

A note on privacy preserving iteratively reweighted least squares

Mijung Park and Max Welling
QUvA Lab. University of Amsterdam
{mijungi.p, welling.max}@gmail.com

Abstract

Iteratively reweighted least squares (IRLS) is a widely-used method in machine learning to estimate the parameters in the generalised linear models. In particular, IRLS for L1 minimisation under the linear model provides a closed-form solution in each step, which is a simple multiplication between the inverse of the weighted second moment matrix and the weighted first moment vector. When dealing with privacy sensitive data, however, developing a privacy preserving IRLS algorithm faces two challenges. First, due to the inversion of the second moment matrix, the usual sensitivity analysis in differential privacy incorporating a single datapoint perturbation gets complicated and often requires unrealistic assumptions. Second, due to its iterative nature, a significant cumulative privacy loss occurs. However, adding a high level of noise to compensate for the privacy loss hinders from getting accurate estimates. Here, we develop a practical algorithm that overcomes these challenges and outputs privatised and accurate IRLS solutions. In our method, we analyse the sensitivity of each moments separately and treat the matrix inversion and multiplication as a post-processing step, which simplifies the sensitivity analysis. Furthermore, we apply the *concentrated differential privacy* formalism, a more relaxed version of differential privacy, which requires adding a significantly less amount of noise for the same level of privacy guarantee, compared to the conventional and advanced compositions of differentially private mechanisms.

1 Introduction

Differential privacy (DP) algorithms provide strong privacy guarantees by typically perturbing some statistics of a given dataset, which appear in the outputs of an algorithm [1]. The amount of noise added for the perturbation is set in order to compensate for any difference in the probability of any outcome of an algorithm by adding a single individual's datapoint to or removing it from the data. So, in order to develop a DP algorithm, one first needs to analyse the maximum difference in the probability of any outcome, which often called *sensitivity*, to set the level of additive noise.

In this note, we're interested in developing a privacy preserving iteratively reweighted least squares (IRLS) method. In the compressed sensing literature [2], IRLS is used for solving the L-1 minimisation problem, in which a closed-form update of parameters in each step is available. This IRLS solution in each step is a simple multiplication between the inverse of the weighted second moment matrix and the weighted first moment vector. Due to the inverse of the second moment matrix, analysing the sensitivity becomes challenging. Previous work [3] assumes each feature of each datapoint is *i.i.d.* drawn from a standard normal distribution, and analysed the sensitivity of the inverse of the second moment matrix. Unfortunately, the assumption on each features being independent is often not realistic.

Another challenge in developing a privacy preserving IRLS method comes from the iterative nature of the IRLS algorithm. The conventional DP composition theorem (Theorem 3.16 in [1]) states that multiple iterations of a ϵ -DP algorithm faces a linearly degrading privacy, which yields $J\epsilon$ -DP after

J iterations. A more advanced composition theorem (Theorem 3.20 in [1]) yields $(\sqrt{2J\log(1/\delta)}\epsilon + J\epsilon(e^\epsilon - 1), \delta)$ -DP. The new variable δ (stating the mechanism's failure probability) that needs to be set to a very small value, which makes the cumulative privacy loss still relatively high. To compensate for the privacy loss, one needs to add a significant amount of noise to the IRLS solution to avoid revealing any individual information from the output of the algorithm.

In this note, we tackle these challenges by : (1) we analyse the sensitivity of the weighted second moment matrix and the weighted first moment vector separately and perturb each moment by adding noise consistent with its own sensitivity. Then, we do the multiplication of the inverse of perturbed second moment matrix and the perturbed first moment vector. This inversion and multiplication can be viewed as a post-processing step, which doesn't alter the privacy level. Since we perturb each moment separately, this method does not require any restrictive assumptions on the data. In addition, the noise variance naturally scales with the amount of data. (2) we apply the *concentrated differential privacy* formalism, a more relaxed version of differential privacy, to obtain more accurate estimates for the same cumulative privacy loss, compared to DP and its (ϵ, δ) -relaxation. In the following, we start by describing our privacy preserving IRLS algorithm.

2 Privacy preserving IRLS

Given a dataset which consists of N input-output pairs $\{\mathbf{x}_i, y_i\}_{i=1}^N$ where we assume $\|\mathbf{x}_i\|_2 \leq 1$ and $\|y_i\| \leq 1$. The iteratively reweighted least squares solution has the form:

$$\hat{\boldsymbol{\theta}}_{irls}^{(t)} = (X^\top S X)^{-1} X^\top S \mathbf{y} := B^{-1} A \quad (1)$$

where $X \in \mathbb{R}^{N \times d}$ is a design matrix in which the i th row is the transposed i th input \mathbf{x}_i^\top (of length d), and \mathbf{y} is a column vector of outputs. We denote $B = \frac{1}{N} X^\top S X$, and $A = \frac{1}{N} X^\top S \mathbf{y}$. Here S is a diagonal matrix with diagonal $\mathbf{s} = |\mathbf{y} - X \hat{\boldsymbol{\theta}}^{(t-1)}|^{p-2}$. Here we set $p = 1$ and compute $L1$ norm constrained least squares. To avoid dividing by 0, we set

$$\mathbf{s}_i = \frac{1}{\max(1/\delta, |\mathbf{y}_i - X_i \hat{\boldsymbol{\theta}}^{(t-1)}|)}, \quad (2)$$

where X_i is the i th row. δ sets the sparsity (number of non-zero values) of the IRLS solution.

We will perturb each of these statistics A and B by certain amounts, such that each statistic is ϵ' -differentially private in each iteration.

ϵ' -differentially private moment A by Laplace mechanism. For perturbing A , we use the Laplace mechanism. To use the Laplace mechanism, we first need to analyse the following $L1$ -sensitivity:

$$\begin{aligned} \Delta A &:= \max_{\mathcal{D}, \tilde{\mathcal{D}} \in \mathbb{N}^{|\mathcal{X}|}, \|\mathcal{D} - \tilde{\mathcal{D}}\|_1 = 1} \left\| \frac{1}{N} X^\top S \mathbf{y} - \frac{1}{N} \tilde{X}^\top \tilde{S} \tilde{\mathbf{y}} \right\|_1 = \max_{\mathbf{x}_k, \tilde{\mathbf{x}}_k} \left\| \frac{1}{N} \mathbf{x}_k s_k y_k - \frac{1}{N} \tilde{\mathbf{x}}_k \tilde{s}_k \tilde{y}_k \right\|_1, \\ &\leq \frac{1}{N} \sum_{l=1}^d |\mathbf{x}_{k,l} s_k y_k| + \frac{1}{N} \sum_{l=1}^d |\tilde{\mathbf{x}}_{k,l} \tilde{s}_k \tilde{y}_k|, \text{ triangle inequality} \\ &\leq \frac{s_k}{N} \sum_{l=1}^d |\mathbf{x}_{k,l}| + \frac{\tilde{s}_k}{N} \sum_{l=1}^d |\tilde{\mathbf{x}}_{k,l}|, \text{ since } |y_k| \leq 1 \text{ and } |\tilde{y}_k| \leq 1, \\ &\leq \frac{2\delta\sqrt{d}}{N}, \text{ since } s_k \leq \delta \text{ and } |\mathbf{x}_l|_1 \leq \sqrt{d}. \end{aligned} \quad (3)$$

Hence, the following Laplace mechanism produces ϵ' -differentially private moment of A :

$$\tilde{A} = A + (Y_1, \dots, Y_d), \quad (4)$$

where $Y_i \sim^{i.i.d.} \text{Laplace}(\frac{2\delta\sqrt{d}}{N\epsilon'})$.

ϵ' -differentially private moment A by Gaussian mechanism. One could perturb the first moment by using the Gaussian mechanism. To use the Gaussian mechanism, one needs to analyse the L2-sensitivity, which is $\Delta_2 A = 2\delta/N$ straightforwardly coming from Eq.(3)

$$\tilde{A} = A + (Y_1, \dots, Y_d), \quad (5)$$

where $Y_i \sim^{i.i.d.} \text{Gaussian}(0, \sigma^2)$, where $\sigma \geq c\Delta_2 A/\epsilon'$ for $c^2 \geq 2\log(1.25/\delta)$.

ϵ' -differentially private moment B . We perturb B by adding Wishart noise following [4], which provides strong privacy guarantees and significantly higher utility than other methods (e.g., [5–7]) as illustrated in [4] when perturbing positive definite matrices.

To draw Wishart noise, we first draw Gaussian random variables:

$$\mathbf{z}_i \sim \mathcal{N}\left(0, \frac{\delta}{2\epsilon'N} I_d\right), \text{ for } i = \{1, \dots, d+1\}, \quad (6)$$

and construct a matrix $Z := [\mathbf{z}_1, \dots, \mathbf{z}_{d+1}] \in \mathbb{R}^{d \times (d+1)}$

$$\tilde{B} := B + ZZ^\top \quad (7)$$

then \tilde{B} is a ϵ' -differentially private second moment matrix. Proof follows the paper [4]. The matrix ZZ^\top is a sample from a Wishart distribution $W(ZZ^\top | \frac{\delta}{2\epsilon'N} I_d, d+1)$. The probability ratio between a noised-up version \tilde{B} given a dataset \mathcal{D} (where B is the exact second moment matrix given \mathcal{D}) and given a neighbouring dataset \mathcal{D}' (where B' is the exact second moment matrix given \mathcal{D}') is given by

$$\frac{W(\tilde{B} - B | \frac{\delta}{2\epsilon'N} I_d, d+1)}{W(\tilde{B} - B' | \frac{\delta}{2\epsilon'N} I_d, d+1)} = \frac{\exp(-\frac{\epsilon'N}{\delta} \text{tr}(\tilde{B} - B))}{\exp(-\frac{\epsilon'N}{\delta} \text{tr}(\tilde{B} - B'))}, \quad (8)$$

$$= \exp\left(\frac{\epsilon'N}{\delta} \text{tr}(B - B')\right), \quad (9)$$

$$= \exp\left(\frac{\epsilon'N}{\delta} \frac{1}{N} \text{tr}(s_k \mathbf{x}_k \mathbf{x}_k^\top - \tilde{s}_k \tilde{\mathbf{x}}_k \tilde{\mathbf{x}}_k^\top)\right), \quad (10)$$

$$= \exp\left(\frac{\epsilon'}{\delta} (s_k \mathbf{x}_k^\top \mathbf{x}_k - \tilde{s}_k \tilde{\mathbf{x}}_k^\top \tilde{\mathbf{x}}_k)\right), \quad (11)$$

$$\leq \exp(\epsilon'), \text{ since } 0 \leq \mathbf{x}_k^\top \mathbf{x}_k \leq 1, \text{ and } 0 \leq s_k \leq \delta.$$

3 Concentrated differential privacy for IRLS

Here we adopt a relaxed version of DP, the so-called concentrated differential privacy (CDP) in order to significantly lower the amounts of noise to add to the moments without compromising on cumulative privacy loss over several iterations.

According to Theorem 3.5 in [8], any ϵ -DP algorithm is $(\epsilon(\exp(\epsilon) - 1)/2, \epsilon)$ -CDP. Furthermore, theorem 3.4 states that J-composition of (μ, τ) -CDP mechanism guarantees $(\sum_{i=1}^J \mu_i, \sqrt{\sum_{i=1}^J \tau_i^2})$ -CDP. Suppose we perturb some key statistic in each IRLS iteration using the Laplace mechanism.

Denote the difference in statistic given dataset x and y by $\Delta S := S(x) - S(y)$. The conventional composition theorem says that I should add $\text{Lap}(\Delta S J/\epsilon)$ in each iteration to ensure ϵ -DP after J iterations. Now suppose we perturb the key statistic in each iteration by adding Laplace noise drawn from $\text{Lap}(\Delta S/\epsilon')$, which, according to Theorem 3.5 in [8], gives us a $(\epsilon'(\exp(\epsilon') - 1)/2, \epsilon')$ -CDP solution. According to Theorem 3.4 in [8], after J iterations, we obtain a $(J\epsilon'(\exp(\epsilon') - 1)/2, \sqrt{J}\epsilon')$ -CDP solution. What we want to make sure is if the expected privacy loss is equal to our privacy budget ϵ , i.e., $J\epsilon'(\exp(\epsilon') - 1)/2 = \epsilon$. Using Taylor's expansion, we can rewrite the left hand side by $J\epsilon'(1 + \epsilon' + \sum_{j=2}^{\infty} \frac{\epsilon'^j}{j!} - 1)/2 = \epsilon$, which we can lower bound by ignoring the infinite sum, $J\epsilon'^2/2 \leq \epsilon$. Hence, the largest ϵ' should be less than equal to $\sqrt{2\epsilon/J}$.

This says, in each iteration, the key statistic should be perturbed by adding Laplace noise drawn from $\text{Lap}(\sqrt{J}\Delta S/\sqrt{2\epsilon})$, in order to obtain a $(\epsilon, \sqrt{2\epsilon})$ -CDP solution after J iterations. In the IRLS algorithm, we have two statistics to perturb in each iteration. Suppose we perturb each statistic to ensure ϵ' -DP. Then, we can modify the result above by replacing J with $2J$ for the IRLS algorithm. Hence, each perturbation should result in ϵ' -DP parameter, where $\epsilon' := \sqrt{\frac{2\epsilon}{2J}} = \sqrt{\frac{\epsilon}{J}}$. This gives us the ϵ -CDP IRLS algorithm below.

Algorithm 1 $(\epsilon, \sqrt{2\epsilon})$ -CDP IRLS algorithm via moment perturbation

Require: Dataset \mathcal{D}

Ensure: ϵ -IRDP least squares solution after J -iteration

- (1) Compute $A = \frac{1}{N}X^\top \mathbf{y}$ and add either Laplace or Gaussian noise by Eq.(4) or Eq.(5)
 - (2) Compute $B = \frac{1}{N}X^\top X$ and add Wishart noise by Eq.(7)
 - (3) Compute the ϵ -CDP least squares solution by $\boldsymbol{\theta}_{cdpirls} := \tilde{B}^{-1}\tilde{A}$.
-

4 Experiments

Our simulated dataset consists of N datapoints, each with d dimensional covariates, generated using *i.i.d.* draws $\mathbf{x}_i \sim \mathcal{N}(0, I_d)$, then normalise X such that the largest squared L2 norm is 1. We generated the true parameter $\boldsymbol{\theta} \in \mathbb{R}^d$ from $\mathcal{N}(0, I_d)$. We generated each observation y_i from $\mathcal{N}(X\boldsymbol{\theta}, \sigma^2 I)$, where $\sigma^2 = 0.01$. We also normalised Y such that the largest squared L2-norm is 1.

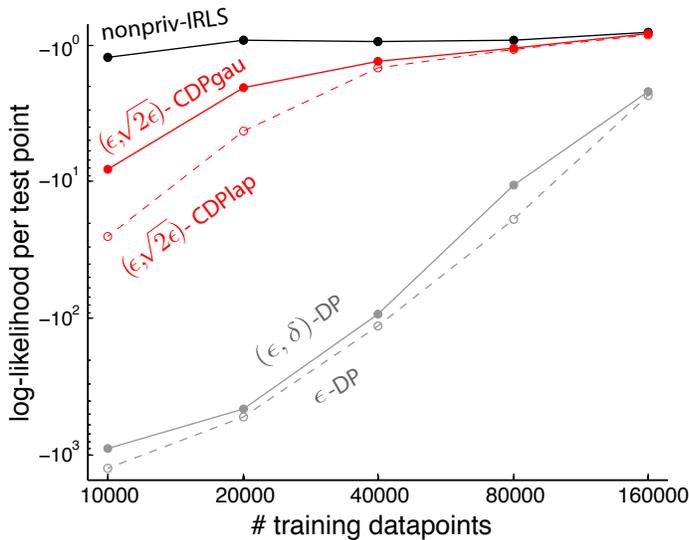


Figure 1: We tested $(\epsilon, \sqrt{2\epsilon})$ -CDP-IRLS (gau: Gaussian mechanism for mean perturbation, lap: Laplace mechanism for mean perturbation), ϵ -DP-IRLS (using the conventional composition theorem), and (ϵ, δ) -DP-IRLS (using the Advanced composition theorem) for $d = 10$ and $\epsilon = 0.9$ with varying N for which we generated 20 independent datasets. For each IRLS solution, we computed the log-likelihood of test data (10% of training data), then divided by the number of test points to show the log-likelihood per test point. CDP-IRLS requires significantly less data than DP-IRLS for the same level of expected privacy.

Acknowledgements

This work is supported by Qualcomm.

References

- [1] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9:211–407, August 2014.
- [2] R. Chartrand and Wotao Yin. Iteratively reweighted algorithms for compressive sensing. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3869–3872, March 2008.
- [3] Or Sheffet. Private approximations of the 2nd-moment matrix using existing techniques in linear regression. *CoRR*, abs/1507.00056, 2015.
- [4] Rutgers Hafiz Imtiaz, Anand D. Sarwate. Symmetric matrix perturbation for differentially-private principal component analysis. In *ICCASP*, 2016.
- [5] Cynthia Dwork, Kunal Talwar, Abhradeep Thakurta, and Li Zhang. Analyze gauss: optimal bounds for privacy-preserving principal component analysis. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 11–20, 2014.
- [6] Moritz Hardt and Eric Price. The noisy power method: A meta algorithm with applications. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2861–2869. Curran Associates, Inc., 2014.
- [7] Kamalika Chaudhuri, Anand Sarwate, and Kaushik Sinha. Near-optimal differentially private principal components. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 989–997. Curran Associates, Inc., 2012.
- [8] C. Dwork and G. N. Rothblum. Concentrated Differential Privacy. *ArXiv e-prints*, March 2016.