Information Privacy Tradeoff

Amir Ali Ahmadi and Cole Becker

September 2022

1 The Problem

We consider the following decision problem, we denote Information Privacy Tradeoff Problem (IPT)

Input: Consider a dataset $\mathcal{D} = \{d_i\}_{i \leq m} \subseteq \mathbb{R}^n$ where $d_i \in \{0,1\}^n$, and denote $D \in \mathbb{R}^{m \times n}$ as the matrix containing this data. We can imagine for example that m is the number of people in the dataset, and $d_i \in \mathbb{R}^n$ represents a binary feature vector of information on n features for the ith person. In addition, you are given a non-negative weight vector $w \in \mathbb{R}^n$ representing the relative importance of each the n features, some number $r \in (0,1)$ and an index $k \in \{1, ..., m\}$. Finally we also take in a non-negative number $l \in \mathbb{R}$.

Question: Is there a way to select a subset of features of d_k, indexed by $x \in \{0,1\}^n$ such that $w^T x \geq l$ and the subset features of d_k are shared by at least rm people in the dataset (d_k included).

1.1 Formulation of IPT

Taking all the same inputs from the above problem, let us first define a helper matrix $K \in \mathbb{R}^{m \times n}$ as

$$K_{ij} = \begin{cases} 0 & \text{if } D_{ij} = D_{kj} = (d_k)_j \\ 1 & \text{otherwise} \end{cases}$$

We note that IPT can be reduced to the following non-convex optimization problem

$$\begin{aligned} \mathrm{IPT}_k = \quad & \underset{x}{\text{maximize}} && w^T x \\ & \text{subject to} && ||Kx||_0 < m(1-r) \\ & && x_i \in \{0,1\} \end{aligned} \tag{IPT_k}$$

where the answer to the decision question is given by whether $\mathrm{IPT}_k \geq l$. To see why the constraint in this problem is equivalent to the requirement that the subset x of d_k's features is shared by at least rm people, note that

$$(Kx)_i = \begin{cases} 0 & \text{if } d_i \text{ shares the same } x \text{ features with } d_k \\ > 0 & \text{otherwise} \end{cases}$$

so the number of 0-elements in Kx must be greater than mr or equivalently, the number of non-zero elements must be less than $m(1-r)$

Convex Upper Bound

1.2 Complexity of IPT

Consider the Knapsack Problem (KNAP):

Input:

Question

2 Other questions to consider

1. given an r,w,D, what is $\mathrm{IPT} = \min_k \mathrm{IPT}_k$