One-way Functions

Krikamol Muandet
Empirical Inference Department
Max Planck Institute for Intelligent Systems

November 11, 2014


U-HAUL

One-way Functions

Informally, a function f is a one-way function if

1. The description of f is publicly known and does not require any secret information for its operation.
2. Given x, it is easy to compute $f(x)$.
3. Given y, in the range of f, it is hard to find an x such that $f(x) = y$.

One-way Functions

Informally, a function f is a one-way function if

1. The description of f is publicly known and does not require any secret information for its operation.
2. Given x, it is easy to compute $f(x)$.
3. Given y, in the range of f, it is hard to find an x such that $f(x) = y$.

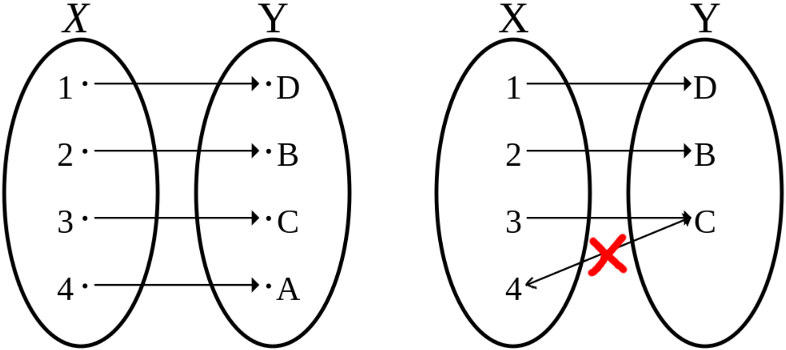
Formal Definition

A function f is *one-way* if f can be computed by a polynomial time algorithm, but for every randomized algorithm A that runs in time polynomial in $n = |x|$, every polynomial $p(n)$, and all sufficiently large n

$$Pr\left[f(A(f(x))) = f(x)\right] < \frac{1}{p(n)}$$

where the probability is over the choice of x from the uniform distribution on $\{0,1\}^n$, and the randomness of A.

One-way Functions



The existence of one-way functions would imply $P \neq NP$.

Examples of Conjectured One-way Functions

One-way functions

1. Factoring problem: $f(p, q) = pq$ for randomly chosen primes p, q.
2. Discrete logarithm problem:

$$f(p, g, x) = \langle p, g, g^x \pmod{p} \rangle$$

   for g a generator of $\mathbb{Z}_p^*$ for some prime p.
3. Discrete root extraction problem.
4. Subset sum problem: $f(a, b) = \langle \sum_i a_i b_i, b \rangle$, for $a_i \in \{0, 1\}$ and n-bit integers b_i.
5. Quadratic residue problem.

Not one-way functions

1. Constant function: $f(x) = 0$.
2. Many-to-one functions (not sufficient to be one-way!).

Privacy in Machine Learning

Two important scenarios:

1. **Interactive** : A “query-response model”
2. **Non-interactive** : Given a dataset $X = (X_1, \ldots, X_n)$, the goal is to produce a sanitized dataset $Z = (Z_1, \ldots, Z_k)$.

The goal is to construct a learning algorithm with a “privacy guarantee”.

Example

We may reveal that smoking correlates to lung cancer, but not that any individual has lung cancer.

Differential Privacy[1]

"Nothing about an individual should be learnable from the database that cannot be learned without access to the database." – Dalenius (1977)

[1]Dwork, C. (2006). Differential privacy. In 33rd International Colloquium on Automata, Languages and Programming. 1–12.

Differential Privacy[1]

"Nothing about an individual should be learnable from the database that cannot be learned without access to the database." – Dalenius (1977)

Definition (Dwork, C. 2006)

A randomized function $\mathcal{K}$ gives ϵ-differential privacy if for all data sets D_1 and D_2 differing on at most one element, and all $S \subseteq \mathrm{Range}(\mathcal{K})$,

$$\frac{\Pr[\mathcal{K}(D_1) \in S]}{\Pr[\mathcal{K}(D_1) \in S]} \leq \exp(\epsilon)\,.$$

[1] Dwork, C. (2006). Differential privacy. In 33rd International Colloquium on Automata, Languages and Programming. 1–12.

Differential Privacy[1]

"Nothing about an individual should be learnable from the database that cannot be learned without access to the database." – Dalenius (1977)

Definition (Dwork, C. 2006)

A randomized function $\mathcal{K}$ gives ϵ-differential privacy if for all data sets D_1 and D_2 differing on at most one element, and all $S \subseteq \mathrm{Range}(\mathcal{K})$,

$$\frac{\Pr[\mathcal{K}(D_1) \in S]}{\Pr[\mathcal{K}(D_1) \in S]} \leq \exp(\epsilon)\,.$$

Differential privacy-preserving mechanism:

1. Data perturbation: $(x_1, y_1), \ldots, (x_n, y_n) \Rightarrow (P_1, y_1), \ldots, (P_n, y_n)$.
2. Perturbing the solution of learning problem.
3. Perturb the optimization problem (Chaudhuri, 2008)
4. Exponential mechanism (McSherry and Talwar, 2007)
5. etc.

[1] Dwork, C. (2006). Differential privacy. In 33rd International Colloquium on Automata, Languages and Programming. 1–12.

Statistical Perspective [2]

- The "query-response" model is considered unrealistic by statisticians.
- Emphasize a role of statistical minimax theory

$$R_n(\mathcal{P}) = \inf_{\hat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P[\ell(\hat{\theta}, \theta)].$$

- Density estimation: $X_1, \ldots, X_n \longrightarrow \hat{p} \longrightarrow \hat{p}^* \longrightarrow Z_1, \ldots, Z_k$.
- Wasserman and Zhou (2010) showed that $\hat{p}^*$ has the same rate of convergence as $\hat{p}$.
- Evaluate "differential privacy" $\Leftrightarrow$ "small ball probabilities"

[2] Wasserman, Larry (2012) "Minimaxity, Statistical Thinking and Differential Privacy," Journal of Privacy and Confidentiality: Vol. 4: Iss. 1, Article 3.

Question?