3iA Côte d'Azur
Institut interdisciplinaire d'intelligence artificielle

Ayşe Ünsal

Adversarial learning & Differential privacy

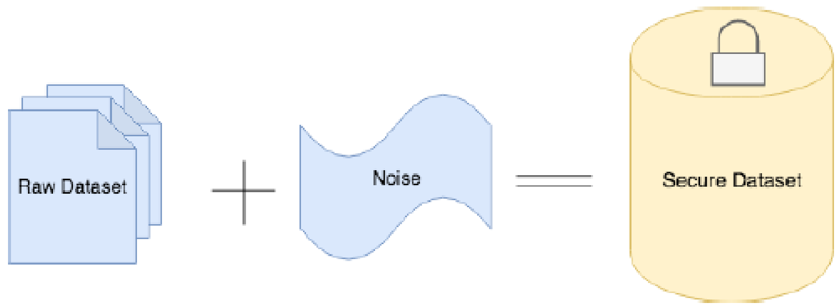
EURECOM
https://www.eurecom.fr

Federated Learning and Privacy

- FL is a machine learning procedure to train a model where the data is distributed over independent providers (or clients)
  - Contributions from individuals with confidential/ sensitive data poses privacy concerns
  - Requirement for privacy guarantee for individuals
  - Criteria for preserving privacy →differential privacy

Differential Privacy

- DP: the absence or presence of a single database item does not affect the outcome of the analysis.



Definition (Dwork'08)

A randomized function K provides ϵ-differential privacy if for all neighboring data sets D_1 and D_2 and all $S \subseteq Range(K)$

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

What if DP is the attack tool?

- What if adversary’s aware of and tries to benefit from DP to strengthen his/her attack?
  - An adversary is able to modify (add, replace, delete, etc.) the published information
  - Modification is differentially private
- Summary: Conflicting goals of the adversary
  - To maximize the possible damage
  - To Minimize detection.

EURECOM
Sophia Antipolis

Problem Formulation

- On the defender's end: preserve differential privacy.
- Trade-off between
  - the attack (the change in the output applied by the adversary),
  - the privacy parameter ε,
  - the sensitivity of the system$\rightarrow$ the amount of change that any single argument to the system can change its output.
- Research goals
  - Generalize this trade-off to all possible types of queries and changes that can be applied onto the system output by the adversary
  - Determine a threshold for detecting the attacker (alternatively, for the attacker to remain undetected)

References

1. C. Dwork, “Differential Privacy”. Automata, Languages and Programming, pgs. 1-12, 2006
2. J. Giraldo, A.A. Cardenas, M. Kantarcıoglu and J. Katz, “Adversarial Classification Under Differential Privacy”, NDSS 2020
3. C. Dwork, F. McSherry, K. Nissim and A. Smith, “Calibrating Noise to Sensitivity in Private Data Analysis”, TCC 2006
4. P. Cuff and L. Yu, “Differential Privacy as a Mutual Information Constraint”, ACM CCS 2016