Pseudonymization

Direct identifiers are replaced by arbitrary ones, which should be kept separately.

“the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information.”

(WG) Pseudonymisation is also addressed to clarify some pitfalls and misconceptions: pseudonymisation is not a method of anonymisation. It merely reduces the linkability of a dataset with the original identity of a data subject, and is accordingly a useful security measure.

Anonymization

Irreversible

No longer personal data

NOT SO SIMPLE ARTICLE 29 DATA PROTECTION WORKING PARTY (0829/14/EN WP216)

Several anonymization techniques may be envisaged, there is no prescriptive standard in EU legislation.

Importance should be attached to contextual elements: account must be taken of "all" the means "likely reasonably" to be used for identification by the controller and third parties, paying special attention to what has lately become, in the current state of technology, "likely reasonably" (given the increase in computational power and tools available).

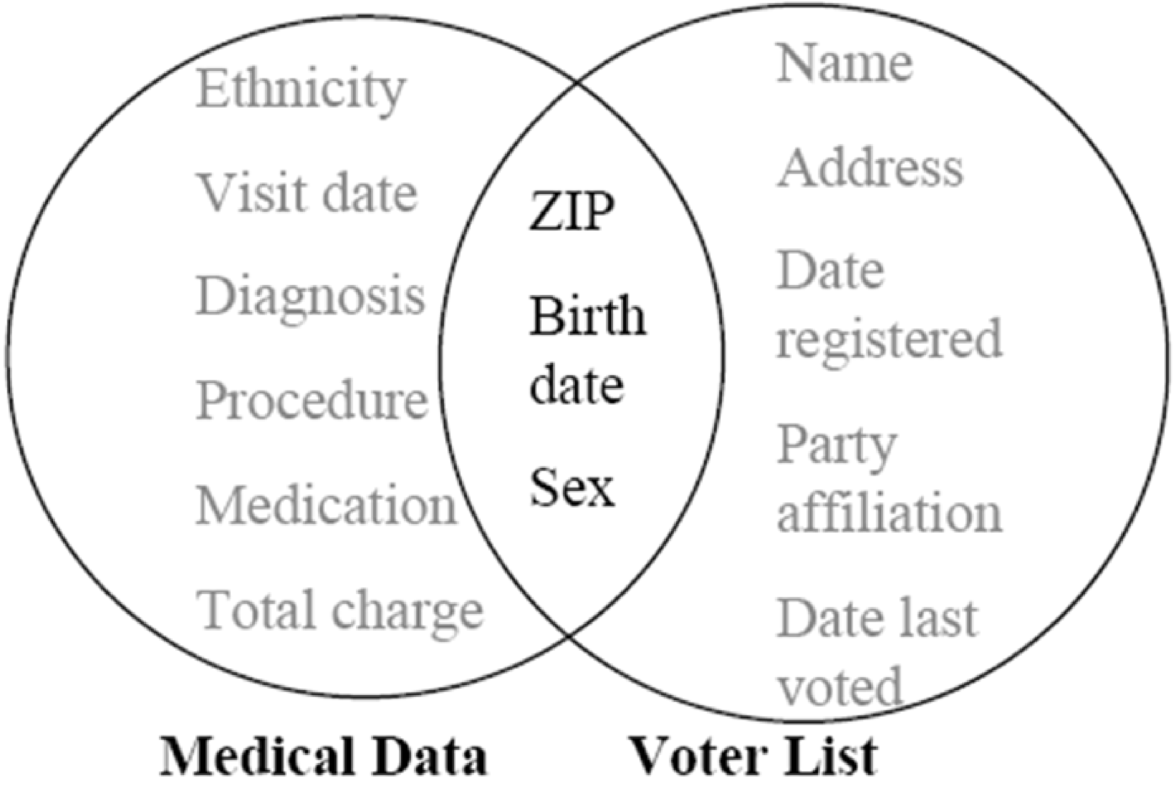
A risk factor is inherent to anonymisation: this risk factor is to be considered in assessing the validity of any anonymisation technique - including the possible uses of any data that is "anonymised" by way of such technique - and severity and likelihood of this risk should be assessed.

PROBLEMS

An effective anonymisation solution prevents all parties from singling out an individual in a dataset, from linking two records within a dataset (or between two separate datasets) and from INFERRING any information in such dataset

Inferring?

PSEUDONYMIZATION - LINK ATTACKS



DIFFERENT FLAVORS

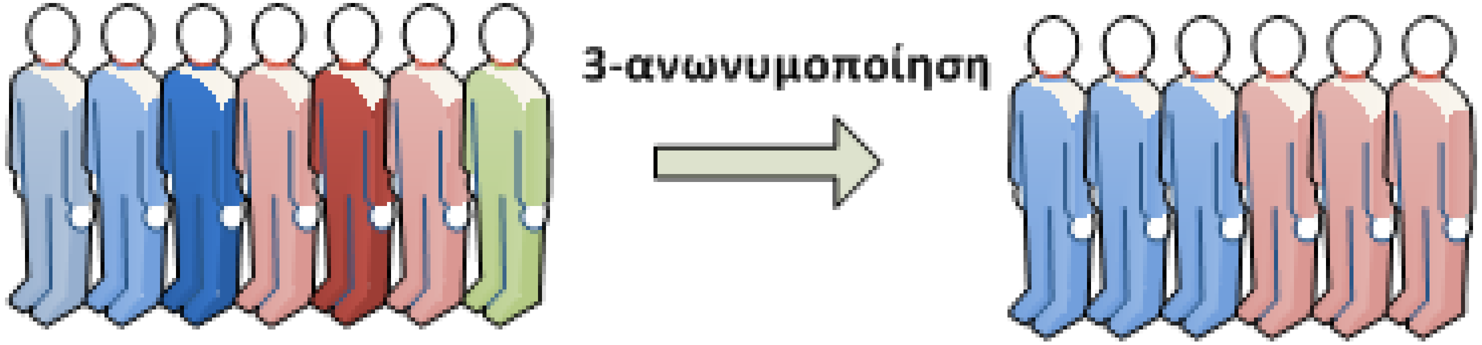
- k-anonymity
- k^m-anonymity
- l-diversity
- t-closeness
- Differential privacy

K-ANONYMITY

Suggested in 2001

Intuitively each person is hidden in a group of similar person

Straightforward interpretation of GDPR



K-ANONYMITY

Each entry becomes indistinguishable from other *k*-1 entries

- *k*-anonymity is achieved through suppression and generalization

id	Zipcode	Age	National.	Disease
1	13053	28	Russian	Heart Disease
2	13068	29	American	Heart Disease
3	13068	21	Japanese	Viral Infection
4	13053	23	American	Viral Infection
5	14853	50	Indian	Cancer
6	14853	55	Russian	Heart Disease
7	14850	47	American	Viral Infection
8	14850	49	American	Viral Infection
9	13053	31	American	Cancer
10	13053	37	Indian	Cancer
11	13068	36	Japanese	Cancer
12	13068	35	American	Cancer

~~id~~	Zipcode	Age	National.	Disease
~~1~~	130**	<30	∗	Heart Disease
~~2~~	130**	<30	∗	Heart Disease
~~3~~	130**	<30	∗	Viral Infection
~~4~~	130**	<30	∗	Viral Infection
~~5~~	1485*	≥40	∗	Cancer
~~6~~	1485*	≥40	∗	Heart Disease
~~7~~	1485*	≥40	∗	Viral Infection
~~8~~	1485*	≥40	∗	Viral Infection
~~9~~	130**	3∗	∗	Cancer
~~10~~	130**	3∗	∗	Cancer
~~11~~	130**	3∗	∗	Cancer
~~12~~	130**	3∗	∗	Cancer

DIFFERENTIAL PRIVACY

a mathematical definition for the privacy loss associated with any data release drawn from a statistical database

The intuition for the definition of ε-differential privacy is that a person's privacy cannot be compromised by a statistical release if their data are not in the database. Therefore, with differential privacy, the goal is to give each individual roughly the same privacy that would result from having their data removed.

Key idea: **the results of a query or an algorithm should be more or less the same whether a single individual’s record participates in the input or not**

DIFFERENTIAL PRIVACY

[DWORK'06]

Randomized Mechanism *K* provides individual, if individual's data effects the of ℛ only "little"

"adjacent" means "differ in one individual's entry"

K: 𝒟 → ℛ ensures *ε*-DP if for all adjacent datasets D_1, D_2 and for all subsets *S* of ℛ:

$$\frac{\Pr[K(D_1) \in S]}{\Pr[K(D_2) \in S]} \leq e^{\varepsilon}$$

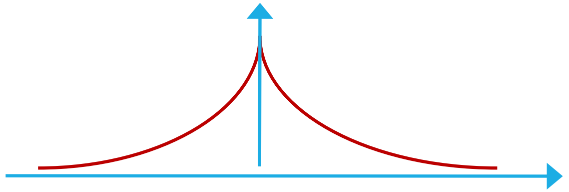
HOW TO CREATE A RANDOMIZED MECHANISM?

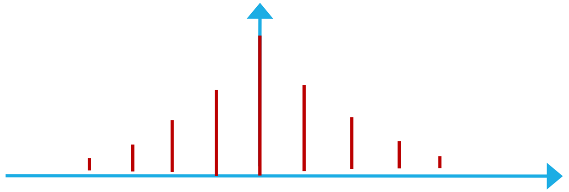
How to randomize a count query?

- Preserve accuracy?
- Guarantee ε-differential privacy

For every value that is output:

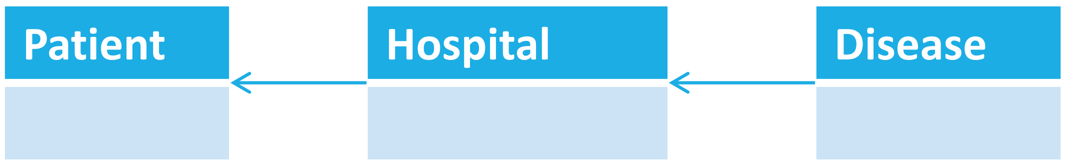
- Add Laplacian noise Lap(ε/s):
- Or Geometric noise for discrete case:





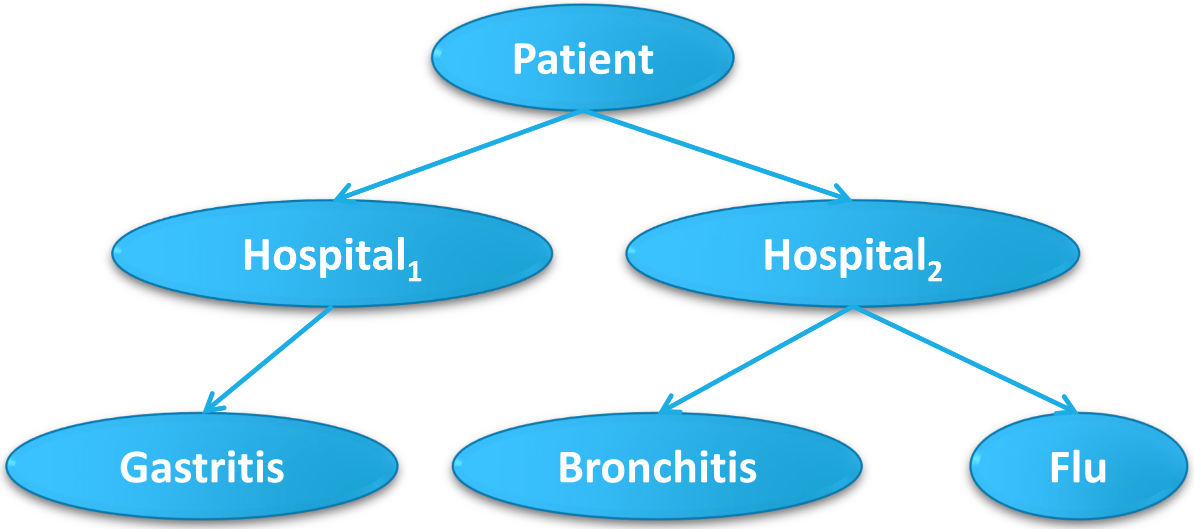
EXAMPLE

- record = unordered, non-recursive, attribute tree
- XML / DB with various tables



TREE-STRUCTURED DATA

- record = unordered, non-recursive, attribute tree
- XML / DB with multiple tables



IMPORTANT APPLICATIONS

- **U.S. Census Bureau (2020 Census):**
  For the first time, the Bureau applied differential privacy to protect individuals in published census tables, ensuring strong privacy guarantees while still enabling demographic analysis.
- **Apple (iOS & macOS):**
  Uses DP for collecting statistics on emoji usage, Safari domains, and keyboard typing patterns to improve services without identifying individuals.
- **Google (Chrome & GBoard):**
  - **Chrome:** Uses DP in telemetry collection (e.g., which domains crash browsers).
  - **GBoard (keyboard):** Applies local differential privacy when collecting typing and emoji usage to improve autocomplete and prediction.
- **HIPAA and EMA** prescribe rule-based anonymization, reflecting general population statistics

PROBLEMS IN PRACTICE

- What are best parameters?
- What is best guarantee?
- Data is usually sparse
- Deciding on quasi identifiers is far from trivial
  - Lack of documented experience
- Low tolerance for reduced quality data

AMNESIA

Guarantees:

- K-anonymity
- Km-anonymity
- Object relational datasets
- Disk based algorithm

API

ReST and command line API exist to help programmers

Upcoming

- Differential privacy
- DICOM metadata anonymization

Popularity

- 150K+ visitors
- 500k visits