Noise Analysis for Linear Regression Reconstruction Attacks

Chhavi Yadav, Tatsuki Koga

March 18, 2021

1 Introduction

1.1 Motivation

"To be left alone is the most precious thing one can ask of the modern world." - Anthony Burgess

The goal of most privacy related research is to protect the privacy of an individual while providing public summary statistics of the dataset. Consider the following setting - a dataset curator wants to release statistics about how a sensitive attribute is related to non-sensitive attributes. An adversary with access to a noisy version of this information can fabricate attacks to reconstruct the sensitive attribute. In this work, we aim to understand the effect of different kinds of noise on the reconstruction by framing it as an optimization problem.

1.2 Previous Work

Reconstruction attacks aim to identify all or almost all individuals' secret data given published information. Such information includes the joint marginal distribution of secret data with other public attributes (contingency tables), or the parameters of machine learning models like logistic regression, linear regression, or support vector machine. A recent comprehensive survey for reconstruction attacks was done by Dwork et al. [2017].
For binary secret data, Dinur and Nissim [2003] first proposed a greedy algorithm for reconstruction attack with a search space of full 2^n possible choices. Dinur and Nissim [2003] also showed that answering $\Omega(n \log n)$ queries with $o(\sqrt{n})$ noise per query is not private. Dwork et al. [2007] reduced the number of queries to $\Omega(n)$ and allowed some queries to be arbitrarily noisy. Dwork and Yekhanin [2008] addressed the computational complexity of the attack and improved it significantly. The attack was extended by Kasiviswanathan et al. [2013] to M-estimators with a differentiable loss function. M-estimators is a general class of estimators that are obtained by minimizing the objective below:

$$\arg\min_{\theta} \sum_{i=1}^{n} \ell(\theta; x_i),$$

where ℓ is the loss function and x_i is i^{th} individual's data. By definition, logistic & linear regression are M-estimators. Specifically, Kasiviswanathan et al. [2013] derived a lower bound on the amount of noise needed to perturb an M-estimator such that the reconstruction attack fails in the small data setting.

1.3 Contributions

Despite Kasiviswanathan et al. [2013]'s lower bound in the small data setting, it is still unclear what kind of noise preserves privacy while maintaining the usefulness of the estimator in real world applications. Also, previous works exclude bias terms from the linear regression model, which may harm the accuracy of the model. We investigate these issues by carrying out the reconstruction attack with randomized linear regression coefficients and biases using real-life medical data. We deploy various kinds of noise to randomize the regressors.
To summarize, our contributions are threefold. We :

1. reformulate the reconstruction attack primal and dual problems to include noisy bias terms from linear regression models,
2. quantitatively evaluate how different noises (Gaussian, Laplacian & Truncated normal) influence the privacy as well as utility, and
3. empirically show that truncated normal distribution achieves the best privacy-utility tradeoff.

1.4 Organization of the Paper

Section 2 talks about the mathematical formulation of the problem in detail. It starts by stating the problem and introducing necessary notation. Sections 2.1, 2.2 & 2.3 lay down the precise primal, dual formulations and KKT conditions respectively. Section 3 briefly mentions the approaches we used. Sections 4.1 and 4.2 discuss the dataset and experimental setup. Closely following is section 4.3 which scrutinizes the results in detail. Lastly, section 5 provides a summary of our report & suggests some potential directions for the exploratory reader.

2 Statement of the Problem

Consider a database D of n individuals. Each individual $i \in [n]$ has public data $x_i \in \mathbb{R}^d$ and a secret attribute $s_i \in \mathbb{R}$ associated with itself. $D = (\mathbf{X} \mid \mathbf{s})$ where $\mathbf{X} \in \mathbb{R}^{n \times d}$ and $\mathbf{s} \in \mathbb{R}^{n \times 1}$. $\mathbf{X}_{(k)} \in \mathbb{R}^n$ corresponds to the k^{th} column of $\mathbf{X}$. For each $j \in [d]$, the linear regression model coefficient $\theta_j \in \mathbb{R}$ and intercept $b_j \in \mathbb{R}$ are obtained, which predict the value of s_i by $\hat{s}_i = \theta_j x_{ij} + b_j$. This is equivalent to $\hat{s}_i = \begin{pmatrix} \theta_j \\ b_j \end{pmatrix} \cdot \begin{pmatrix} x_{ij} \\ 1 \end{pmatrix}$. In matrix form, the linear regression predicts $\hat{\mathbf{s}} \in \mathbb{R}^n$ by $\hat{\mathbf{s}} = \begin{pmatrix} \mathbf{X}_{(j)} & \mathbf{1} \end{pmatrix} \begin{pmatrix} \theta_j \\ b_j \end{pmatrix}$. We further define $\mathbf{X}'_{(j)} = \begin{pmatrix} \mathbf{X}_{(j)} & \mathbf{1} \end{pmatrix} \in \mathbb{R}^{n \times 2}$
Before publishing $\theta = (\theta_1, \ldots, \theta_d)$ and $\mathbf{b} = (b_1, \ldots, b_d)$, the database curator adds noise to θ and $\mathbf{b}$ to prevent adversaries from attacking the secret attribute, e.g., publishes $\tilde{\theta}_j = \theta_j + z_j$ for all $j \in [d]$, where z_j is i.i.d. random noise according to some distribution (e.g., Gaussian, Laplacian). The database curator makes public noisy coefficients $\tilde{\theta} = (\tilde{\theta}_1, \ldots, \tilde{\theta}_d)$ and biases $\tilde{\mathbf{b}} = (\tilde{b}_1, \ldots, \tilde{b}_d)$.
A simple reconstruction attack is the optimization problem stated below. Our goal is to identify the effect of different kinds of noisy perturbations on the reconstructed s.

2.1 Primal

For each feature $j \in [d]$, the linear regression tries to solve $\min_{\theta_j, b_j} \|\mathbf{s} - \mathbf{X}'_{(j)} \begin{pmatrix} \theta_j \\ b_j \end{pmatrix} \|_2^2$. The solution satisfies the maximum likelihood estimation equation:

$$\mathbf{X}'^{\top}_{(j)} \mathbf{s} - \mathbf{X}'^{\top}_{(j)} \mathbf{X}'_{(j)} \begin{pmatrix} \theta_j \\ b_j \end{pmatrix} = 0.$$

This gives rise to the optimization problem for the reconstruction attack as stated below:

$$\min_{\mathbf{s}} \quad \left\| \text{vert}\left(\mathbf{X}'^{\top}_{(1)}, \ldots, \mathbf{X}'^{\top}_{(d)}\right) \mathbf{s} - \text{vert}\left(\mathbf{X}'^{\top}_{(1)} \mathbf{X}'_{(1)} \begin{pmatrix} \tilde{\theta}_1 \\ \tilde{b}_1 \end{pmatrix} \tilde{\theta}_1, \ldots, \mathbf{X}'^{\top}_{(d)} \mathbf{X}'_{(d)} \begin{pmatrix} \tilde{\theta}_d \\ \tilde{b}_d \end{pmatrix}\right) \right\|_2, \tag{1}$$

where $\text{vert}(\cdot, \ldots, \cdot)$ denotes vertical concatenation of the argument matrices.
The problem in (1) is equivalent to the optimization problem below and we see it as the primal problem:

$$\begin{aligned} \min \quad & \|\mathbf{y}\|_2 \\ \text{s.t.} \quad & \mathbf{y} = \mathbf{A}\mathbf{s} - \mathbf{z}, \end{aligned} \tag{2}$$

where $\mathbf{A} = \text{vert}\left(\mathbf{X}'^{\top}_{(1)}, \ldots, \mathbf{X}'^{\top}_{(d)}\right)$ and $\mathbf{z} = \text{vert}\left(\mathbf{X}'^{\top}_{(1)} \mathbf{X}'_{(1)} \begin{pmatrix} \tilde{\theta}_1 \\ \tilde{b}_1 \end{pmatrix} \tilde{\theta}_1, \ldots, \mathbf{X}'^{\top}_{(d)} \mathbf{X}'_{(d)} \begin{pmatrix} \tilde{\theta}_d \\ \tilde{b}_d \end{pmatrix}\right)$.

2.2 Dual

The Lagrange dual problem is the following:

$$\begin{aligned} \max \quad & \mathbf{z}^{\top} \nu \\ \text{s.t.} \quad & \|\nu\|_2 \leq 1 \\ & \mathbf{A}^{\top} \nu = 0. \end{aligned} \tag{3}$$

We use the fact that the conjugate function of ℓ_2-norm is the indicator function of the ℓ_2-norm unit ball.

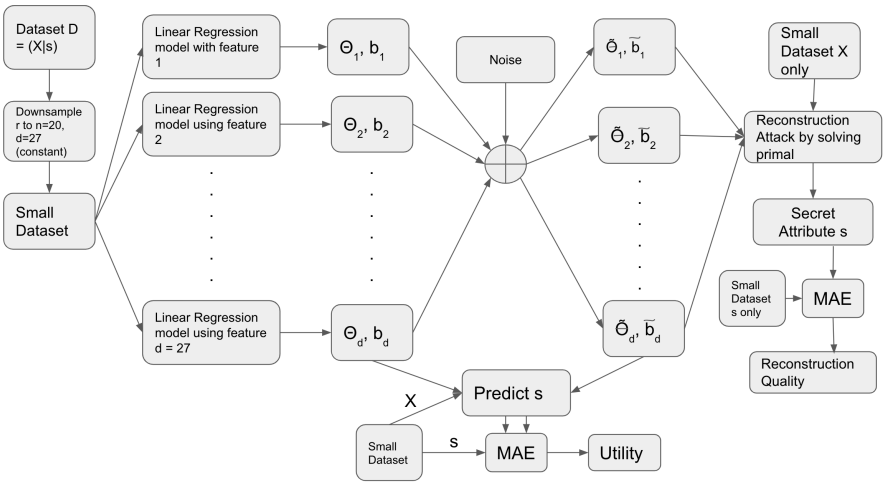

Figure 1: Flowchart showing the complete methodology used in our report.

2.3 KKT Conditions

Since the primal problem 2 does not have inequality constraints, the KKT conditions are as below:

$$\mathbf{y} - \mathbf{A}\mathbf{s} + \mathbf{z} = 0 \tag{4}$$

$$\mathbf{A}^\top \nu = 0 \tag{5}$$

$$\frac{\mathbf{y}}{\|\mathbf{y}\|_2} + \nu = 0. \tag{6}$$

3 Approaches

Our approach for the reconstruction attack is to solve the primal optimization problem in (2). It can be solved easily with open-source optimization packages. Solving the problem is also known to be computationally efficient. A possible alternative is to carry out an exhaustive search, but it is nearly impossible in terms of the computational complexity since $\mathbf{s}$ is real-valued. Our end-to-end methodology is illustrated in flowchart 1.

4 Experiments & Results

4.1 Dataset

We use the Cancer Mortality Rate Dataset[1] for our investigation. This dataset comprises of 32 independent variables. We remove any features with missing values and textual data leading to $d = 27$ independent variables. These are public. A complete list is available in Table 1. The dependent variable to be predicted is the death rate of a county. It is also the secret attribute in our case. We standardize each feature separately as a preprocessing step. The total number of samples in the dataset, n is 3047.

4.2 Experiments

Figure 1 demonstrates our end-to-end pipeline. In short, we first fit 27 linear regression models, one for each feature, to predict the death rate. Noisy versions of these model parameters are made public. An adversary uses these noisy parameters & the noiseless public data features to predict the secret attribute, death rate. We plug-in different noise models, change their parameters to observe the effect on reconstruction quality and utility.

The aforementioned attack provides guaranteed good reconstruction only for small-data regime where $d \approx n$. Instead of directly picking the first d samples from the dataset, we take a more principled approach. We analyze how the prediction error of noiseless regressors changes with dataset size and select the size with the lowest mean absolute error (MAE) for further experimentation (Fig. 2). Dataset is downsampled randomly and experiments are repeated 50 times for each size.

[1] https://data.world/exercises/linear-regression-exercise-1/workspace/file?filename=cancer_reg.csv

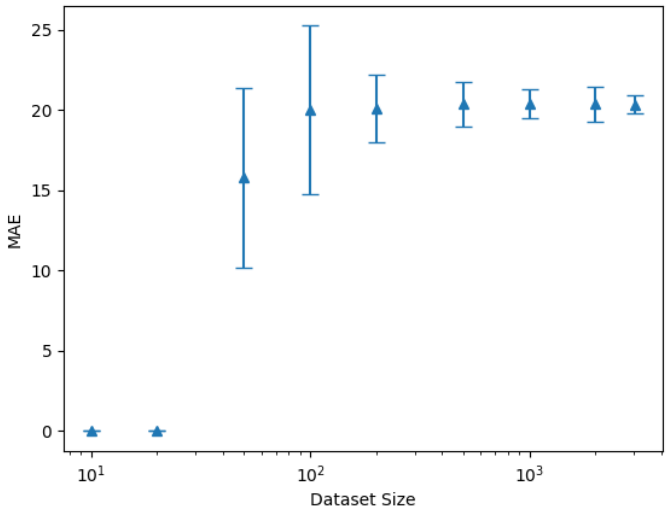

Figure 2: Plot showing the effect of sample size on error. We downsample the dataset to the lowest MAE size for our small-data setting reconstruction attack.

After finding the optimal dataset size, we fit regressors on randomly downsampled datasets of this size (10 times) and publish their corrupted versions. We then reconstruct the death rate whilst acting as an adversary by solving the primal problem in (2). MAE is again used to assess difference between reconstructed and original death rates.

To investigate the effect of noise on reconstruction, we consider 3 types of noise distributions commonly used in privacy literature like [Dwork et al., 2006, McSherry and Talwar, 2007, Dwork and Roth, 2014] : Gaussian, Laplacian & Truncated-Normal (TN). All of them are centered at 0. We vary the σ and *scale* parameters logarithmically for gaussian, laplacian and TN noises respectively. This translates to different levels of corruption in published regressors. We set the interval of TN as $(-0.05, 0.05)$.

4.3 Results

For all noise models, reconstruction of the secret variable is almost perfect for noise levels $\leqslant 0.001$. The quality degrades as variance of the noise distributions increases. With increasing variance, the noise distributions tend to a dirac delta function which has all its mass concentrated on a single point (0 in our case) with high probability and hence effectively no randomization. The degradation, as measured by MAE between reconstructed and original death rate, is exponential in nature.

Degradation in reconstruction quality for gaussian and laplacian noise models is almost 50 times higher as compared to the TN noise models. This suggests that for the same perturbation budget, one should use gaussian or laplacian noise model to achieve higher privacy.

Of major importance in privacy research is the privacy-utility tradeoff. The question at hand is what level of noise is enough to sufficiently corrupt the regressors (which makes it hard for an adversary to reconstruct the secret attribute) while still leaving them useful enough to achieve a good performance on the original prediction task (utility). We measure this using MAE ratios between noisy and noiseless regressors on the secret attribute prediction task. As can be seen from the right plots in Fig. 3 and 4, as noise levels increase, the utility sinks by at least 3 times for all 27 regressors. This is suggested by the increase in MAE ratio with increasing noise levels. On the other hand, even with sufficiently high noise levels, the utility for TN-corrupted regressors is well maintained. An MAE Ratio ≈ 1 in Fig. 5 hints at it. Hence TN models are better at maintaining a good privacy-utility tradeoff as compared to gaussian and laplacian noise models.

Tables for all the graphs are present in appendix B.

5 Conclusion & future works

To conclude, we aim at understanding the behavior of the proposed reconstruction attack on noisy linear regressors in the small-data regime. We reformulate the attack primal and dual problems to include bias terms from the regressors. We conduct experiments on real-world data using 3 noise models : Gaussian, Laplacian and Truncated-Normal. We observe a unanimous exponential degradation in the attack efficacy for noise levels

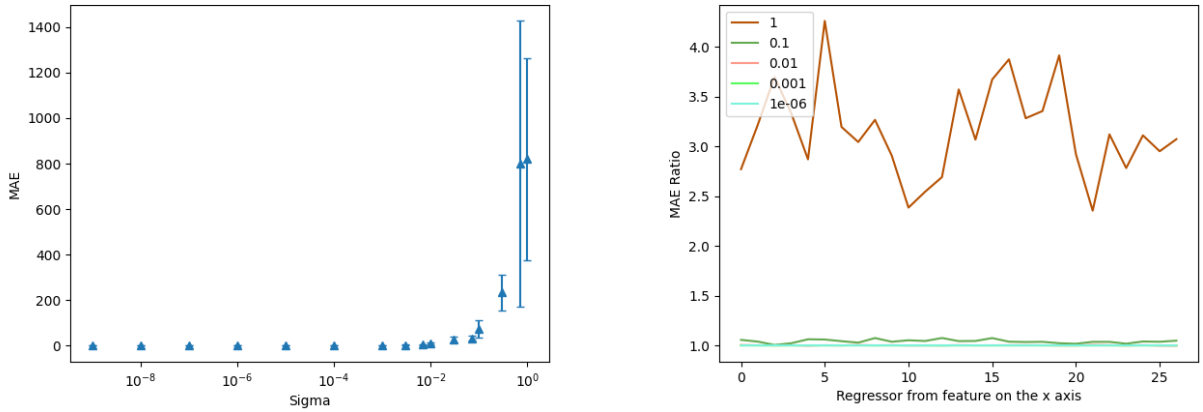

Figure 3: Left Plot: MAE between death rate and its reconstructed version using corrupted published regressors. The noise is controlled by varying σ in the gaussian noise model. Right Plot: Ratio between prediction MAE of noisy and noiseless regressors.

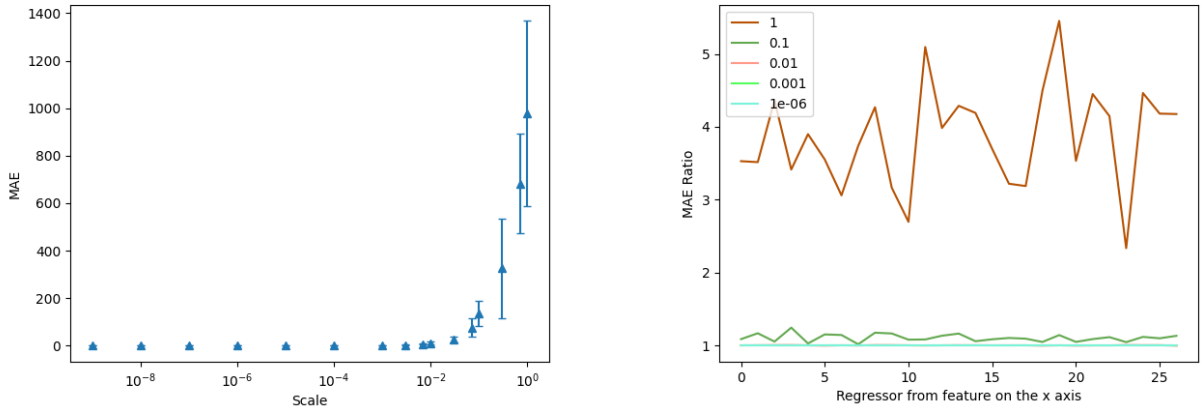

Figure 4: Left Plot: MAE between death rate and its reconstructed version using corrupted published regressors. The noise is controlled by varying *scale* in the laplace noise model. Right Plot: Ratio between prediction MAE of noisy and noiseless regressors.

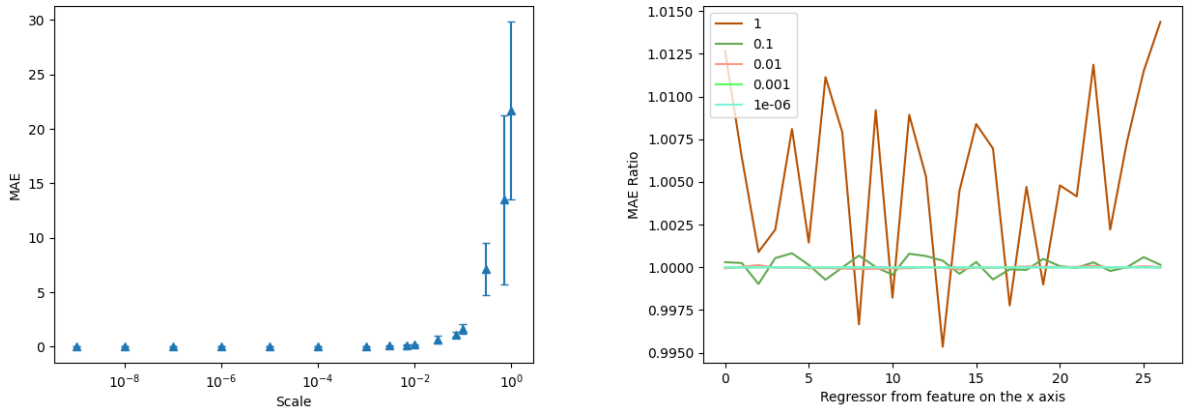

Figure 5: Left Plot: MAE between death rate and its reconstructed version using corrupted published regressors. The noise is controlled by varying *scale* in the truncated-normal noise model. Right Plot: Ratio between prediction MAE of noisy and noiseless regressors.

$\geqslant 0.001$. For smaller noise levels, the reconstructions are remarkably accurate. We also note that the choice of noise model depends on the aim of the publisher. If the aim is to only guarantee high privacy, gaussian or laplacian should be noise models of choice. However, if the goal is to maintain a good privacy-utility balance, Truncated-Normal noise model should be used.

This type of work has many open questions. On similar lines, one can investigate the effect of noise models for other instances of M-estimators like logistic regression. One can also dig deeper into the effectiveness of truncated normal distribution as a noise model, since it yields the best privacy-utility tradeoff but is relatively uncommon in privacy literature. An interesting possible direction is to create reconstruction attacks using robust optimization techniques when the public data can also be noisy.

6 Task Assignment

Both authors contributed equally to the tasks. Tasks include investigating the previous work, reformulating the primal and deriving dual formulations, processing the dataset, coding linear regression to obtain θ, perturbing θs using different noise models, coding the convex optimization problem in cvxpy and writing the report.

References

I. Dinur and K. Nissim. Revealing information while preserving privacy. In *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, PODS '03, pages 202–210, New York, NY, USA, June 2003. Association for Computing Machinery. ISBN 978-1-58113-670-8. doi: 10.1145/773153.773173. URL https://doi.org/10.1145/773153.773173.

C. Dwork and A. Roth. The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, Aug. 2014. ISSN 1551-305X. doi: 10.1561/0400000042. URL https://doi.org/10.1561/0400000042.

C. Dwork and S. Yekhanin. New Efficient Attacks on Statistical Disclosure Control Mechanisms. 5157, Aug. 2008. URL https://www.microsoft.com/en-us/research/publication/new-efficient-attacks-on-statistical-disclosure-control-mechanisms/.

C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating Noise to Sensitivity in Private Data Analysis. In S. Halevi and T. Rabin, editors, *Theory of Cryptography*, Lecture Notes in Computer Science, pages 265–284, Berlin, Heidelberg, 2006. Springer. ISBN 978-3-540-32732-5. doi: 10.1007/11681878_14.

C. Dwork, F. McSherry, and K. Talwar. *The Price of Privacy and the Limits of LP Decoding*. June 2007. ISBN 978-1-59593-631-8. URL https://www.microsoft.com/en-us/research/publication/the-price-of-privacy-and-the-limits-of-lp-decoding/.

C. Dwork, A. Smith, T. Steinke, and J. Ullman. Exposed! A Survey of Attacks on Private Data. *Annual Review of Statistics and Its Application*, 4(1):61–84, 2017. doi: 10.1146/annurev-statistics-060116-054123. URL https://doi.org/10.1146/annurev-statistics-060116-054123. _eprint: https://doi.org/10.1146/annurev-statistics-060116-054123.

S. P. Kasiviswanathan, M. Rudelson, and A. Smith. The power of linear reconstruction attacks. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, SODA '13, pages 1415–1433, USA, Jan. 2013. Society for Industrial and Applied Mathematics. ISBN 978-1-61197-251-1.

F. McSherry and K. Talwar. Mechanism Design via Differential Privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 94–103, Oct. 2007. doi: 10.1109/FOCS.2007.66. ISSN: 0272-5428.

A Dataset

Table 1: Dataset Independent Variables

Feature
medincome
pctbachdeg25_over
pctasian
avgdeathsperyear
pctwhite
studypercap
birthrate
pctempprivcoverage
medianagemale
pctunemployed16_over
pctprivatecoverage
percentmarried
pctbachdeg18_24
pctnohs18_24
medianagefemale
pcths25_over
pctpubliccoverage
pctpubliccoveragealone
popest2015
medianage
pctotherrace
incidencerate
avganncount
pcths18_24
pctblack
povertypercent
pctmarriedhouseholds

B Result Tables

Table 2: Dataset Size vs. MAE

Dataset Size	MAE $\pm$ σ
10	0 ± 0
20	0 ± 0
50	15.8 ± 5.605
100	20.01 ± 5.241
200	20.07 ± 2.114
500	20.36 ± 1.392
1000	20.42 ± 0.898
2000	20.37 ± 1.087
3047	20.34 ± 0.564

Table 3: σ vs. Error for Gaussian Noise

Sigma/Scale	Gaussian MAE $\pm\sigma$	Laplace MAE $\pm\sigma$	TN MAE$\pm\sigma$
1	819.18 ± 443.037	976.86 ± 390.848	21.69 ± 8.124
0.7	799.5 ± 627.341	683.43 ± 208.622	13.48 ± 7.724
0.3	233.76 ± 78.517	325.25 ± 209.462	7.1 ± 2.387
0.1	73.19 ± 36.932	135.72 ± 53.591	1.62 ± 0.471
0.07	32.33 ± 10.216	76.48 ± 37.556	1.14 ± 0.223
0.03	27.57 ± 11.343	26.05 ± 11.012	0.65 ± 0.364
0.01	10.11 ± 4.775	10.9 ± 5.196	0.19 ± 0.051
0.007	5.28 ± 1.472	6.92 ± 2.499	0.15 ± 0.04
0.003	2.68 ± 1.365	3.12 ± 1.128	0.08 ± 0.029
0.001	0.82 ± 0.324	1.01 ± 0.633	0.03 ± 0.012
0.0001	0.09 ± 0.049	0.1 ± 0.033	0 ± 0
1e-05	0.01 ± 0.001	0.01 ± 0.003	0 ± 0
1e-06	0 ± 0	0 ± 0	0 ± 0
1e-07	0 ± 0	0 ± 0	0 ± 0
1e-08	0 ± 0	0 ± 0.001	0 ± 0
1e-09	0 ± 0	0 ± 0	0 ± 0