Deriving an Optimal Noise Adding Mechanism for Privacy-Preserving Machine Learning

Mohit Kumar[1,2(✉)], Michael Rossbory[2], Bernhard A. Moser[2], and Bernhard Freudenthaler[2]

[1] Faculty of Computer Science and Electrical Engineering, University of Rostock, Rostock, Germany
mohit.kumar@uni-rostock.de

[2] Software Competence Center Hagenberg, Hagenberg, Austria
{Michael.Rossbory,Bernhard.Moser,Bernhard.Freudenthaler}@scch.at

Abstract. Differential privacy is a standard mathematical framework to quantify the degree to which individual privacy in a statistical dataset is preserved. We derive an optimal (ϵ, δ)–differentially private noise adding mechanism for real-valued data matrices meant for the training of models by machine learning algorithms. The aim is to protect a machine learning algorithm from an adversary who seeks to gain an information about the data from algorithm's output by perturbing the value in a sample of the training data. The fundamental issue of trade-off between privacy and utility is addressed by presenting a novel approach consisting of three steps: (1) the sufficient conditions on the probability density function of noise for (ϵ, δ)–differential privacy of a machine learning algorithm are derived; (2) the noise distribution that, for a given level of entropy, minimizes the expected noise magnitude is derived; (3) using entropy level as the design parameter, the optimal entropy level and the corresponding probability density function of the noise are derived.

Keywords: Privacy · Noise adding mechanism · Machine learning

1 Introduction

The data on which a machine learning or a data analytics algorithm operates might be owned by more than one party and a party may be unwilling to share its real data. The reason being that an algorithm's output may result in a leakage of private or sensitive information regarding the data. Differential privacy [3,5] is a standard framework to quantify the degree to which the data privacy of

The research reported in this paper has been partly supported by EU Horizon 2020 Grant 826278 "Serums" and the Austrian Ministry for Transport, Innovation and Technology, the Federal Ministry for Digital and Economic Affairs, and the Province of Upper Austria in the frame of the COMET center SCCH.

G. Anderst-Kotsis et al. (Eds.): DEXA 2019 Workshops, CCIS 1062, pp. 108–118, 2019.
https://doi.org/10.1007/978-3-030-27684-3_15

each individual in the dataset is preserved while releasing the algorithm output. Differential privacy is a property of an algorithm's data access mechanism and remains immune to any post-processing on the output of the algorithm. Machine learning methods such as deep neural networks have delivered remarkable results in a wide range of application domains. However, their training requires large datasets which might be containing sensitive information that need to be be protected from *model inversion* attack [6] and such issues have been addressed within the framework of differential privacy [1,14].

The classical approach for attaining differential privacy for a real-valued function, where the function represents mathematically a machine learning algorithm, is to perturb the function output via adding noise calibrated to the global *sensitivity* of the function [4]. Adding of required amount (for attaining a given level of privacy) of noise would result in a loss of algorithm's accuracy and thus it is important to study the trade-off between privacy and accuracy [2,10]. A general framework to provide utility guarantees for a single count query, subject to ϵ–differential privacy, was studied in [11]. A similar study taking a minimax model of utility for information consumers has been made in [12]. For single real-valued query function, a staircase-shaped probability density function was suggested in [8] for an optimal ϵ–differentially private noise adding mechanism. The approach was extended to the vector real-valued query function in [7]. For integer-valued query functions, the optimal mechanisms in (ϵ, δ)–differential privacy were studied in [9]. For single real-valued query function, the trade-off between privacy and utility in $(0, \delta)$–differential privacy was studied in [10].

Despite the fact that random noise adding mechanism has been widely used in privacy-preserving machine learning via output perturbation, there remains the challenge of studying privacy-utility trade-off for the algorithms performing a learning of the models with the matrix data (where e.g. rows corresponds to features and columns corresponds to samples). The aim is to protect a machine learning algorithm from an adversary who seeks to gain an information about the data from algorithm's output by perturbing the value in an element of the training data matrix. There is no standard approach to optimally design (ϵ, δ)–differentially private noise adding mechanism for real-valued data matrices used by a machine learning algorithm for the model training purpose. This study fills this gap by providing a general random noise adding mechanism for real-valued data matrices such that the mechanism, subject to (ϵ, δ)–differential privacy of a machine learning algorithm, minimizes the expected noise magnitude. To the best knowledge of authors, this is the first study of its kind to provide entropy based approach for resolving the privacy-utility trade-off for real-valued data matrices.

2 Sufficient Conditions for Differential Privacy

Consider a dataset consisting of N number of samples with each sample having p number of attributes. Assuming the data as numeric, the dataset can be represented by a matrix, say $\mathrm{Y} \in \mathbb{R}^{p \times N}$. The machine learning algorithms

typically train a model using available dataset. A given machine learning algorithm, training a model using data matrix Y, can be represented by a mapping, $\mathcal{A} : \mathbb{R}^{p\times N} \rightarrow \mathbf{M}$, where $\mathbf{M}$ is the model space. That is, for a given dataset Y, the algorithm builds a model $\mathcal{M} \in \mathbf{M}$ such that $\mathcal{M} = \mathcal{A}(\mathrm{Y})$. The privacy of data can be preserved via adding a suitable random noise to data matrix before the application of algorithm $\mathcal{A}$ on the dataset. This will result in a private version of algorithm $\mathcal{A}$ which is formally defined by Definition 1.

Definition 1 (A Private Algorithm on Data Matrix). *Let* $\mathcal{A}^{+} : \mathbb{R}^{p\times N} \rightarrow Range(\mathcal{A}^{+})$ *be a mapping defined as*

$$\mathcal{A}^{+}(\mathrm{Y}) = \mathcal{A}(\mathrm{Y} + \mathrm{V}),\ \mathrm{V} \in \mathbb{R}^{p\times N} \tag{1}$$

where V *is a random noise matrix with* $f_{\mathrm{v}_j^i}(v)$ *being the probability density function of its* (j,i)*–th element* v_j^i*;* v_j^i *and* $\mathrm{v}_j^{i'}$ *are independent from each other for* $i \neq i'$*; and* $\mathcal{A} : \mathbb{R}^{p\times N} \rightarrow \mathbf{M}$ *(where* $\mathbf{M}$ *is the model space) is a given mapping representing a machine learning algorithm. The range of* $\mathcal{A}^{+}$ *is as*

$$Range(\mathcal{A}^{+}) = \left\{\mathcal{A}(\mathrm{Y} + \mathrm{V}) \mid \mathrm{Y} \in \mathbb{R}^{p\times N}, \mathrm{V} \in \mathbb{R}^{p\times N}\right\}. \tag{2}$$

We intend to protect the algorithm $\mathcal{A}^{+}$ from an adversary who seeks to gain an information about the data from algorithm's output by perturbing the values in a sample of the dataset. We seek to attain differential privacy for algorithm $\mathcal{A}^{+}$ against the perturbation in an element of Y, say (j_0, i_0)–th element, such that magnitude of the perturbation is upper bounded by a scalar d. The d–adjacency [13] definition for two real matrices is provided in Definition 2.

Definition 2 (d–Adjacency for Data Matrices). *Two matrices* $\mathrm{Y}, \mathrm{Y}' \in \mathbb{R}^{p\times N}$ *are* d*–adjacent if for a given* $d \in \mathbb{R}_{+}$*, there exist* $i_0 \in \{1, 2, \cdots, N\}$ *and* $j_0 \in \{1, 2, \cdots, p\}$ *such that* $\forall i \in \{1, 2, \cdots, N\}, j \in \{1, 2, \cdots, p\}$*,*

$$\left|\mathrm{y}_j^i - \mathrm{y}_j'^i\right| \leq \begin{cases} d, \text{ if } i = i_0, j = j_0 \\ 0, \quad \text{otherwise} \end{cases}$$

where y_j^i *and* $\mathrm{y}_j'^i$ *denote the* (j,i)*–th element of* Y *and* Y′ *respectively. Thus,* Y *and* Y′ *differ by only one element and the magnitude of the difference is upper bounded by* d*.*

Definition 3 ((ϵ, δ)–Differential Privacy for $\mathcal{A}^{+}$). *The algorithm* $\mathcal{A}^{+}(\mathrm{Y})$ *is* (ϵ, δ)*–differentially private if*

$$Pr\{\mathcal{A}^{+}(\mathrm{Y}) \in \mathcal{O}\} \leq \exp(\epsilon) Pr\{\mathcal{A}^{+}(\mathrm{Y}') \in \mathcal{O}\} + \delta \tag{3}$$

for any measurable set $\mathcal{O} \subseteq Range(\mathcal{A}^{+})$ *and for* d*–adjacent matrices pair* $(\mathrm{Y}, \mathrm{Y}')$*.*

Result 1 (Sufficient Conditions for (ϵ, δ)–Differential Privacy). *The following conditions on the probability density function of noise $\mathrm{v}_j^i \in \mathbb{R}$ are sufficient to attain (ϵ, δ)–differential privacy by algorithm $\mathcal{A}^+$ (Definition 1):*

$$\int_{\Theta} f_{\mathrm{v}_j^i}(v)\ \mathrm{d}v \geq 1 - \delta, \quad \textit{where} \tag{4}$$

$$\Theta \stackrel{\text{def}}{=} \left\{ v \mid \sup_{\hat{d} \in [-d,d]} \frac{f_{\mathrm{v}_j^i - \hat{d}}(v)}{f_{\mathrm{v}_j^i}(v)} \leq \exp(\epsilon),\ f_{\mathrm{v}_j^i}(v) \neq 0,\ \mathrm{v}_j^i \in \mathbb{R} \right\}. \tag{5}$$

Proof. The proof follows an approach similar to that of [13]. Define a set $\mathbf{S} \subseteq \mathbb{R}^{p \times N}$ as

$$\mathbf{S} = \{\mathrm{Y} + \mathrm{V} \mid \mathcal{A}(\mathrm{Y} + \mathrm{V}) \in \mathcal{O}\}. \tag{6}$$

Further, define $\mathbf{S}_j^i \subseteq \mathbb{R}$ as the set of (j,i)–th elements of members in $\mathbf{S}$, i.e.,

$$\mathbf{S}_j^i = \left\{ \mathrm{y}_j^i + \mathrm{v}_j^i \mid \mathcal{A}(\mathrm{Y} + \mathrm{V}) \in \mathcal{O} \right\}. \tag{7}$$

We have

$$\begin{aligned} Pr\{\mathcal{A}^+(\mathrm{Y}) \in \mathcal{O}\} &= Pr\{\mathcal{A}(\mathrm{Y} + \mathrm{V}) \in \mathcal{O}\} && (8)\\ &= Pr\{\mathrm{Y} + \mathrm{V} \in \mathbf{S}\} && (9)\\ &= \prod_{j=1}^{p} \prod_{i=1}^{N} Pr\{\mathrm{y}_j^i + \mathrm{v}_j^i \in \mathbf{S}_j^i\}. && (10) \end{aligned}$$

Considering Y and Y′ as d–adjacent matrices, there must exist an index, say (j_0, i_0), at which Y and Y′ differ in value. Equality (10) can be expressed as

$$Pr\{\mathcal{A}^+(\mathrm{Y}) \in \mathcal{O}\} = Pr\{\mathrm{y}_{j_0}^{i_0} + \mathrm{v}_{j_0}^{i_0} \in \mathbf{S}_{j_0}^{i_0}\} \prod_{j,i,j \neq j_0, i \neq i_0} Pr\{\mathrm{y}_j^i + \mathrm{v}_j^i \in \mathbf{S}_j^i\} \tag{11}$$

Now, consider

$$\begin{aligned} &Pr\{\mathrm{y}_{j_0}^{i_0} + \mathrm{v}_{j_0}^{i_0} \in \mathbf{S}_{j_0}^{i_0}\}\\ &= Pr\{\mathrm{y}_{j_0}^{i_0} + \mathrm{v}_{j_0}^{i_0} \in \mathbf{S}_{j_0}^{i_0} \mid \mathrm{v}_{j_0}^{i_0} \in (\mathbb{R} \setminus \Theta)\} + Pr\{\mathrm{y}_{j_0}^{i_0} + \mathrm{v}_{j_0}^{i_0} \in \mathbf{S}_{j_0}^{i_0} \mid \mathrm{v}_{j_0}^{i_0} \in \Theta\}. \end{aligned} \tag{12}$$

It follows from the d–adjacency that there exists a $\hat{d} \in [-d, d]$ such that

$$\mathrm{y}_{j_0}^{i_0} = \mathrm{y}_{j_0}'^{i_0} - \hat{d}.$$

Thus,

$$\begin{aligned} &Pr\{\mathrm{y}_{j_0}^{i_0} + \mathrm{v}_{j_0}^{i_0} \in \mathbf{S}_{j_0}^{i_0}\}\\ &= Pr\{\mathrm{y}_{j_0}'^{i_0} - \hat{d} + \mathrm{v}_{j_0}^{i_0} \in \mathbf{S}_j^{i_0} \mid \mathrm{v}_{j_0}^{i_0} \in (\mathbb{R} \setminus \Theta)\} + Pr\{\mathrm{y}_{j_0}'^{i_0} - \hat{d} + \mathrm{v}_{j_0}^{i_0} \in \mathbf{S}_{j_0}^{i_0} \mid \mathrm{v}_{j_0}^{i_0} \in \Theta\}\\ &= \int_{\{\mathrm{y}_{j_0}'^{i_0} - \hat{d} + \mathrm{v}_{j_0}^{i_0} \mid \mathrm{v}_{j_0}^{i_0} \in (\mathbb{R} \setminus \Theta)\} \cap \mathbf{S}_{j_0}^{i_0}} f_{\mathrm{y}_{j_0}'^{i_0} - \hat{d} + \mathrm{v}_{j_0}^{i_0}}(v)\ \mathrm{d}v\\ &\quad + \int_{\{\mathrm{y}_{j_0}'^{i_0} - \hat{d} + \mathrm{v}_{j_0}^{i_0} \mid \mathrm{v}_{j_0}^{i_0} \in \Theta\} \cap \mathbf{S}_{j_0}^{i_0}} f_{\mathrm{y}_{j_0}'^{i_0} - \hat{d} + \mathrm{v}_{j_0}^{i_0}}(v)\ \mathrm{d}v. \end{aligned} \tag{13}$$

Now, we derive upper bounds on both terms at the right hand side of (13). First, consider

$$\int_{\{\mathrm{y}'^{i_0}_{j_0}-\hat{d}+\mathrm{v}^{i_0}_{j_0}\ |\ \mathrm{v}^{i_0}_{j_0}\in(\mathbb{R}\setminus\Theta)\}\cap\mathbf{S}^{i_0}_{j_0}} f_{\mathrm{y}'^{i_0}_{j_0}-\hat{d}+\mathrm{v}^{i_0}_{j_0}}(v)\ \mathrm{d}v$$
$$\leq \int_{\{\mathrm{y}'^{i_0}_{j_0}-\hat{d}+\mathrm{v}^{i_0}_{j_0}\ |\ \mathrm{v}^{i_0}_{j_0}\in(\mathbb{R}\setminus\Theta)\}} f_{\mathrm{y}'^{i_0}_{j_0}-\hat{d}+\mathrm{v}^{i_0}_{j_0}}(v)\ \mathrm{d}v \tag{14}$$

$$= \int_{\mathbb{R}\setminus\Theta} f_{\mathrm{v}^{i_0}_{j_0}}(v)\ \mathrm{d}v \tag{15}$$

$$= 1 - \int_{\Theta} f_{\mathrm{v}^{i_0}_{j_0}}(v)\ \mathrm{d}v. \tag{16}$$

It follows from the definition of Θ, i.e. from (5), that

$$\int_{\{\mathrm{y}'^{i_0}_{j_0}-\hat{d}+\mathrm{v}^{i_0}_{j_0}\ |\ \mathrm{v}^{i_0}_{j_0}\in\Theta\}\cap\mathbf{S}^{i_0}_{j_0}} f_{\mathrm{y}'^{i_0}_{j_0}-\hat{d}+\mathrm{v}^{i_0}_{j_0}}(v)\ \mathrm{d}v$$
$$\leq \exp(\epsilon)\int_{\{\mathrm{y}'^{i_0}_{j_0}-\hat{d}+\mathrm{v}^{i_0}_{j_0}\ |\ \mathrm{v}^{i_0}_{j_0}\in\Theta\}\cap\mathbf{S}^{i_0}_{j_0}} f_{\mathrm{y}'^{i_0}_{j_0}+\mathrm{v}^{i_0}_{j_0}}(v)\ \mathrm{d}v \tag{17}$$

$$\leq \exp(\epsilon)\int_{\mathbf{S}^{i_0}_{j_0}} f_{\mathrm{y}'^{i_0}_{j_0}+\mathrm{v}^{i_0}_{j_0}}(v)\ \mathrm{d}v \tag{18}$$

$$= \exp(\epsilon) Pr\{\mathrm{y}'^{i_0}_{j_0}+\mathrm{v}^{i_0}_{j_0}\in\mathbf{S}^{i_0}_{j_0}\}. \tag{19}$$

Using (16) and (19) in (13), we have

$$Pr\{\mathrm{y}^{i_0}_{j_0}+\mathrm{v}^{i_0}_{j_0}\in\mathbf{S}^{i_0}_{j_0}\} \leq 1 - \int_{\Theta} f_{\mathrm{v}^{i_0}_{j_0}}(v)\ \mathrm{d}v + \exp(\epsilon) Pr\{\mathrm{y}'^{i_0}_{j_0}+\mathrm{v}^{i_0}_{j_0}\in\mathbf{S}^{i_0}_{j_0}\}. \tag{20}$$

Under condition (4), inequality (20) leads to

$$Pr\{\mathrm{y}^{i_0}_{j_0}+\mathrm{v}^{i_0}_{j_0}\in\mathbf{S}^{i_0}_{j_0}\} \leq \delta + \exp(\epsilon) Pr\{\mathrm{y}'^{i_0}_{j_0}+\mathrm{v}^{i_0}_{j_0}\in\mathbf{S}^{i_0}_{j_0}\}. \tag{21}$$

Using (21) in (11), we have

$$Pr\{\mathcal{A}^{+}(\mathrm{Y})\in\mathcal{O}\}$$
$$\leq \delta \prod_{j,i,j\neq j_0,i\neq i_0} Pr\{\mathrm{y}^{i}_{j}+\mathrm{v}^{i}_{j}\in\mathbf{S}^{i}_{j}\}$$
$$+\ \exp(\epsilon) Pr\{\mathrm{y}'^{i_0}_{j_0}+\mathrm{v}^{i_0}_{j_0}\in\mathbf{S}^{i_0}_{j_0}\} \prod_{j,i,j\neq j_0,i\neq i_0} Pr\{\mathrm{y}^{i}_{j}+\mathrm{v}^{i}_{j}\in\mathbf{S}^{i}_{j}\} \tag{22}$$

$$\leq \delta + \exp(\epsilon) Pr\{\mathrm{y}'^{i_0}_{j_0} + \mathrm{v}^{i_0}_{j_0} \in \mathbf{S}^{i_0}_{j_0}\} \prod_{j,i,j\neq j_0,i\neq i_0} Pr\{\mathrm{y}^{i}_{j} + \mathrm{v}^{i}_{j} \in \mathbf{S}^{i}_{j}\} \tag{23}$$

$$= \delta + \exp(\epsilon) Pr\{\mathrm{y}'^{i_0}_{j_0} + \mathrm{v}^{i_0}_{j_0} \in \mathbf{S}^{i_0}_{j_0}\} \prod_{j,i,j\neq j_0,i\neq i_0} Pr\{\mathrm{y}'^{i}_{j} + \mathrm{v}^{i}_{j} \in \mathbf{S}^{i}_{j}\} \tag{24}$$

$$= \delta + \exp(\epsilon) Pr\{\mathrm{Y}' + \mathrm{V} \in \mathbf{S}\} \tag{25}$$

$$= \delta + \exp(\epsilon) Pr\{\mathcal{A}\left(\mathrm{Y}' + \mathrm{V}\right) \in \mathcal{O}\} \tag{26}$$

$$= \delta + \exp(\epsilon) Pr\{\mathcal{A}^{+}\left(\mathrm{Y}'\right) \in \mathcal{O}\}. \tag{27}$$

That is, the condition (3) is satisfied and hence the result is proved.

Remark 1 (Sufficient Conditions for ϵ–Differential Privacy). The sufficient conditions for ϵ–differential privacy follow from (4) with $\delta = 0$ as

$$\int_{\Theta} f_{\mathrm{v}^{i}_{j}}(v)\ \mathrm{d}v = 1, \tag{28}$$

$$\Theta = \left\{ v \mid \sup_{\hat{d}\in[-d,d]} \frac{f_{\mathrm{v}^{i}_{j}-\hat{d}}(v)}{f_{\mathrm{v}^{i}_{j}}(v)} \leq \exp(\epsilon),\ f_{\mathrm{v}^{i}_{j}}(v) \neq 0,\ \mathrm{v}^{i}_{j} \in \mathbb{R} \right\}. \tag{29}$$

where Θ is defined as in (5). The equality in (28) is due to the fact that the integral of any probability density function over a subset can't exceed unity.

3 An Optimal Differentially Private Noise

Result 2 (Minimum Magnitude for a Given Entropy Level). *The probability density function of noise that, for a given level of entropy, minimizes the expected noise magnitude is given as*

$$f^{*}_{\mathrm{v}^{i}_{j}}(v; h) = \frac{1}{\exp(h-1)} \exp(-\frac{2|v|}{\exp(h-1)}), \tag{30}$$

where h is the given entropy level. The expected noise magnitude is given as

$$E_{f^{*}_{\mathrm{v}^{i}_{j}}}\left[|v|\right](h) = \frac{1}{2}\exp(h-1). \tag{31}$$

Proof. We seek to solve

$$f^*_{\mathrm{v}^i_j}(v;h) = \arg \min_{f_{\mathrm{v}^i_j}(v)} \int_{\mathbb{R}} |v| f_{\mathrm{v}^i_j}(v)\ \mathrm{d}v \tag{32}$$

subject to

$$\int_{\mathbb{R}} f_{\mathrm{v}^i_j}(v)\ \mathrm{d}v = 1 \tag{33}$$

$$-\int_{\mathbb{R}} \log\left(f_{\mathrm{v}^i_j}(v)\right) f_{\mathrm{v}^i_j}(v)\ \mathrm{d}v = h. \tag{34}$$

Introducing Lagrange multiplier λ_1 for (33) and λ_2 for (34), the following Lagrangian is obtained:

$$\begin{aligned}\mathcal{L}(f_{\mathrm{v}^i_j}, \lambda_1, \lambda_2) = &\int_{\mathbb{R}} |v| f_{\mathrm{v}^i_j}(v)\ \mathrm{d}v + \lambda_1 \left(\int_{\mathbb{R}} f_{\mathrm{v}^i_j}(v)\ \mathrm{d}v - 1\right) \\ &+ \lambda_2 \left(h + \int_{\mathbb{R}} \log\left(f_{\mathrm{v}^i_j}(v)\right) f_{\mathrm{v}^i_j}(v)\ \mathrm{d}v\right).\end{aligned}$$

The functional derivative of $\mathcal{L}$ with respect to $f_{\mathrm{v}^i_j}$ is given as

$$\frac{\delta \mathcal{L}}{\delta f_{\mathrm{v}^i_j}} = |v| + \lambda_1 + \lambda_2 \left(1 + \log\left(f_{\mathrm{v}^i_j}(v)\right)\right). \tag{35}$$

Setting $\delta\mathcal{L}/\delta f_{\mathrm{v}^i_j}$ equal to zero, we have

$$f_{\mathrm{v}^i_j}(v) = \exp(-1 - \frac{\lambda_1}{\lambda_2}) \exp(-\frac{|v|}{\lambda_2}),\ \lambda_2 \neq 0. \tag{36}$$

Setting $\partial\mathcal{L}/\partial\lambda_1$ equal to zero and then solving using (36), we get

$$f_{\mathrm{v}^i_j}(v) = \frac{1}{2\lambda_2} \exp(-\frac{|v|}{\lambda_2}),\ \lambda_2 > 0. \tag{37}$$

Setting $\partial\mathcal{L}/\partial\lambda_2$ equal to zero and then solving using (37), we get the optimal value of λ_2 as

$$\lambda^*_2 = \frac{1}{2} \exp(h - 1). \tag{38}$$

Using the optimal value of λ^*_2 in (37), the optimal expression for $f_{\mathrm{v}^i_j}(v)$ is obtained as in (30). As $\lambda^*_2 > 0$, $\mathcal{L}$ is convex in $f_{\mathrm{v}^i_j}$ and thus $f^*_{\mathrm{v}^i_j}$ corresponds to the minimum. Finally, the expected noise magnitude for $f^*_{\mathrm{v}^i_j}$ is given by (31).

Result 3 (An Optimal ϵ–Differentially Private Noise). *The probability density function of noise that minimizes the expected noise magnitude together with satisfying the sufficient conditions for ϵ–differential privacy is given as*

$$f^*_{\mathrm{v}^i_j}(v) = \frac{\epsilon}{2d} \exp(-\frac{\epsilon}{d}|v|). \tag{39}$$

The optimal value of expected noise magnitude is given as

$$E_{f^*_{\mathrm{v}^i_j}}[|v|] = \frac{d}{\epsilon}. \tag{40}$$

Proof. Let h^* be the entropy of the optimal probability density function of the noise satisfying the sufficient condition for ϵ–differential privacy. It follows from Result 2 that the expression for optimal probability density function is given as

$$f^*_{\mathrm{v}^i_j}(v; h^*) = \frac{1}{\exp(h^* - 1)} \exp(-\frac{2|v|}{\exp(h^* - 1)}). \tag{41}$$

Now, we have

$$\sup_{\hat{d} \in [-d,d]} \frac{f^*_{\mathrm{v}^i_j - \hat{d}}(v; h^*)}{f^*_{\mathrm{v}^i_j}(v; h^*)} = \exp(\frac{2d}{\exp(h^* - 1)}). \tag{42}$$

Since $f^*_{\mathrm{v}^i_j}(v; h^*)$ satisfies the sufficient conditions (28–29), we have

$$\exp(\frac{2d}{\exp(h^* - 1)}) \leq \exp(\epsilon). \tag{43}$$

That is,

$$\frac{1}{2} \exp(h^* - 1) \geq \frac{d}{\epsilon}. \tag{44}$$

The left hand side of (44) is equal to the expected noise magnitude for $f^*_{\mathrm{v}^i_j}(v; h^*)$. That is,

$$E_{f^*_{\mathrm{v}^i_j}}[|v|](h^*) \geq \frac{d}{\epsilon}. \tag{45}$$

It follows from (45) that the minimum possible value of expected noise magnitude is equal to the right hand side of (45). The value of h^*, resulting in the minimum expected noise magnitude, is given as

$$h^* = 1 + \log\left(2\frac{d}{\epsilon}\right). \tag{46}$$

The value of h^* is put into (41) to obtain (39). The optimal density function (39) satisfies the sufficient conditions (28–29) for $\Theta = \mathbb{R}$.

Result 3 justifies the widely used Laplacian distribution for ϵ–differential privacy.

Result 4 (An Optimal (ϵ, δ)–Differentially Private Noise). *The probability density function of noise that minimizes the expected noise magnitude together with satisfying the sufficient conditions for (ϵ, δ)–differential privacy is given as*

$$f^*_{\mathrm{v}^i_j}(v) = \begin{cases} \delta Dirac\delta(v), & v = 0 \\ (1-\delta)\frac{\epsilon}{2d}\exp(-\frac{\epsilon}{d}|v|), & v \in \mathbb{R} \setminus \{0\} \end{cases} \tag{47}$$

where $Dirac\delta(v)$ is Dirac delta function satisfying $\int_{-\infty}^{\infty} Dirac\delta(v)\ \mathrm{d}v = 1$. The optimal value of expected noise magnitude is given as

$$E_{f^*_{\mathrm{v}^i_j}}[|v|] = (1-\delta)\frac{d}{\epsilon}. \tag{48}$$

Proof. It is obvious that the optimal noise density function (39) satisfies the sufficient conditions (4–5) with $\Theta = \mathbb{R}$ for any $\delta \in [0, 1]$ and thus attain (ϵ, δ)–differential privacy for any $\delta \in [0, 1]$. However, in this case (i.e. when $\Theta = \mathbb{R}$ and $\delta > 0$), the lower bound on $\int_\Theta f_{\mathrm{v}^i_j}(v)\ \mathrm{d}v$ in (4) is not tight. Therefore, we need to derive an optimal density function for (ϵ, δ)–differential privacy taking $\Theta \subset \mathbb{R}$. Let $v_0 \in \mathbb{R}$ be a point which is excluded from $\mathbb{R}$ to define Θ, i.e.,

$$\Theta = \mathbb{R} \setminus \{v_0\}. \tag{49}$$

We extend the solution space for optimization by considering the discontinuous distributions having an arbitrary probability mass r at an arbitrary point v_0. Let $f_{\mathrm{v}^i_j}\left(v; v_0, r, q_{\mathrm{v}^i_j}(v)\right)$ be an arbitrary density function defined as

$$f_{\mathrm{v}^i_j}\left(v; v_0, r, q_{\mathrm{v}^i_j}(v)\right) = \begin{cases} rDirac\delta(v - v_0), & v = v_0 \\ (1-r)q_{\mathrm{v}^i_j}(v), & v \in \Theta \end{cases} \tag{50}$$

Here, $q_{\mathrm{v}^i_j}(v)$ is an arbitrary density function with a continuous cumulative distribution function and satisfying the sufficient conditions (28–29) for ϵ–differential privacy. As $q_{\mathrm{v}^i_j}(v)$ is an arbitrary density function, the expected noise magnitude for $q_{\mathrm{v}^i_j}(v)$ must be greater than or equal to the optimal value (40), i.e.,

$$\int_{\mathbb{R}} |v| q_{\mathrm{v}^i_j}(v)\ \mathrm{d}v \geq \frac{d}{\epsilon} \tag{51}$$

$$\int_{\Theta} |v| q_{\mathrm{v}^i_j}(v)\ \mathrm{d}v + \underbrace{\int_{\{v_0\}} |v| q_{\mathrm{v}^i_j}(v)\ \mathrm{d}v}_{=0} \geq \frac{d}{\epsilon}. \tag{52}$$

Here, the integral over a single point is equal to zero because of a continuous cumulative distribution function associated to $q_{\mathrm{v}^i_j}(v)$. Thus,

$$\int_{\Theta} |v| q_{\mathrm{v}^i_j}(v)\ \mathrm{d}v \geq \frac{d}{\epsilon}, \tag{53}$$

where equality occurs if $q_{\mathrm{v}_j^i}(v)$ is equal to (39). Also

$$\int_{\Theta} q_{\mathrm{v}_j^i}(v)\ \mathrm{d}v = \int_{\mathbb{R}} q_{\mathrm{v}_j^i}(v)\ \mathrm{d}v - \int_{\{v_0\}} q_{\mathrm{v}_j^i}(v)\ \mathrm{d}v \tag{54}$$

$$= 1. \tag{55}$$

Thus

$$\int_{\Theta} f_{\mathrm{v}_j^i}\left(v; v_0, r, q_{\mathrm{v}_j^i}(v)\right)\ \mathrm{d}v = 1 - r. \tag{56}$$

For the density function (50) to satisfy condition (4), we must have

$$1 - r \geq 1 - \delta. \tag{57}$$

The expected noise magnitude for the density function (50) is given as

$$E_{f_{\mathrm{v}_j^i}}\left[|v|\right](v_0, r, q_{\mathrm{v}_j^i}(v)) = r \underbrace{|v_0|}_{\geq 0} + \underbrace{(1-r)}_{\geq 1-\delta} \underbrace{\int_{\Theta} |v| q_{\mathrm{v}_j^i}(v)\ \mathrm{d}v}_{\geq d/\epsilon}. \tag{58}$$

It follows immediately that $E_{f_{\mathrm{v}_j^i}}[|v|]$ is minimized together with satisfying the sufficient conditions (4–5) with the following optimal choices for $(v_0, r, q_{\mathrm{v}_j^i}(v))$: $v_0^* = 0$, $r^* = \delta$, and $q_{\mathrm{v}_j^i}^*(v) = \frac{\epsilon}{2d} \exp(-\frac{\epsilon}{d}|v|)$. The result is proved after putting the optimal values into (50).

4 Concluding Remarks

This paper has stated an approach to derive an optimal (ϵ, δ)–differentially private noise adding mechanism for privacy-preserving machine learning. This is the first study to address the fundamental issue of trade-off between privacy and utility for matrix-valued query functions. Using noise entropy level as a design parameter for resolving the privacy-utility trade-off is a novel idea that would be further explored in our future work to link differential privacy with information-theoretic machine learning.

References

1. Abadi, M., et al.: Deep learning with differential privacy. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, CCS 2016, pp. 308–318. ACM, New York (2016)
2. Balle, B., Wang, Y.: Improving the Gaussian mechanism for differential privacy: analytical calibration and optimal denoising. CoRR abs/1805.06530 (2018)
3. Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., Naor, M.: Our data, ourselves: privacy via distributed noise generation. In: Vaudenay, S. (ed.) EUROCRYPT 2006. LNCS, vol. 4004, pp. 486–503. Springer, Heidelberg (2006). https://doi.org/10.1007/11761679_29

4. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: Halevi, S., Rabin, T. (eds.) TCC 2006. LNCS, vol. 3876, pp. 265–284. Springer, Heidelberg (2006). https://doi.org/10.1007/11681878_14
5. Dwork, C., Roth, A.: The algorithmic foundations of differential privacy. Found. Trends Theor. Comput. Sci. **9**(3–4), 211–407 (2014)
6. Fredrikson, M., Jha, S., Ristenpart, T.: Model inversion attacks that exploit confidence information and basic countermeasures. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, CCS 2015, pp. 1322–1333. ACM, New York (2015)
7. Geng, Q., Kairouz, P., Oh, S., Viswanath, P.: The staircase mechanism in differential privacy. IEEE J. Sel. Topics Signal Process. **9**(7), 1176–1184 (2015)
8. Geng, Q., Viswanath, P.: The optimal noise-adding mechanism in differential privacy. IEEE Trans. Inf. Theory **62**(2), 925–951 (2016)
9. Geng, Q., Viswanath, P.: Optimal noise adding mechanisms for approximate differential privacy. IEEE Trans. Inf. Theory **62**(2), 952–969 (2016)
10. Geng, Q., Ding, W., Guo, R., Kumar, S.: Optimal noise-adding mechanism in additive differential privacy. CoRR abs/1809.10224 (2018)
11. Ghosh, A., Roughgarden, T., Sundararajan, M.: Universally utility-maximizing privacy mechanisms. SIAM J. Comput. **41**(6), 1673–1693 (2012)
12. Gupte, M., Sundararajan, M.: Universally optimal privacy mechanisms for minimax agents. In: Proceedings of the Twenty-Ninth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2010, pp. 135–146. ACM, New York (2010)
13. He, J., Cai, L.: Differential private noise adding mechanism: basic conditions and its application. In: 2017 American Control Conference (ACC), pp. 1673–1678, May 2017
14. Phan, N., Wang, Y., Wu, X., Dou, D.: Differential privacy preservation for deep auto-encoders: an application of human behavior prediction. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI 2016, pp. 1309–1316. AAAI Press (2016)