Differential Privacy for Regularised Linear Regression

Ashish Dandekar, Debabrota Basu, and Stéphane Bressan

School of Computing, National University of Singapore, Singapore.
(ashishdandekar, debabrota.basu)@u.nus.edu, steph@nus.edu.sg

Abstract. We present ϵ-differentially private functional mechanisms for variants of regularised linear regression, LASSO, Ridge, and elastic net. We empirically and comparatively analyse their effectiveness. We quantify the error incurred by these ϵ-differentially private functional mechanisms with respect to the non-private linear regression. We show that the functional mechanism is more effective than the state-of-art differentially private mechanism using input perturbation for the three main regularised linear regression models. We also discuss caveats in the functional mechanism, such as non-convexity of the noisy loss function, which causes instability in the results.

Keywords: linear regression, data privacy, differential privacy

1 Introduction

Dwork et al. proposed *differential privacy* [5] to quantify the privacy of a mechanism.

Let $\mathcal{D}$ denotes a universe of d-dimensional real-valued datapoints and the corresponding real-valued responses of a statistical machine learning algorithm M. An element from this universe can be represented by a pair $D = (X, Y)$ where $X \in \mathbb{R}^{n \times d}$ is a matrix and $Y \in \mathbb{R}^n$ is a vector. We use $\|\cdot\|_p$ to represent the L_p norm of a vector. Let us call a *neighbouring* dataset a dataset D' that differs from the dataset D by one datapoint. Differential privacy quantifies the privacy of a randomized algorithm, referred to as a mechanism $\mathcal{M}$, run on those datasets.

Definition 1. *[5] A randomized algorithm $\mathcal{M}$ with domain $\mathcal{D}$ is ϵ-differentially private if for all $S \in Range(\mathcal{M})$ and $D, D' \in \mathcal{D}$ such that D and D': are neighbouring datasets*

$$\log\left(\left|\frac{Pr(\mathcal{M}(D) \in S)}{Pr(\mathcal{M}(D') \in S)}\right|\right) \leq \epsilon$$

Differential privacy introduces noise at selected stages of a statistical machine learning algorithm, for instance in the output of the model [4] or in the input of the algorithm [9] or in the loss function as with the *functional mechanism* [16]. The calibration of these mechanisms requires the computation of *sensitivity* of the function.

Definition 2. *The sensitivity of a function $f : \mathcal{D} \rightarrow \mathbb{R}^k$ is defined as:*

$$\varDelta_f = \max_{x \sim y} \|f(x) - f(y)\|_1$$

Laplace mechanism [4] is a widely used privacy-preserving mechanism. It achieves differential privacy by adding random noise from a Laplace distribution. For a given privacy level ϵ, Laplace distribution is calibrated such that the mean is zero and the scale $\frac{\varDelta_f}{\epsilon}$. The Laplace mechanism calibrated in this way satisfies ϵ-differential privacy [5].

We present ϵ-differentially private functional mechanisms for variants of regularised linear regression, namely LASSO, Ridge, and elastic net. Linear regression [10] is a widely used statistical machine learning model. It uses a linear hypothesis to map a set of predictor attributes of a datapoint to the corresponding response. In matrix notation, linear regression is parameterized by $\theta \in \mathbb{R}^d$ such that $X\theta = Y$. In order to find the optimal value of θ, training step in linear regression minimizes *mean squared loss*, $l_\theta(T)$, over the training data T, as defined in Equation 1.

$$\theta^* = \arg\min_\theta l_\theta(T) = \arg\min_\theta (X\theta - Y)^2 \tag{1}$$

Properties of the coefficient of quadratic term in Equation 1 determine the convexity of the optimisation problem. The optimization problem is made convex by adding a regularisation term to the objective function in Equation 1. With a regularisation term that is proportional to the L_2 norm of the parameters the new optimisation is called *Ridge regression* [8]. It is defined in Equation 2.

$$\theta^* = \arg\min_\theta (X\theta - Y)^2 + \lambda\|\theta\|_2^2 \tag{2}$$

With a regularisation term that is proportional to the L_1 norm of the parameters the new optimisation is called *LASSO regression* [14]. With a regularisation term that is proportional to a convex combination of L_1 and L_2 norms of parameters the new optimisation is called *Elastic net regression* [17].

In Section 3, we extend the work of the authors of [16] for Ridge regression and present the functional mechanism [16] for linear regression and three of its regularised variants, namely, Ridge, LASSO and Elastic net.

In Section 4, we comparatively evaluate the performance of these four mechanisms and their differentially private variants on two datasets with different correlations and sparsity. We observe that the functional mechanism applied to the regularised linear regression yields similar performance results and that the private linear regression models perform worse than the non-private linear regression models. We compare the effectiveness of the functional mechanism with an *input perturbation mechanism* [9]. For a given privacy level, ϵ, we empirically show, for the three main regularised linear regression models, that the functional mechanism is more effective than the state-of-art differentially private mechanism using input perturbation, DPME [9]. We extend the analysis in [16] to empirically study the robustness of the functional mechanism. The key observation in our experiments is that all the private linear regression models are unstable. Our analysis shows that the reason for such an instability is inherent to the functional mechanism. In reference to these experimental evidences, we conclude by puting forth (Section 5) the need of designing a differentially private mechanism that produces a convex noisy loss function in order to provide both stable and private output for linear regression models.

The extended version of this paper is available at [2].

2 Related Work

Linear regression [10] is a fundamental yet a widely used machine learning model. Variants of linear regression, Ridge [8] and LASSO [14], are used to reduce correlation in the data features and to avoid overfitting. Elastic net [17] regression uses convex combination of regularisation terms that are used in Ridge and LASSO. For a detailed presentation and discussion of regularisation and regression analysis, interested readers can refer to [10].

Differential Privacy [5] is a probabilistic framework that quantifies the privacy of a randomized function or algorithm. Existing deterministic machine learning models can be randomized by introducing calibrated random noise. The resultant randomized *mechanism* can be shown to satisfy constraints of differential privacy. Dwork et. al. propose the Laplace mechanism [4], which perturbs the output of a machine learning model by explicitly adding scaled random noise from the Laplace distribution. The Gaussian mechanism [5] and the K-norm mechanism [7] are differentially private mechanisms that are also based on the idea of output perturbation with noise from different distributions. Lei [9] proposes differentially private M-estimators, which perturbs the histogram of input data using a scaled noise and further uses the noisy histogram to train the models. Zhang et. al. [16] propose a differentially private *functional mechanism* that adds a properly scaled Laplace noise to the coefficients of loss function in the polynomial basis. Hall et.al. [6] also propose a differentially private *functional mechanism* that adds a properly scaled noise drawn from the Gaussian process to the coefficients of loss function in the kernel basis.

Zhang et. al. instantiate their functional mechanism on linear regression and logistic regression. In order to alleviate the non-convexity caused in the loss function due to addition of random noise, they use Ridge regularised linear and logistic regressions. Yu et. al. [15] achieve differential privacy in the elastic net logistic regression by controlling the coefficient of regularisation term. The regularisation term in their proposal is inversely proportional to the number of datapoints. It causes reduction in regularisation as the number of datapoints increases. Therefore, their proposed mechanism is not applicable for large datasets. Talwar et. al. [12] propose a differentially private variant of Frank-Wolfie optimisation algorithm to perform LASSO regression. This method adds noise in the optimisation algorithm instead of adding it to the objective function.

3 Functional Mechanism for regularised linear regression

Functional mechanism [16], which is a privacy-preserving mechanism, introduces random noise in the loss function of a machine learning algorithm. Optimisation of such a noisy loss function leads to the parameters that are different than true optimal parameters. In this way, we indirectly get noisy outputs from the machine learning model without explicitly adding noise to the outputs. In this section, we elucidate the details related to the functional mechanism.

For a given machine learning model, the loss function l_θ can be expanded in the polynomial basis, using Stone-Weierstrass theorem [13], as a function of parameter θ as given in Equation 3 where $t = (x, y)$ denotes a datapoint in training dataset T, $\varPhi_j$ denotes the set of polynomials with degree j and $\lambda_{t\phi}$ denote respective coefficients.

$$l_\theta(T) = \sum_{t \in T} \sum_{j=0}^{J} \sum_{\phi \in \varPhi_j} \lambda_{t\phi} \phi(\theta) \tag{3}$$

Lemma 1. *[16] Upper bound on sensitivity of the loss function of a machine learning model is given by:*

$$\varDelta_l = 2 \max_t \sum_{j=1}^{J} \sum_{\phi \in \varPhi_j} \|\lambda_{t\phi}\|_1$$

Using Lemma 1 and the Laplace mechanism, Zhang et. al. [16] devise an algorithm, namely the functional mechanism, that adds noise to loss function of a machine learning model. For a given privacy level ϵ, they use Laplace mechanism calibrated with the sensitivity calculation in Lemma 1 to induce noise in the coefficients of the Taylor expansion of the loss function. Parameters of the machine learning model are estimated by optimizing the noisy loss function. They prove that this algorithm satisfies ϵ-differential privacy. Please refer to Algorithm 1 in [16] for the details.

Elastic net regression [17] adds the regularisation term which is a convex combination of L_1 regularisation term and L_2 regularisation term. The optimisation problem for elastic net regression with functional mechanism is given in Equation 4. $l'_\theta(T)$ denotes the noisy loss function obtained by applying the functional mechanism on the loss function for linear regression, as stated in Equation 1.

$$\theta^* = \arg\min_\theta l'_\theta(T) + \lambda(\alpha\|\theta\|_2^2 + (1-\alpha)\|\theta\|_1) \tag{4}$$

The additive regularisation term is proportional to the norm of the parameters and it does not depend on the training dataset. Therefore, regularisation does not change the sensitivity of the loss function. We use this observation and the sensitivity calculation for linear regression in [16] to compute the sensitivity of Elastic net regression. We present the result below.

Assuming that all features of datapoints are normalized such that each of the feature value lies in $[-1, 1]$, L_1 sensitivity of the loss function in Equation 4 is given by:

$$\varDelta_l = 2(d^2 + 2d + 1)$$

4 Empirical Performance Evaluation

We comparatively and empirically evaluate functional mechanism for regularised linear regressions: namely Ridge, LASSO, and elastic net. We present the result analysis in this section.

We conduct experiments on a microdata sample of US Census in 2000 provided by IPUMS International [1]. The census dataset consists of 1% sample of the original census data. We consider a subset of $316,276$ records of the heads of households in our dataset. Each record has 9 attributes, namely, *Age, Gender, Race, Marital Status, Family Size, Education, Employment Status, House type, Income*. Regression analysis is performed using *Income* as the response variable and the rest of the attributes as predictor variables.

We use Python® 2.7.6 with the SCS [11] solver from CVXPY [3] package.

We report the results as the aggregates over 50 experimental runs. For every experimental run, we randomly hold out 20% of the data for testing and use the rest 80% of the data for training regression models. We normalize each of these features such that their values lie in $[-1, 1]$. We use *root mean squared error (RMSE)* [10] as the metric to comparatively evaluate effectiveness. For given value of ϵ, the model with smallest value of RMSE is the most effective model.

We comparatively evaluate eight regression problems: linear regression (LR), Ridge regression (RG), LASSO regression (LS), elastic net regression (EN), and their private versions. We call the regression model obtained using the functional mechanism *functional regression*. For every regularised regression model, we set, by cross-validation, the regularisation coefficient, λ, that yields smallest testing error.

Figure 1 shows the comparative evaluation of the functional mechanisms with an *input perturbation mechanism*, differentially private M-estimators (DPME) [9]. Discretisation of a large number of attributes leads to a large discrete space that causes prohibitive computation cost. Due to concentration of data around subsets of features, a large discrete space also leads to sparse histograms [9]. In order to alleviate the sparsity, we follow [9] and evaluate the performance on a simpler regression model. We show the comparative study on the census dataset where we predict *Income* of a person using *Age, Gender, Race* and *Education Status*. The results are presented in Figure 1. Solid lines represent the *mean* RMSE over 50 runs. For a given value of ϵ, we observe that the functional mechanism provides lower RMSE for all regularised linear regressions. Thus, we show that the functional mechanism is more effective than DPME.

Now we present the comparative evaluation functional regularised regressions. Figure 2 shows the boxplot of functional elastic net regression for different values of ϵ's. We note the presence of a large number of outliers in the result. We observe similar results for the rest of the functional regressions. In order to avoid this bias due to the outliers, we choose to plot the *median* instead of the *mean*. Figure 3 shows the comparative evaluation of the variants of regularised linear regression for the census quality dataset. In the plot, solid line represents median over 50 experimental runs and the shaded region covers RMSE values that lie between 20^{th} and 80^{th} percentile. Smaller values of ϵ's induce higher noise in the

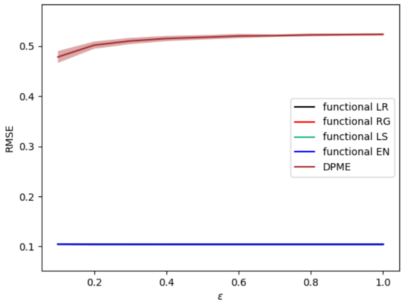

Fig. 1: RMSE of regularised linear regressions for varying values of ϵ's for DPME [9] and the functional mechanism

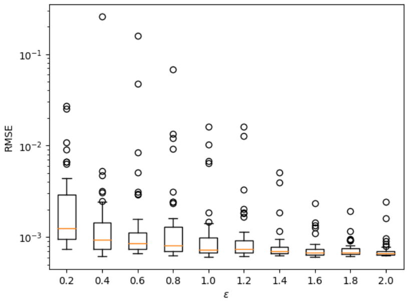

Fig. 2: Boxplot of RMSE of elastic net regression with functional mechanism for different values of ϵ for the census dataset.

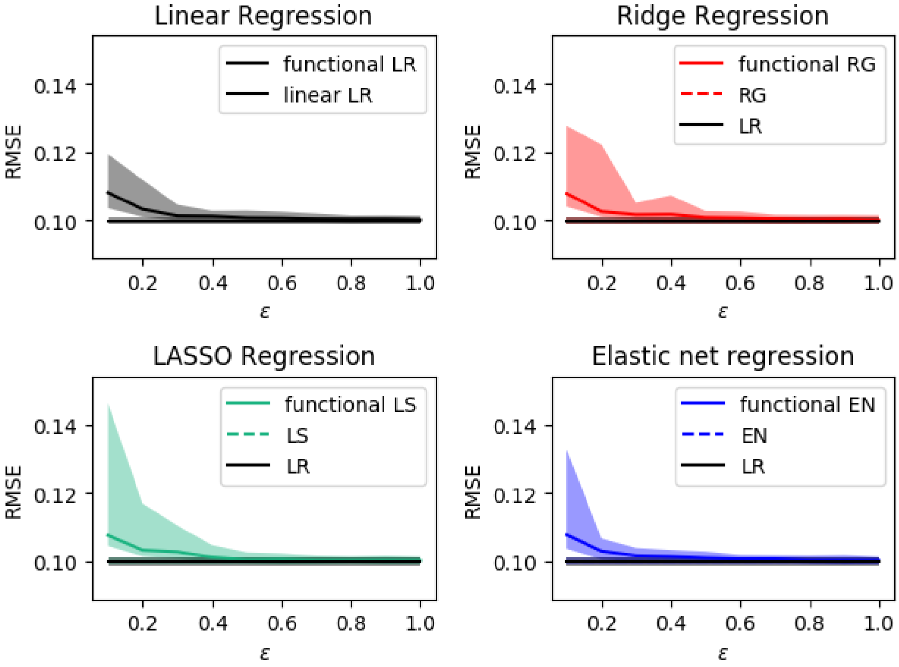

Fig. 3: Comparison of different regressions for the census dataset with *median* as the aggregate

function, which in turn results in higher privacy. Therefore, we observe higher RMSE for smaller values of ϵ's. As the value of ϵ increases, the effectiveness of the functional regression approaches the the effectiveness of the non-private counterpart.

One observation that is common in all the empirical evaluation in Figure 2 and Figure 3 is the instability in the results. M-estimator is a robust statistic [**?**]. We observe the stability in the performance of DPME as compared to the functional regressions. We find that the reasons for this instability are rooted in the functional mechanism it self. We complete the result analysis by the discussion of these possible reasons.

The coefficient of the quadratic term in Equation 1, X^tX, is a symmetric matrix. It loses its symmetric property after adding random noise from the Laplace distribution. A standard way to make a given matrix A symmetric is to use $(A + A^t) * 0.5$. This way of symmetrization of noisy X^tX indirectly incurs addition two Laplace random variables. Addition of two Laplace random variable does not follow Laplace distribution. Therefore, in order to maintain the integrity of the functional mechanism, we can not make X^tX symmetric in the conventional way.

Linear regression works on the assumption that the attributes in a dataset are independent of each other. Independence among the attributes makes X^tX a positive definite matrix. Positive definite matrices make the optimisation convex and guarantees optimality of the solution. Noisy loss function fails to guarantee convexity of the objective problem, and hence the optimality of the solution. A similar observation is made by Lei [9] while perturbing the histograms of input data by adding the calibrated noise. In order to make the objective function convex, Zhang et. al. [16] calculate the spectral decomposition of X^tX and consider the projection of parameters onto the eigenspace spanned by eigenvectors with positive eigenvalues. They do not provide any analytical justification which guarantees differential privacy after pruning the non-positive eigenspace.

Functional mechanism proves that the loss function generated by any two neighbouring datasets satisfies differential privacy. Composition of a differentially private function with a deterministic function, called as post-processing [5], remains differentially private. An optimisation problem solver calculates an approximate solution when the objective function is not convex. Therefore, differential privacy of a loss function is not preserved by the optimisation algorithm itself.

5 Conclusion and Future Works

We present the construction of differentially private versions of the three linear regression models, Ridge, LASSO and Elastic net, using the functional mechanism. We empirically and comparatively evaluate the effectiveness of the private and non-private versions on a census datasets. For a given privacy level, ϵ, we observe that the functional mechanism is more effective than DPME [9] for regularized linear regression. We extend the analysis in [16] to empirically study the robustness of the functional mechanism. As expected, we invariably observe that the private versions are less effective than their non-private counterparts. The key observation from these experiments is that all these private regularised regression methods are equally unstable, and that private linear regression is comparatively more unstable. We analyse the loss of symmetry of the covariance matrix and the non-convexity of the loss function after adding the noise as the principal reasons of this instability. This opens up the need of designing a privacy-preserving mechanism that would retain these properties for private linear regression.

Acknowledgement

This research is supported by the National Research Foundation, Prime Ministers Office, Singapore, under its Corporate Laboratory@University Scheme, National University of Singapore, and Singapore Telecommunications Ltd.

References

1. Minnesota population center. integrated public use microdata series international: Version 5.0. https://international.ipums.org. (2009)
2. Dandekar, A., Basu, D., Bressan, S.: Differential privacy for regularised linear regression. Tech. Rep. TRA6/18, National University of Singapore (Jun 2018), https://dl.comp.nus.edu.sg/handle/1900.100/7051
3. Diamond, S., Boyd, S.: CVXPY: A Python-embedded modeling language for convex optimization. Journal of Machine Learning Research 17(83), 1–5 (2016)
4. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: Theory of Cryptography Conference. pp. 265–284. Springer (2006)
5. Dwork, C., Roth, A., et al.: The algorithmic foundations of differential privacy. Foundations and Trends® in Theoretical Computer Science 9(3–4), 211–407 (2014)
6. Hall, R., Rinaldo, A., Wasserman, L.: Differential privacy for functions and functional data. Journal of Machine Learning Research (JMLR) 14(Feb), 703–727 (2013)
7. Hardt, M., Talwar, K.: On the geometry of differential privacy. In: Proceedings of the forty-second ACM Symposium on Theory of Computing (STOC). ACM (2010)
8. Hoerl, A.E., Kennard, R.W.: Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12(1), 55–67 (1970)
9. Lei, J.: Differentially private m-estimators. In: Advances in Neural Information Processing Systems. pp. 361–369 (2011)
10. Murphy, K.P.: Machine Learning: A Probabilistic Perspective. The MIT Press (2012)
11. O'Donoghue, B., Chu, E., Parikh, N., Boyd, S.: Conic optimization via operator splitting and homogeneous self-dual embedding. Journal of Optimization Theory and Applications 169 (June 2016)
12. Talwar, K., Thakurta, A.G., Zhang, L.: Nearly optimal private lasso. In: Advances in Neural Information Processing Systems (NIPS). pp. 3025–3033 (2015)
13. Thomas, G.B., Weir, Maurice D.and Hass, J.: Thomas calculus (2016)
14. Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) pp. 267–288 (1996)
15. Yu, F., Rybar, M., Uhler, C., Fienberg, S.E.: Differentially-private logistic regression for detecting multiple-snp association in gwas databases. In: International Conference on Privacy in Statistical Databases. pp. 170–184. Springer (2014)
16. Zhang, J., Zhang, Z., Xiao, X., Yang, Y., Winslett, M.: Functional mechanism: regression analysis under differential privacy. Proceedings of the VLDB Endowment 5(11), 1364–1375 (2012)
17. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67(2), 301–320 (2005)