# Towards Attack-Resilient Geometric Data Perturbation

Keke Chen [*]    Gordon Sun [†]    Ling Liu [‡]

## Abstract

Data perturbation is a popular technique for privacy-preserving data mining. The major challenge of data perturbation is balancing privacy protection and data quality, which are normally considered as a pair of contradictive factors. We propose that selectively preserving only the task/model specific information in perturbation would improve the balance. Geometric data perturbation, consisting of random rotation perturbation, random translation perturbation, and noise addition, aims at preserving the important geometric properties of a multidimensional dataset, while providing better privacy guarantee for data classification modeling. The preliminary study has shown that random geometric perturbation can well preserve model accuracy for several popular classification models, including kernel methods, linear classifiers, and SVM classifiers, while it also revealed some security concerns to random geometric perturbation. In this paper, we address some potential attacks to random geometric perturbation and design several methods to reduce the threat of these attacks. Experimental study shows that the enhanced geometric perturbation can provide satisfactory privacy guarantee while still well preserving model accuracy for the discussed data classification models.

## 1 Introduction

Data perturbation is one of the most popular approaches to privacy preserving data mining [3, 6, 21]. It is especially useful for data owners to publish data while preserving privacy-sensitive information. Typical examples include publishing micro data for research purpose or outsourcing the data to the third party that provides data mining services. A data perturbation procedure can be simply described as follows. Before the data owner publishes the data, they change the data in certain way to disguise the sensitive information while preserving the particular data property that is critical for building meaningful data mining models. Several perturbation techniques have been proposed for mining purpose recently, among which the most popular one is randomization approach [3, 6] and K-anonymization [21]. Different from the randomization approach that focuses on single-dimensional perturbation and assumes independency between data columns, random rotation perturbation approach [4] and condensation approach [1] try to perturb data while preserving *multidimensional information*.

Perturbation techniques are often evaluated with two basic metrics, the loss of privacy and the loss of information. An ideal data perturbation algorithm aims at minimizing both privacy loss and information loss. However, the two factors are not well-balanced in most existing perturbation techniques [3, 2, 5, 1, 17]. Perturbing data has to lose part of the original information, and there is no perturbation technique that can preserve all information contained in the original dataset. We realized that in order to better balance the two factors we need to focus on preserving the task/model-specific information, e.g., the specific information for data classification models. Bearing this principle in mind, we developed the random rotation perturbation technique [4] aiming at preserving the multidimensional information, such as distance and maniford-based geometric classification boundary. Geometric perturbation, including random rotation perturbation, extends privacy-preserving classification to three important categories of classification models, namely, kernel methods, linear classifiers, and SVM with the popular kernels. These classifiers, if trained with geometrically perturbed data, have similar model accuracy to those trained with the original data.

Effective data perturbation should also consider the potential attacks to the perturbation. The research on randomization technique shows that it is not sufficient to study the naive estimation solely based on the intensity of perturbation, i.e., the difference between the perturbed data and the original. There exist attacks that can utilize the published information of perturbation and the perturbed data to approximately reconstruct the original dataset [16, 10]. In the initial study, we have noticed some potential attacks to random rotation perturbation [4]. It is thus critical to thoroughly study the potential attacks, develop the evaluation methods, and enhance the basic rotation perturbation with additional components, in order to provide better privacy guarantee in terms of attacks.

In this paper, we revisit the basic random rotation perturbation technique, analyze some discovered attacks,

---

[*]College of Computing, Georgia Institute of Technology, kekechen@cc.gatech.edu, Yahoo!, kchen@yahoo-inc.com

[†]Yahoo! gzsun@yahoo-inc.com

[‡]College of Computing, Georgia Institute of Technology lingliu@cc.gatech.edu

and propose a general framework for evaluating the threats and optimizing perturbation in terms of the addressed attacks. Concretely, this paper has three major contributions.

First, we extend the single-column variance-based privacy metric to multidimensional privacy evaluation and develop a generic privacy evaluation model, with which the resilience of perturbation to most potential attacks can be quantitatively evaluated.

Second, some attacks to random rotation perturbation are addressed and analyzed. To systematically study the potential attacks, we categorize the attacks according to three different levels of knowledge the attacker may have about the original dataset. Concretely, naive estimation addresses the attacks that use no information about the original dataset, reconstruction-based attack assumes the attacker knows sufficient information about column distributions, and distance-inference attack is based on a few known original data points and their possible images in the perturbed dataset.

Third, components, such as random translation and noise, are added to the basic rotation perturbation to address the attacks. The generic privacy evaluation model is used to optimize the enhanced geometric perturbation.

The rest of the paper is organized as follows. In section 2, we briefly review the basic random rotation perturbation and its related issues. In section 3, the generic multidimensional privacy evaluation model is defined, which works as the major tool for quantitatively studying the threat of most attacks to a given perturbation. In section 4, some attacks to rotation perturbation are addressed and analyzed in detail, and the additional components are appended to improve the privacy guarantee, which forms an enhanced geometric data perturbation. In section 5, we present the experimental results, showing how effective geometric data perturbation can be in preserving both privacy and model accuracy for several data classifiers.

## 2  Preliminary

Before discussing the attacks to geometric perturbation, we would like to briefly review the previous work on rotation perturbation [4]. The basic perturbation can be defined as follows. Let the original dataset with $d$ columns and $N$ records represented as $X_{d \times N}$ for mathematical convenience. A rotation perturbation is defined by $G(X) = RX$, where $R_{d \times d}$ is a randomly generated $d \times d$ orthogonal matrix [7], for which we use the intuitive name "rotation perturbation" instead. We also use $X_i$ to denote the column $i$ of dataset $X$. Rotation perturbation preserves some important geometric properties of dataset, such as distance, inner product, and any multidimensional geometric surface or manifolds. In the following discussion, we use capitals to represent matrices, lower case characters to represent scalar variables, and bold lower cases to represent vectors.

Why is random rotation perturbation so useful to privacy-preserving data classification? There are two important features of random rotation perturbation. First of all, if we want to preserve the task/model-specific information in data perturbation in order to achieve better balance between data quality and privacy guarantee, random rotation will be a good candidate for preserving the important geometric properties that are critical to many popular classifiers. We proved that three categories of classifiers, namely, kernel methods, linear classifiers, and SVM classifiers with the popular kernels, are "invariant" to rotation perturbation − a rotation-invariant classifier, if trained and tested with rotation perturbed data, will have similar model accuracy to that trained and tested with the original data.

Second, rotation perturbation is safe enough if no information about the original dataset is known. With rotation perturbation, the attacker cannot estimate the original data solely from the perturbed data without any additional knowledge about the original dataset. This makes rotation perturbation perfect for outsourcing secure data sources for privacy-preserving data classification modeling, where no information about the original data source can possibly be obtained by attackers.

However, data sources involved in privacy-preserving data mining often have some (or all) data columns with well-known statistical properties. For example, a column "Age" could have Gaussian distribution with approximately predictable maximum and minimum values, and some of its values might also have strong correlation with some disease symptoms. Such information could be obtained from other similar data sources, such as k-anonymized version of the original data [21]. Furthermore, some particular points, e.g., outliers, could be distinguished in the original dataset [5], and their mapping images can be detected in the perturbed dataset with high probability, which can help to infer the perturbation matrix $R$. We have addressed some potential attacks including the ICA-based data reconstruction [4]. However, without a clear categorization of attackers' knowledge about the original data, it would be inefficient in discussing the methods countering the various attacks. In the following sections, we will discuss the potential attacks to rotation perturbation in detail, in terms of different levels of knowledge attackers may have. Then, with certain security assumption the data owner can get for the published data, s/he can decide to employ certain level of perturbation optimization or be advised not to use ge-

ometric perturbation. We start with multidimensional privacy evaluation, which will be the basic tool in the following analysis.

## 3 Multidimensional Privacy Evaluation

Before the concrete analysis of the potential attacks, we should define an evaluation model to quantitatively evaluate the effectiveness of the attacks. Since attackers try to reduce the privacy guarantee of a specific perturbation, it would be ideal to design a privacy evaluation model that can conveniently incorporate any attack evaluation. We aim at designing such a model in this section.

Unlike the popular randomization methods, where multiple columns are perturbed separately, random geometric perturbation needs to perturb *all* columns together. Therefore, the privacy quality of all columns is correlated under one single perturbation and should be evaluated under a unified metric. Our approach to evaluating the privacy quality of random rotation perturbation consists of two steps: first, we define a unified general-purpose privacy metric that is effective for any multidimensional perturbation technique. Second, we present the methodology of using the privacy evaluation model to evaluate potential attacks for geometric data perturbation.

**Conceptual Multidimensional Privacy Evaluation Model** Since in practice different columns(attributes) may have different privacy concern, we consider that the general-purpose privacy metric $\Phi$ for entire dataset is based on **column privacy metric**. An abstract privacy model is defined as follows. Let **p** be the column privacy metric vector $\mathbf{p} = [p_1, p_2, \ldots, p_d]$, and there are **privacy weights** associated to the $d$ columns, respectively, denoted as $\mathbf{w} = (w_1, w_2, \ldots, w_d)$. $\Phi = \Phi(\mathbf{p}, \mathbf{w})$ uses the two vectors to define the privacy guarantee. In summary, the design of specific privacy model should determine the three factors **p**, **w**, and the function $\Phi$.

We will leave the discussion about one concrete design of **p** later, and define the other two factors first. The first design idea is to take the column importance into unification of different column privacy. Intuitively, the more important the column is, the higher level of privacy guarantee will be required for the perturbed data column. Since **w** is used to denote the importance of columns in terms of preserving privacy, we use $p_i/w_i$ to represent the *weighted column privacy* of column $i$.

The second intuition is the concept of *minimum privacy guarantee* and *average privacy guarantee* among all columns. Normally, when we measure the privacy quality of a multidimensional perturbation, we need to pay more attention to the column that has the lowest weighted column privacy, because such a column could become the breaking point of privacy. Hence, we design the first composition function $\Phi_1 = \min_{i=1}^{d}\{p_i/w_i\}$ and call it the minimum privacy guarantee. Similarly, the *average privacy guarantee* of the multi-column perturbation, defined by $\Phi_2 = \frac{1}{d}\sum_{i=1}^{d} p_i/w_i$, could be another interesting measure.

**Variance-based Unified Column Privacy Metric** Intuitively, for a data perturbation approach, the quality of preserved privacy can be understood as the difficulty level of estimating the original data from the perturbed data. Therefore, how different the *estimated data* is from the original data could be an intuitive measure. We use a variance-of-difference (VoD) based approach, which is derived from the naive variance-based evaluation [3] with more general setting.

Let the difference between the original column data and the estimated data be a random variable $\mathbf{D}_i$. Without any knowledge about the original data, the mean and variance of the difference present the quality of the estimation. Since the mean of difference can be easily removed if the attacker can estimate the original distribution of column, we use only the variance of the difference (VoD) as the primary metric to determine the level of difficulty in estimating the original data.

$VoD$ is formally defined as follows. Let $\mathbf{X}_i$ be a random variable representing the column $i$, $\mathbf{X}'_i$ be the estimated result of $\mathbf{X}_i$, and $\mathbf{D}_i$ be $\mathbf{D}_i = \mathbf{X}' - \mathbf{X}$. Let $E[\mathbf{D}_i]$ and $Var(\mathbf{D}_i)$ denote the mean and the variance of $\mathbf{D}$ respectively. Then $VoD$ for column $i$ is $Var(\mathbf{D}_i)$. Let an estimation of certain value, say $x_i$, be $x'_i$ in $\mathbf{X}'_i$, $\sigma = \sqrt{Var(\mathbf{D}_i)}$, and $c$ denote confidence parameter depending on the distribution of $\mathbf{D}_i$ and the corresponding confidence level. The corresponding original value $x_i$ in $\mathbf{X}_i$ is located in the range defined below:

$$[x'_i - E[\mathbf{D}_i] - c\sigma, x'_i - E[\mathbf{D}_i] + c\sigma]$$

Without considering $E[\mathbf{D}_i]$, the width of the estimation range, $2c\sigma$, presents the difficulty of guessing the original value, which proportionally reflects the level of privacy guarantee. For simplicity, we often use only $VoD$ or $\sigma$ to represent the privacy level.

$VoD$ only defines the privacy guarantee for single column. As we have discussed, we need to evaluate the privacy of all perturbed columns together. The single-column $VoD$ does not work across different columns since different column value ranges may result in very different $VoD$s. Therefore, the same amount of VoD is not equally effective for columns with different value ranges. One straightforward method to unify the different value ranges is via *normalization* over the original dataset and the perturbed dataset. Normalization can be done with max/min normalization or standardized normalization [20]. We use max/min normalization in this paper.

**Incorporating Attack Evaluation:** Since the variance-based model evaluates the accuracy of "estimated" values, it is convenient to incorporate attack evaluation into privacy evaluation. In general, let $X$ be the normalized original dataset, $P$ be the perturbed dataset, and $O$ be the estimated/observed dataset. We calculate $VoD(\mathbf{X}_i, \mathbf{O}_i)$ for the column $i$ in terms of different attacks. Here, we summarize the evaluation of the inference attacks to rotation perturbation [4] that will be further described in detail in the later sections.

1. Naive Estimation: $O = P$;

2. ICA-based Reconstruction: Independent Component Analysis (ICA) is used to estimate $R$. Let $\hat{R}$ be the estimate of $R$, and align the estimated data $\hat{R}^{-1}P$ with the known column distributions and statistics to get the dataset $O$;

3. Distance-based Inference: knowing a set of special points in $X$ that can be mapped to certain set of points in $P$, so that the mapping helps to get the estimated rotation $\hat{R}$, and then $O = \hat{R}^{-1}P$.

## 4 Analysis of Some Attacks and Optimization of Geometric Perturbation

The higher the inference level is, the more knowledge about the original dataset the attacker needs and thus the more complicated the attack might be. In the following sections, we analyze each inference attack, quantify the effectiveness of the attack with the generic privacy evaluation model, and extend the relatively weak rotation perturbation to the full version of geometric perturbation that is more resilient to the discussed attacks.

The complete version of geometric perturbation to the normalized dataset $X$ is defined as $G(X) = RX + \Psi + \Delta$, where $\Psi$ is a random *translation matrix* and $\Delta$ is a noise matrix.

DEFINITION 1. *Let* $\mathbf{t}$ *be a random vector* $\mathbf{t} = [t_1, t_2, \ldots, t_d]^t$, $0 \le t_i \le 1$, *and* $\mathbf{1} = [1, 1, \ldots, 1]^t$. *A translation matrix* $\Psi$ *is defined by* $\Psi = [\mathbf{t}, \mathbf{t}, \ldots, \mathbf{t}]$, *i.e.,* $\Psi = \mathbf{t}\mathbf{1}^t$.

$\Delta = [\delta_1, \delta_2, \ldots, \delta_N]$, where $\delta_i$ is $d$-dimensional Gaussian random vector, each element of which follows the same distribution but is generated independently (i.e., the i.i.d. noise). In this paper, we use Gaussian noise $N(0, \sigma^2)$.

The additional components $\Psi$ and $\Delta$ are used to address the weakness of rotation perturbation, while still preserving the data quality for classification modeling. Concretely, the random translation matrix addresses the attack to rotation center, and the noise addition addresses the distance-inference attack. We also design

an iterative randomized method to maximize the resilience to naive estimation and ICA-based reconstruction. Below we analyze the attacks to the components of geometric perturbation and also discuss the trade-offs between privacy guarantee and data quality (model accuracy).

### 4.1 Naive Estimation to Rotation Perturbation
With the $VoD$ metric over the normalized data, we can formally analyze the privacy guarantee provided by the rotation perturbed data, if no additional information is known by the attacker. Let $X$ be the normalized dataset, $X'$ be the rotation of $X$, and $I_d$ be the $d$-dimensional identity matrix. Thus, VoD of column $i$ can be evaluated by

$$\begin{aligned} Cov(\mathbf{X}' - \mathbf{X})_{(i,i)} &= Cov(R\mathbf{X} - \mathbf{X})_{(i,i)} \qquad (4.1) \\ &= ((R - I_d)Cov(\mathbf{X})(R - I_d)^T)_{(i,i)} \end{aligned}$$

Let $r_{ij}$ represent the element $(i,j)$ in the matrix $R$, and $c_{ij}$ be the element $(i,j)$ in the covariance matrix of $\mathbf{X}$. The VoD for $i$th column is computed as follows.

$$Cov(\mathbf{X}' - \mathbf{X})_{(i,i)} = \sum_{j=1}^{d}\sum_{k=1}^{d} r_{ij}r_{ik}c_{kj} - 2\sum_{j=1}^{d} r_{ij}c_{ij} + c_{ii} \tag{4.2}$$

When the random rotation matrix is generated following the Haar distribution, a considerable number of matrix entries are approximately independent normal distribution $N(0, 1/d)$ [14]. The full discussion about the numerical characteristics of random rotation matrix will be out of the scope of this work. For simplicity and easy understanding, we assume that all entries in random rotation matrix approximately follow independent normal distribution $N(0, 1/d)$. Therefore, random rotations will make $VoD_i$ changing around the mean value $c_{ii}$ as shown in the following equation.

$$E[VoD_i] \sim \sum_{j=1}^{d}\sum_{k=1}^{d} E[r_{ij}]E[r_{ik}]c_{kj} - 2\sum_{j=1}^{d} E[r_{ij}]c_{ij} + c_{ii} = c_{ii}$$

It means that the original column variance could substantially influence the result of random rotation. However, the expectation of VoDs is not the only factor determining the final privacy guarantee. We should also look at the variance of VoDs. If the variance of $VoD$s is considerably large, we still get great chance to find a rotation with high VoDs in a set of sample random rotations, and the larger the $Var(VoD_i)$ is, the more likely the randomly generated rotation matrices can provide a high privacy level. With the approximately independency assumption, we have

$$Var(VoD_i) \sim \sum_{i=1}^{d}\sum_{j=1}^{d} Var(r_{ij})Var(r_{ik})c_{ij}^2$$

$$+4\sum_{j=1}^{d}Var(r_{ij})c_{ij}^2$$

$$\sim\quad O(1/d^2\sum_{i=1}^{d}\sum_{j=1}^{d}c_{ij}^2+4/d\sum_{j=1}^{d}c_{ij}^2).$$

The above result shows that $Var(VoD_i)$ seems approximately related to the average of the squared covariance entries, with more influence from the row $i$ of covariance matrix. Therefore, by looking at the covariance matrix of the original dataset and estimate the $Var(VoD_i)$, we can estimate the chance of finding a random rotation that can give high privacy guarantee.

In Equation 4.2, we also notice that the $i$-th row vector of rotation matrix, i.e., the values $r_{i*}$, plays a dominating role in calculating $VoD_i$. Since swapping rows of a rotation matrix will result in another rotation matrix, it is possible to simply swap the rows of $R$ to locally improve the privacy guarantee. This drives us to propose the row-swapping based local optimization method for finding a better rotation from a given rotation matrix, which greatly reduces the computational cost in randomized search. We define the method as follows. Let $\{(1),(2),\ldots,(d)\}$ be a permutation of the sequence $\{1,2,\ldots,d\}$. The goal is to find a permutation of rows that maximizes the minimum (or average) privacy guarantee.

$$argmax_{\{(1),(2),\ldots,(d)\}}\{min_{1\le i\le d}\{$$

$$(\sum_{j=1}^{d}\sum_{k=1}^{d}r_{(i)j}r_{(i)k}c_{kj}-2\sum_{j=1}^{d}r_{(i)j}c_{ij}+c_{ii})/w_i\}\}\quad(4.3)$$

**4.2  ICA-based Attack** Naive estimation is the basic attack trying to find the original value directly from the perturbed data, which will be ineffective to carefully perturbed data. In this section, we introduce a high-level attack based on data reconstruction. The basic method trying to reconstruct $X$ from the perturbed data $RX$ would be Independent Component Analysis (ICA) technique derived from signal processing [12].

The ICA model can be applied to estimate the independent components (the row vectors) of the original dataset $X$, from the perturbed data, if the following conditions are satisfied:

1. The source row vectors are independent;

2. All source row vectors should be non-Gaussian with possible exception of one row;

3. The number of observed row vectors must be at least as large as the independent source row vectors.

4. The transformation matrix $R$ must be of full column rank.

For rotation matrices, the 3rd and 4th conditions are always satisfied. However, the first two conditions although practical for signal processing, are often not satisfied in data classification. Furthermore, there are a few more difficulties in applying the above ICA-based attack. First of all, even ICA can be done successfully, the order of the original independent components cannot be preserved or determined through ICA [12]. Formally, any permutation matrix $P$ and its inverse $P^{-1}$ can be substituted in the model to give $X'=RP^{-1}PX$. ICA could possibly give the estimate for some permutated source $PX$. Thus, we cannot identify the particular column if the original column distributions are unknown. Second, even if the ordering of columns can be identified, ICA reconstruction does not guarantee to preserve the variance of the original signal − the estimated signal is often scaled up but we do not know how much the scaling is unless we know the original value range of the column. Therefore, without knowing the basic statistics of original columns, ICA-attack is not effective.

However, as we have mentioned earlier, such column statistics are not impossible to get in similar datasets provided for privacy-preserving data mining. We assume the attackers know the basic statistics, including the max/min values and the probability density function (PDF), or empirical PDF of each column. The enhanced ICA-based attack can be described as follows.

1. Run ICA algorithm to get a reconstructed dataset;

2. For each reconstructed column $\mathbf{O}_i$ and each original column $\mathbf{X}_j$, we scale $\mathbf{O}_i$ with the max/min values of $\mathbf{X}_j$, and compare the PDFs of the scaled $\mathbf{O}_i$ and $\mathbf{X}_j$ to find the closest match;

The important step is "PDF Alignment" to find the match between original columns and the perturbed columns. A straightforward method is to calculate the difference between the two PDF functions. Let $f(x)$ and $g(x)$ be the original PDF and the PDF of the reconstructed column, respectively. A typical method to define the difference of PDFs employs the following function.

$$\Delta PDF=\int|f(x)-g(x)|dx\qquad(4.4)$$

In practice, for easy manipulation we discretize the PDF function into bins. It is then equivalent to use the discretized version: $\sum_{i=1}^{n}|f(b_i)-g(b_i)|$, where $b_i$ is the discretized bin $i$. However, the evaluation is not accurate if the values in the two columns are not in the same range as shown in Figure 1. Hence, the reconstructed PDF needs to be translated and scaled to match the range, which requires the maximum/minimum values of the original column to be known, too.
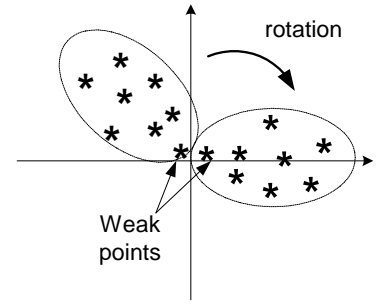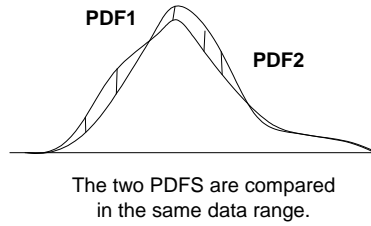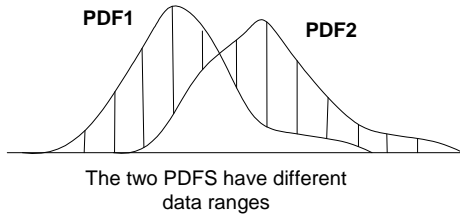
Figure 1: Comparing PDFs in different ranges results in large error. (The lined areas are calculated as the difference between the PDFs.)



Figure 2: Weak points around the default rotation center.

The above procedure describes how to use ICA and additional knowledge about the original dataset to precisely reconstruct the original dataset. Note if the four conditions for effective ICA are exactly satisfied and the basic statistics and PDFs are all known, the basic rotation perturbation will be broken by the enhanced ICA-based attack. In practice, we can test if the first two conditions for effective ICA are satisfied to decide whether we can safely use rotation perturbation. Since the first and second conditions are not satisfied for most datasets in data classification, precise ICA reconstruction cannot be achieved. Under this circumstance, we observed that different rotation perturbations may result in different privacy guarantee and the goal is to find one rotation that is satisfactorily resilient to the enhanced ICA-based attacks. We use the following method to evaluate the resilience against the enhanced ICA-based attacks.

Without loss of generality, we suppose that the level of confidence for an attack is primarily based on the PDF similarity between the two matched columns. Let $O$ be the reconstruction of the original dataset $X$. $\Delta PDF(\mathbf{O}_i, \mathbf{X}_j)$ represents the PDF difference of the column $i$ in $O$ and the column $j$ in $X$. Let $\{(1), (2), \ldots, (d)\}$ be a permutation of the sequence $\{1, 2, \ldots, d\}$, which means a match from the original column $i$ to $(i)$. An optimal match minimizes the sum of PDF differences of all pair of matched columns. We define the minimum privacy guarantee based on the optimal match as follows.

$$p^{min} = min\{\frac{1}{w_k} VoD(\mathbf{X}_k, \mathbf{O}_{(k)}), 1 \le k \le d\} \quad (4.5)$$

where $\{(1), (2), \ldots, (d)\} = argmin_{\{(1),(2),\ldots,(d)\}} \sum_{i=1}^{d} \Delta PDF(\mathbf{X}_i, \mathbf{O}_{(i)})$ Similarly, we can define the average privacy guarantee $p^{avg}$ based on an optimal match between all columns as well.

With the above measures, we are able to estimate how resilient a rotation perturbation is to the ICA-based attacks that incorporate the knowledge of column statistics. We observed in experiments that, although the ICA method may effectively reduce the privacy guarantee for certain rotation perturbations, we can always find some rotation matrices so that they can provide satisfactory privacy guarantee to ICA-based attacks.

**4.3 Attacks to Rotation Center** The basic rotation perturbation uses the origin as the rotation center. Therefore, the points around the origin will be still close to the origin after the perturbation, which leads to weaker privacy protection over these points. The attack to rotation center is another kind of naive estimation. We address this problem with random translation perturbation. The sophisticated attack to the enhanced perturbation would utilize the ICA technique. Therefore, we discuss this problem after we presented the ICA-based attack.

A random translation vector (matrix) has been defined earlier, in Section 4. Concretely, each dimensional value of the random translation vector $\mathbf{t}$ is uniformly drawn from the range $[0, 1]$, so that the center hides in the normalized data space, resilient to estimation. There are two candidates for the extended perturbation.

$$Transformation(1): \quad G(X) = R(X + \Psi) \quad (4.6)$$

or

$$Transformation(2): \quad G(X) = RX + \Psi = R(X + R^{-1}\Psi) \quad (4.7)$$

It is easy to verify that $R^{-1}\Psi$ is also a translation matrix. Thus, the two are equivalent. We will use Transformation (2) in the complete version of geometric perturbation.

The effectiveness of random translation to protecting the rotation center is evaluated by how easy it is to estimate $\Psi$ (or $R^{-1}\Psi$). One approach is again via ICA reconstruction. We assume that attackers know the basic column statistics for effective ICA-based attacks. Since translation just moves the mean of PDF function but preserves the shape of PDF, we can still find the

matches by "PDF Alignment" and get the estimated $R$: $\hat{R}$. Then, an estimation to $\mathbf{t}$ can be done by the following steps.

Take Transformation (1) as example. Let $P$ be the perturbed data. The estimate given by ICA is $\widehat{X + \Psi} = \hat{R}^{-1}P$. Suppose the original column $i$ has the maximum and minimum values $max_i$ and $min_i$, respectively, and $\hat{R}^{-1}P$ has $max_i'$ and $min_i'$, respectively. As the process of ICA shows [12], the reconstruction may scale the original data column with some factor $s$, which can be estimated by $s \approx \frac{max_i' - min_i'}{max_i - min_i}$. Then, the attackers are able to estimate the translation matrix $\Psi$ based on $\hat{R}^{-1}P$. First, the column $i$ of $\hat{R}^{-1}P$ is scaled down to the same span of $X$ by the factor $s$. Then, the translation $t_i$ for column $i$ is estimated by

$$\hat{t}_i \approx min_i' \times s - min_i$$

Apparently, the quality of the estimated $\Psi$ is dependent on the quality of ICA reconstruction. By optimizing the resilience to ICA-based attacks, $\Psi$ will be well protected as well.

**4.3.1  Effect to Model Accuracy** We have shown that random translation can effectively protect the rotation center from attacks. On the other hand, we also need to prove that this additional component will not seriously affect the model accuracy of the three categories of classifiers. Since translation does not change the distance relationship and hyperplane-based class boundary, it is easy to prove that kernel methods, linear classifiers, and SVM classifiers with radial basis function [9] will be invariant to translation.

However, translation does not preserve inner product. Therefore, it would be more complicated to directly prove the classifiers based on inner product, such as the SVM classifiers with polynomial kernels and sigmoid kernels. We will ignore the formal proofs here and show some results in experiments.

**4.4  Distance-inference Attack** In the previous section, we have discussed naive estimation, ICA-based attacks, and attacks to rotation center. In the following discussion, we assume that, besides the information necessary to perform the discussed attacks, the attacker manages to get more knowledge about the original dataset: s/he also knows at least $d+1$ original data records, $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{d+1}\}$. S/he then tries to find the mapping between these points and their images in the perturbed dataset, denoted by $\{\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_{d+1}\}$, to break the rotation and translation perturbation.

With the known points, it is possible to find their images in the perturbed data. Particularly, if a few ($\geq d+1$) original points, such as the "outliers", are known, their images in the perturbed data can

be found with high probability for low-dimensional small datasets ($< 4$ dimensions). With considerable cost, it is not impossible for higher dimensional larger datasets by simple exhaustive search. With the known mapping, the rotation $R$ and translation $\mathbf{t}$ can be precisely calculated if only the geometric perturbation $G(X) = RX + \Psi$ is applied. Therefore, the threat is substantial to the basic geometric perturbation.
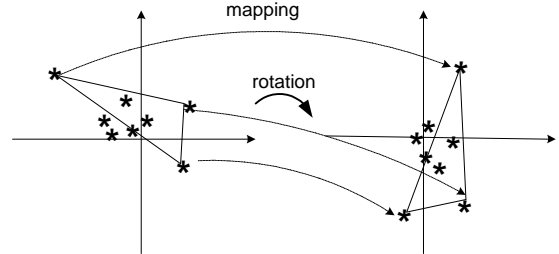


Figure 3: Using known points and distance relationship to infer the rotation matrix.

In order to protect from the distance-inference attack, we introduce an additional noise component $\Delta = [\delta_1, \delta_2, \ldots, \delta_N]$, $\delta_i$ is $d$-dimensional Gaussian random vector, to form the complete version of geometric perturbation, $G(X) = RX + \Psi + \Delta$. Under this perturbation, we analyze how the attacker can estimate the original data with the known points and mappings to decide how intense the noise $\delta_i$ should be.

There are two possible scenarios. In the first scenario, the attacker does not know the exact matching between the known original data records and their images in the perturbed data. The attacker has to figure out the accurate matches with the distance information. However, because the distance relationship has been disturbed, there is low confidence guarantee with plausible matches.

In the second scenario, we assume that the attacker can get (or guess) the right mapping between the original points and their images in the perturbed data: $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{d+1}\} \rightarrow \{\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_{d+1}\}$, where $\mathbf{o}_i$ is perturbed by the noise component, i.e., $\mathbf{o}_i = R\mathbf{x}_i + \mathbf{t} + \delta_i$, and the images are linearly independent. Suppose $\delta_i$ are independently drawn from Gaussian distribution $N(0, \sigma^2)$. We analyze a linear-regression-based attack method for estimating $R$ and deriving $\mathbf{t}$ from the estimated $R$.

**Step 1.** $R$ is estimated as follows. The translation vector $\mathbf{t}$ can be canceled from the perturbation and we get $d$ equations: $\mathbf{o}_i - \mathbf{o}_{d+1} = R(\mathbf{x}_i - \mathbf{x}_{d+1}) + \delta_i - \delta_{d+1}$, $1 \leq i \leq d$. Let $\bar{O} = [\mathbf{o}_1 - \mathbf{o}_{d+1}, \mathbf{o}_2 - \mathbf{o}_{d+1}, \ldots, \mathbf{o}_d - \mathbf{o}_{d+1}]$, $\bar{X} = [\mathbf{x}_1 - \mathbf{x}_{d+1}, \mathbf{x}_2 - \mathbf{x}_{d+1}, \ldots, \mathbf{x}_d - \mathbf{x}_{d+1}]$, and $\bar{\Delta} = [\delta_1 - \delta_{d+1}, \delta_2 - \delta_{d+1}, \ldots, \delta_d - \delta_{d+1}]$. The equations are unified to $\bar{O} = R\bar{X} + \bar{\Delta}$, and estimating $R$ becomes a linear regression problem. Let $\bar{X}^t$ be the transpose of

$\bar{X}$. It follows that the best estimator − the minimum variance unbiased estimator [20, 18] for $R$ is

$$\hat{R} = \bar{O}\bar{X}^t(\bar{X}^t\bar{X})^{-1} \qquad (4.8)$$

.

**Step 2.** With $\hat{R}$, the translation vector $\mathbf{t}$ can also be estimated. Since $\mathbf{o}_i - R\mathbf{x}_i - \delta_i = \mathbf{t}$ and $\delta_i$ has mean value 0, with $\hat{R}$ we have the estimate of $\mathbf{t}$ as

$$\begin{aligned}
\hat{\mathbf{t}} &= \frac{1}{d+1}\{\sum_{i=1}^{d+1}(\mathbf{o}_i - \hat{R}\mathbf{x}_i) - \sum_{i=1}^{d+1}\delta_i\} \\
&\approx \frac{1}{d+1}\sum_{i=1}^{d+1}(\mathbf{o}_i - \hat{R}\mathbf{x}_i)
\end{aligned}$$

. However, $\hat{\mathbf{t}}$ will have considerable variance brought by the components $\hat{R}$ and $\delta_i$.

**Step 3.** With $\hat{R}$ and $\hat{\mathbf{t}}$, the original data $X$ can be estimated. As $O = RX + \Psi + \Delta$, using the estimators $\hat{R}$ and $\hat{\Psi} = [\hat{\mathbf{t}}, \ldots, \hat{\mathbf{t}}]$, we get $\hat{X} = \hat{R}^{-1}(O - \hat{\Psi})$. Due to the variance introduced by $\hat{R}$, $\hat{\Psi}$, and $\Delta$, in practice the attacker may actually need more samples to perform several runs to get several estimated $\hat{X}_i$ , and then uses the mean of $\hat{X}_i$ as the final estimate.

The effective estimation with the above procedure would depend on multiple factors, such as the noise component, and there are strong dependency between $\hat{R}$, $\hat{\Psi}$ and $\hat{X}_i$. Any error in the previous steps can be propagated to the late steps, which makes the noise addition powerful for preventing effective estimation. Furthermore, with the above estimation (attacking) process, we are able to simulate the attack and estimate the actual privacy guarantee to the attack − the unified privacy metric for column $i$ can be calculated with $Var\{\hat{X}_i - X_i\}$.

However, the additional noise component also implies that we have to sacrifice some model accuracy for gaining the stronger privacy protection. We will further study the relationship between the noise level, the privacy guarantee, and the model accuracy in experiments.

**4.5 Other Addressed Attacks** We have studied four kinds of attacks, according to the different levels of knowledge that an attacker may have. The distance-inference attack presents an extreme case that the attacker can know some specific points in the original dataset and their images in the perturbed dataset. AK_ICA [8] investigates a scenario that may also rarely happen in practice. It assumes the attacker can know a significant amount ($\gg d+1$) of the original data points, although the amount is still relatively small compared to the total number of records. These known points might contain significant information, such as the distribution, the covariance matrix of the original dataset.

Therefore, theoretically this information can be used to model the original data. Typical methods, such as Principle Component Analysis (PCA) and ICA, can then be used to reconstruct the original dataset with the approximate information from both the known points and the perturbed data. However, unless the known points can approximately describe the distribution of original dataset, these methods will be not so effective. Furthermore, with the random translation and the additional noise component, the information from the perturbed dataset might be inconsistent with that from the original dataset. As a result, such methods would be ineffective on the full version of geometric perturbation, even though a considerable amount of original points are known. Further studies will be performed on such kind of attacks.

## 5 Randomized Algorithm for Finding Resilient Perturbations

We have analyzed the related inference attacks with the help of multidimensional privacy evaluation model, which allows us to design an algorithm to choose a geometric perturbation resilient to these inference attacks. Considering that a determined algorithm in perturbation optimization may provide extra clue to privacy attackers, we try to randomly optimize the perturbation so that the attacker cannot inference any additional information from the algorithm itself.

Algorithm 1 runs in a given number of iterations, aiming at locally maximizing the minimum privacy guarantee. Initially, a random translation is selected. In each iteration, the algorithm randomly generates a rotation matrix. Local swapping-based optimization of rotation is then applied to find a better rotation matrix against naive estimation, which is then tested by the ICA reconstruction method by the methods defined in Section 4.2. The rotation matrix is accepted as the currently best rotation if it provides higher minimum privacy guarantee than the previous rotations. After the limited number of iterations, finally, the noise component is appended to the perturbation, so that the distance-inference attack cannot reduce the privacy guarantee to a safety level $\phi$, e.g., $\phi = 0.2$. Algorithm 1 outputs the rotation matrix $R_t$, the random translation vector $\mathbf{t}$, the noise level $\sigma^2$, and the minimum privacy guarantee. If the privacy guarantee is lower than an anticipated threshold, the data owner can select not to release the data.

Note that the distance-inference attack is optimized separately. The additional noise component will further reduce the effectiveness of naive estimation and ICA-based attack. Therefore, the actual privacy guarantee will be higher than the evaluated result.
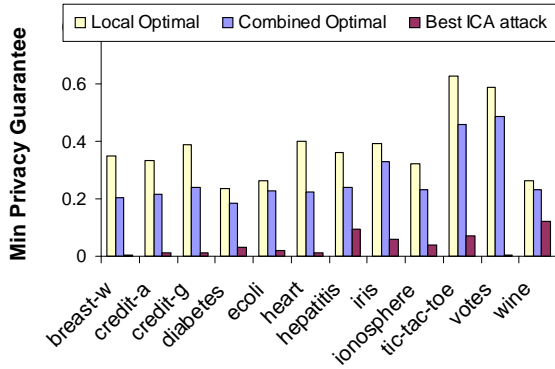
Figure 4: Minimum privacy guarantee generated by local optimization, combined optimization, and the performance of ICA-based attack.
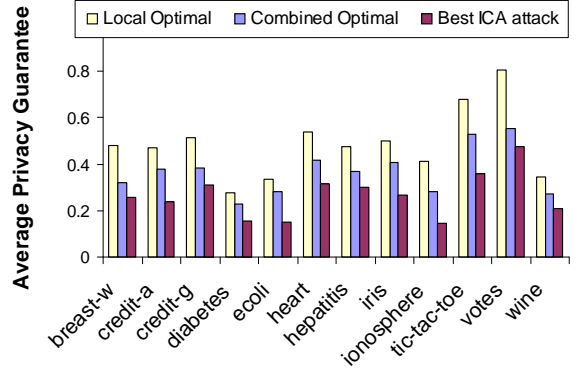


Figure 5: Average privacy guarantee generated by local optimization, combined optimization, and the performance of ICA-based attack.

---

**Algorithm 1** Finding a resilient perturbation ($X_{d \times N}$, **w**, $\phi$, $m$)

**Input**: $X_{d \times N}$:the original dataset, **w**: weights of attributes in privacy evaluation, $\phi$: the expected privacy guarantee in terms of distance-inference attack, $m$: the number of iterations.
**Output**: $R_t$: the selected rotation matrix, $\Psi$: the random translation, $\sigma^2$: the noise level, $p$: privacy quality
calculate the covariance matrix $C$ of $X$;
$p = 0$, and randomly generate the translation $\Psi$;
**for** Each iteration **do**
  randomly generate a rotation matrix $R$;
  swapping the rows of $R$ to get $R_1$, which maximizes $min_{1 \leq i \leq d}\{\frac{1}{w_i}(Cov(R_1 X - X)_{(i,i)}\}$;
  $p_0 =$ the privacy guarantee of $R_1$, $p_1 = 0$;
  **if** $p_0 > p$ **then**
    generate $O$ with ICA;
    scale the columns in $O$ with the maximum/minimum values of original columns;
    $\{(1), (2), \ldots, (d)\} = argmin_{\{(1),(2),\ldots,(d)\}} \sum_{i=1}^{d} \Delta PDF(X_i, O_{(i)})$
    $p_1 = min_{1 \leq k \leq d}\frac{1}{w_k}VoD(X_k, O_{(k)})$
  **end if**
  **if** $p < min(p_0, p_1)$ **then**
    $p = min(p_0, p_1)$, $R_t = R_1$;
  **end if**
**end for**
$p_2 =$ the privacy guarantee to the distance-inference attack with the perturbation $G(X) = R_t X + \Psi + \Delta$.
Tune the noise level $\sigma^2$, so that $p_2 \geq \phi$

---

## 6 Experiments

We design three sets of experiments to evaluate the geometric perturbation approach. The first set shows the optimization of the privacy guarantee in the basic geometric perturbation (without noise addition) in terms of naive estimation and ICA-based attack. In the second set of experiments, we study the effectiveness of translation perturbation to protecting the rotation center, and show that the two kinds of classifiers: SVM with polynomial kernel and sigmoid kernel are also invariant to translation perturbation. The third set of

experiments studies the threat of distance-inference attack and the relationship between the additional noise component of the geometric perturbation, the privacy guarantee, and the model accuracy. All datasets used in the experiments can be found in UCI machine learning database [1]. We also use FastICA package [11] in evaluating the effectiveness of ICA-based attacks.

**6.1 Perturbation Optimization against Naive Estimation and ICA-based Attack** We run the randomized optimization algorithm and show how effective it can generate resilient perturbations. Each column in the experimental dataset is considered equally important in privacy evaluation, thus, the weights are not included in evaluation.

Figure 4 and 5 summarize the evaluation of privacy quality on experimental datasets. The results are obtained in 50 iterations with the optimization algorithm described in Section 5. The "Local Optimal" represents the locally optimized minimum privacy guarantee addressing naive estimation. "Best ICA attack" is the worst perturbation that gives the best ICA attack performance, i.e., getting the lowest privacy guarantee among the 50 perturbations. "Combined Optimal"is the combined optimization result given by Algorithm 1 at the end of 50 iterations. The above values are all standard deviation of the difference between the perturbed dataset (or the estimated dataset) and the original dataset. The "Local Optimal" values can often reach a high level after 50 iterations, which means that the swapping method is very efficient in locally optimizing the privacy quality. The best ICA attacks often result very low privacy guarantee, which means some rotation perturbations are weak to ICA-based attacks. "Combined Optimal" values are much higher than the corresponding ICA-based attacks, which supports our conjecture that we can always find one perturbation

that is sufficiently resilient to ICA-based attacks if the four conditions for perfect ICA reconstruction are not satisfied.

We also show the detail in the course of optimization for two datasets "Diabetes" and "Votes" in Figure 6 and 7, respectively. For both datasets, since the lowest privacy guarantees reduced by ICA-based attacks are lower than the result of swapping-based optimization, the combined optimal result is located in between the curves of best ICA-attacks and the best local optimization result. In the case that ICA-based attacks are not effective, i.e., the "best ICA attack" is higher than local optimization curve, we take the local optimization curve as the combined optimal result.

## 6.2 Effectiveness of Translation Perturbation

In this set of experiments, firstly, we show that it is ineffective to estimate the rotation center if the translation perturbation is appended. As we have mentioned, if the translation vector could be precisely estimated, the rotation center would be exposed. We applied the ICA-based attack to rotation center that is described in Section 4.3. The data in Figure 8 shows $stdev(\hat{\mathbf{t}} - \mathbf{t})$ which is equivalent to the VoD used in multidimensional privacy evaluation model. Compared to the range of the elements in $\mathbf{t} - [0, 1]$, the standard deviations are quite large, so we can conclude that random translation will also be safe to attacks, if we have optimized the resilience of rotation perturbation in terms of ICA-based attacks.

Secondly, we show that the two classifiers, SVM with polynomial kernel, and SVM with sigmoid kernel, are also invariant to translation transformation. Table 1 lists the experimental result on the 12 datasets. We randomly translate each dataset for ten times. The result is the average of the ten runs. For most datasets, the result shows zero or tiny deviation from the standard model accuracy.

Table 1: Experimental result on random translation

| Dataset | SVM(polynomial) | | SVM(sigmoid) | |
|---|---|---|---|---|
| | orig | Tr | orig | Tr |
| breast-w | 96.6 | $0 \pm 0$ | 65.5 | $0 \pm 0$ |
| credit-a | 88.7 | $0 \pm 0$ | 55.5 | $0 \pm 0$ |
| credit-g | 87.3 | $-0.4 \pm 0.4$ | 70 | $0 \pm 0$ |
| diabetes | 78.5 | $0 \pm 0.3$ | 65.1 | $0 \pm 0$ |
| ecoli | 89.9 | $-0.1 \pm 0.5$ | 42.6 | $0 \pm 0$ |
| heart | 91.1 | $-0.2 \pm 0.2$ | 55.6 | $0 \pm 0$ |
| hepatitis | 96.7 | $-0.4 \pm 0.3$ | 79.4 | $0 \pm 0$ |
| ionosphere | 98 | $+0.3 \pm 0$ | 63.5 | $+0.6 \pm 0$ |
| iris | 97.3 | $0 \pm 0$ | 29.3 | $-1.8 \pm 0.4$ |
| tic-tac-toe | 100 | $0 \pm 0$ | 65.3 | $0 \pm 0$ |
| votes | 99.2 | $+0.2 \pm 0.1$ | 65.5 | $-4.7 \pm 0.6$ |
| wine | 100 | $0 \pm 0$ | 39.9 | $0 \pm 0$ |

## 6.3 Tradeoffs in Terms of Distance-inference Attack

We use the geometric perturbation with random noise component : $G(X) = RX + \Psi + \Delta$, to address the potential distance-inference attacks. From the formal analysis, we know that the noise component $\Delta$ can significantly affect the accuracy of distance-inference attack, thus provide certain privacy guarantee. Intuitively, the higher the noise level is, the better the privacy guarantee. However, with the increasing noise level, the model accuracy could be affected, too. In this set of experiments, we first study the relationship between the noise level, represented by the variance $\sigma^2$, and the privacy guarantee, as well as between the noise level and the model accuracy, with three datasets "Diabetes", "Iris", and "Votes". Then, we compare the privacy guarantee and the model accuracy for all of the experimental datasets at certain noise level ($\sigma = 0.1$).

Figure 9 shows, if the attack described in Section 4.4 is addressed with the noise component, the privacy guarantee increases with the increase of noise level. At the noise level $\sigma = 0.1$, the privacy guarantee is almost above 0.2. However, Figure 10 and 11 show the decreasing trend of accuracy for KNN classifier and SVM (RBF kernel) classifier, respectively. With the noise level lower than 0.1, the accuracy of both classifiers is only reduced less than 6%, which is quite acceptable.

We summarize the privacy guarantees at the noise level 0.1 for all experimental datasets [2] in Figure 12, and also the change of model accuracy for KNN, SVM(RBF), and Perceptron in Figure 13. The positive accuracy differences indicate that the perturbation increases the accuracy in some cases. Except the boolean datasets "Votes" and "Tic-tac-toe" are quite sensitive to the noise component, most of the results show that, with small noise addition, we can get satisfactory privacy guarantee with small sacrifice of model accuracy.

## 7 Related Work

Data perturbation changes the data in such a way that it is difficult to estimate the original values from the perturbed data, while information critical to data mining are still preserved. Recently data perturbation techniques have become popular for privacy-preserving data mining [3, 6, 1, 21, 4], due to the relatively low cost to deploy them compared to the cryptographic techniques [19, 22, 23, 15, 13]. However, there are a few challenges in the data-perturbation based privacy-preserving data mining. First, it is commonly recognized that it is critical but difficult to balance the data

---
[2] "Ionosphere" is not included because any combination of known $d$ records results in a singular matrix. Therefore, the attack described in Section 4.4 does not work.
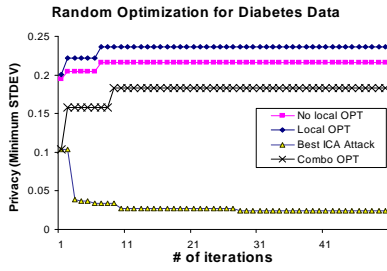
Figure 6: Optimization of perturbation for Diabetes data.
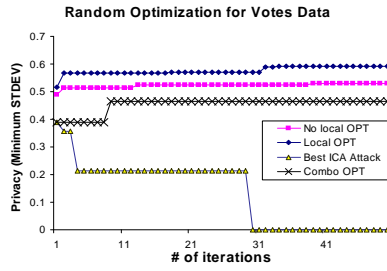


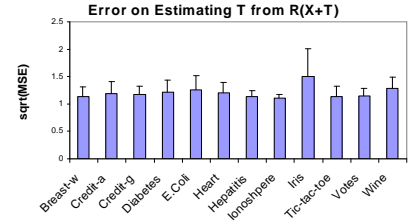Figure 7: Optimization of perturbation for Votes data.



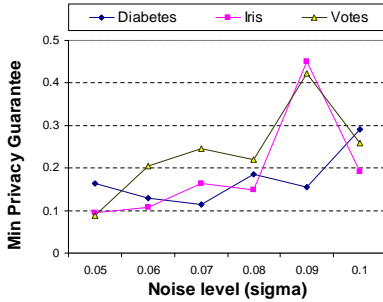Figure 8: Resilience to the attack to random translation



Figure 9: The change of minimum privacy guarantee vs. the increase of noise level for the three datasets.
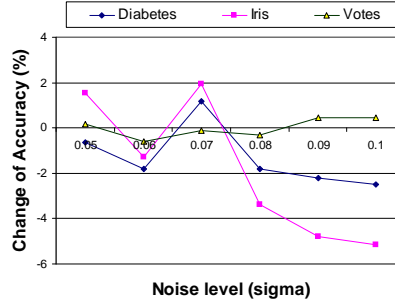


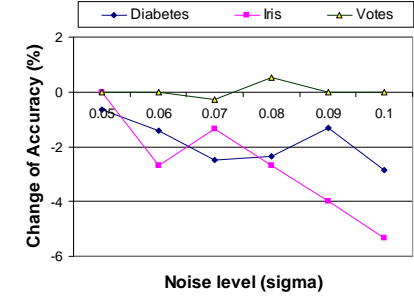Figure 10: The change of accuracy of KNN classifier vs. the increase of noise level.



Figure 11: The change of accuracy of SVM(RBF) classifier vs. the increase of noise level.

quality (affecting the model accuracy) and the data privacy. Second, the potential attacks to the data perturbation methods are not sufficiently investigated. A few works have started to address the privacy breaches to randomization approaches, by applying data reconstruction techniques [10, 16] or the domain knowledge [5]. Third, some approaches, such as randomization approach [3], require to develop new data mining algorithms to mine the perturbed data, which raises extra difficulty in applying the technique. To address these challenges, it is critical to understand the intrinsic relationship between data mining models and the perturbation techniques.

In paper [4], we propose to investigate the perturbation techniques from the perspective of the specific data mining models. We noticed that different data mining tasks/models actually care about different properties of the dataset, which could be statistical information, such as the column distribution and the covariance matrix, geometric properties, such as distance, and so on. Clearly, it is almost impossible to preserve all of the information in the original dataset in data perturbation. Thus, we have to focus on preserving only the task-specific information in the dataset that is critical to the specific data mining task/model, in order to bring better flexibility in optimizing data privacy guarantee. Our initial study on the geometric perturbation approach to data classification [4] has shown that the *task/model-specific data perturbation* can provide

better privacy guarantee and better model accuracy. Furthermore, compared to existing randomization approaches, geometric perturbation does not require to develop new classification algorithms that can utilize the perturbed data to build classification models. We also compared geometric perturbation with condensation approach [1]. The result shows that geometric perturbation can provide much higher privacy guarantee.

## 8 Conclusion

Task/model-oriented perturbation can improve the balance between model accuracy and privacy guarantee. Geometric data perturbation is specifically designed for a bunch of popular data classification models. These classifiers, if trained and tested with the perturbed dataset, can have similar model accuracy compared to those trained and tested with the original dataset. This paper analyzes some potential attacks to geometric perturbation and provides a framework for investigating more attacks and optimizing the perturbation in terms of the attacks. Experimental results show that with a random optimization method, geometric perturbation can provide satisfactory privacy guarantee with little sacrifice of model accuracy, in terms of the discussed attacks.
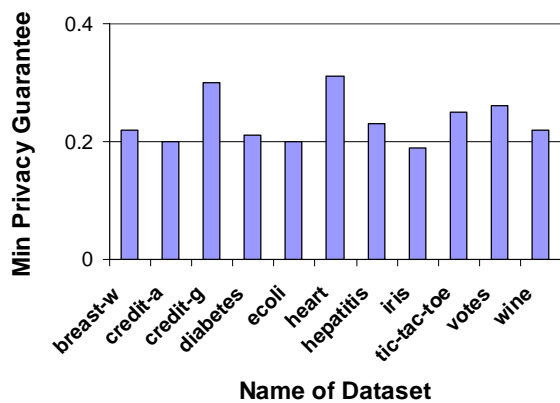
Certainly, there are more potential attacks to be

Figure 12: Minimum privacy guarantee at the noise level $\sigma = 0.1$
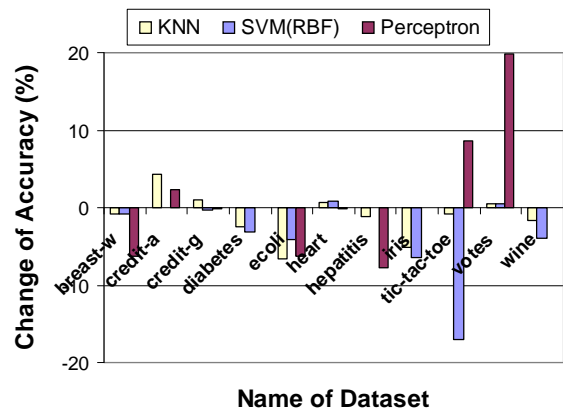


Figure 13: The change of model accuracy at the noise level $\sigma = 0.1$

discovered. We hope that the methodology developed in this paper can be extended to analyze more attacks and to optimize the geometric perturbation as well.

## References

[1] AGGARWAL, C. C., AND YU, P. S. A condensation approach to privacy preserving data mining. *Proc. of Intl. Conf. on Extending Database Technology (EDBT) 2992* (2004), 183–199.

[2] AGRAWAL, D., AND AGGARWAL, C. C. On the design and quantification of privacy preserving data mining algorithms. *Proc. of ACM PODS Conference* (2002).

[3] AGRAWAL, R., AND SRIKANT, R. Privacy-preserving data mining. *Proc. of ACM SIGMOD Conference* (2000).

[4] CHEN, K., AND LIU, L. A random rotation perturbation approach to privacy preserving data classification. *Proc. of Intl. Conf. on Data Mining (ICDM)* (2005).

[5] EVFIMIEVSKI, A., GEHRKE, J., AND SRIKANT, R. Limiting privacy breaches in privacy preserving data mining. *Proc. of ACM PODS Conference* (2003).

[6] EVFIMIEVSKI, A., SRIKANT, R., AGRAWAL, R., AND GEHRKE, J. Privacy preserving mining of association rules. *Proc. of ACM SIGKDD Conference* (2002).

[7] GALLIER, J. *Geometric Methods and Applications for Computer Science and Engineering.* Springer-Verlag, New York, 2000.

[8] GUO, S., AND WU, X. Deriving private information from general linear transformation perturbed data. *CS Technical Report, UNC Charlotte* (March, 2006).

[9] HASTIE, T., TIBSHIRANI, R., AND FRIEDMANN, J. *The Elements of Statistical Learning.* Springer-Verlag, 2001.

[10] HUANG, Z., DU, W., AND CHEN, B. Deriving private information from randomized data. *Proc. of ACM SIGMOD Conference* (2005).

[11] HYVARINEN, A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks 10*, 3 (1999).

[12] HYVARINEN, A., KARHUNEN, J., AND OJA, E. *Independent Component Analysis.* Wiley-Interscience, 2001.

[13] JAGANNATHAN, G., AND WRIGHT, R. N. Privacy-preserving distributed k-means clustering over arbitrarily partitioned data. *Proc. of ACM SIGKDD Conference* (2005).

[14] JIANG, T. How many entries in a typical orthogonal matrix can be approximated by independent normals. *To appear in The Annals of Probability* (2005).

[15] KANTARCIOGLU, K., AND CLIFTON, C. Privacy-preserving distributed mining of association rules on horizontally partitioned data. *IEEE Trans. on Knowledge and Data Eng.* (2004).

[16] KARGUPTA, H., DATTA, S., WANG, Q., AND SIVAKUMAR, K. On the privacy preserving properties of random data perturbation techniques. *Proc. of Intl. Conf. on Data Mining (ICDM)* (2003).

[17] LEFEVRE, K., DEWITT, D. J., AND RAMAKRISHNAN, R. Incognito: Efficient full-domain k-anonymity. *Proc. of ACM SIGMOD Conference* (2005), 49–60.

[18] LEHMANN, E. L., AND CASELLA, G. *Theory of Point Estimation.* Springer-Verlag, 1998.

[19] LINDELL, Y., AND PINKAS, B. Privacy preserving data mining. *Journal of Cryptology 15*, 3 (2000), 177–206.

[20] NETER, J., KUTNER, M. H., NACHTSHEIM, C. J., AND WASSERMAN, W. *Applied Linear Statistical Methods.* WCB/McGraw-Hill, 1996.

[21] SWEENEY, L. k-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10*, 5 (2002).

[22] VAIDYA, J., AND CLIFTON, C. Privacy preserving association rule mining in vertically partitioned data. *Proc. of ACM SIGKDD Conference* (2002).

[23] VAIDYA, J., AND CLIFTON, C. Privacy preserving k-means clustering over vertically partitioned data. *Proc. of ACM SIGKDD Conference* (2003).