View Reviews

Paper ID
4673

Paper Title
Learning Privately over Distributed Features: An ADMM Sharing Approach

Reviewer #1

Questions

1. Contributions: Please list three things this paper contributes (e.g., theoretical, methodological, algorithmic, empirical contributions; bridging fields; or providing an important critical analysis). For each contribution, briefly state the level of significance (i.e., how much impact will this work have on researchers and practitioners in the future?). If you cannot think of three things, please explain why. Not all good papers will have three contributions.
1. This paper studies a new setting where the features are distributed and the goal is to learning models privately.
2. The authors propose a method called ADMM shared algorithm.
3. Using this algorithm, the authors extend to the private case.

2. Detailed comments: Please provide a thorough review of the submission, including its originality, quality, clarity, and significance. Hover over the "?" next to this prompt to see a brief description of these metrics.
Here are the reasons why I tend to reject the paper:
1. From the view of differential privacy, the method in the paper is not theoretically interesting. It is just add some noise to the non-private algorithm which is the most common technique method in DP community. Also, there is no theoretical guarantee, even if the loss function is convex.

2. It is still not interesting for me from the optimization view, although I am not a member of optimization. First, it needs quite a lot assumptions while the authors do not discuss about these assumptions. Are these reasonable assumptions? Also, what is the iteration complexity in Theorem 2? It is hard to understand.

3. Also from the DP view (I am a member in DP) . There are many papers study feature distributed setting or Private ADMM algorithms. However, the authors do not mention them. For example, for private feature distributed learning see [1-3], for (distributed) private ADMM see [3-9]. Why these methods cannot be modified to the problem in this paper?

[1] Uplink Communication Efficient Differentially Private Sparse Optimization with Feature-Wise Distributed Data. AAAI 2018
[2] Heinze-Deml, Christina, Brian McWilliams, and Nicolai Meinshausen. "Preserving Differential Privacy Between Features in Distributed Estimation." arXiv preprint arXiv:1703.00403 (2017).
[3] Optimal Differentially Private ADMM for Distributed Machine Learning
[4] On Privacy-preserving Decentralized Optimization through Alternating Direction Method of Multipliers
[5] Dynamic differential privacy for ADMM-based distributed classification learning
[6] Improving the privacy and accuracy of ADMM-based distributed algorithms.
[7] Recycled admm: Improve privacy and accuracy with less computation in distributed algorithms.
[8] ADMM based privacy- preserving decentralized optimization.
[9] DP-ADMM: ADMM-based Distributed Learning with Differential Privacy.

3. Please provide an "overall score" for this submission.
3: A clear reject. I vote and argue for rejecting this submission.

4. Please provide a "confidence score" for your assessment of this submission.
5: You are absolutely certain about your assessment. You are very familiar with the related work.

5. Improvements: What would the authors have to do for you to increase your score?
1. Explicit Theoretical Guarantee for the private algorithm (when the loss is convex). Most of the papers I mentioned have theoretical guarantee (error analysis with time complexity).
2. Some non-trivial private algorithms.

Reviewer #4

Questions

1. Contributions: Please list three things this paper contributes (e.g., theoretical, methodological, algorithmic, empirical contributions; bridging fields; or providing an important critical analysis). For each contribution, briefly state the level of significance (i.e., how much impact will this work have on researchers and practitioners in the future?). If you cannot think of three things, please explain why. Not all good papers will have three contributions.
- Convergence analysis of ADMM sharing algorithm for linear models with nonlinear loss: low significance
- Private version of ADMM sharing: low significance

2. Detailed comments: Please provide a thorough review of the submission, including its originality, quality, clarity, and significance. Hover over the "?" next to this prompt to see a brief description of these metrics.
This paper deals with the problem of learning over features that are distributed across several parties under privacy constraints. To this end, the authors rely on a classic ADMM sharing algorithm in the context of linear models.

Their first contribution is to analyze the convergence of this algorithm for parallel updates of the model blocks and under non-convex loss. The analysis closely follows that of [18] for sequential updates. I am not very convinced of the relevance of this extension for machine learning. Indeed, the convergence analysis in the case of parallel updates and convex loss is well-studied (see [40]), and using a non-convex loss function together with linear models is not very common. As a matter of fact, the experiments are based on logistic regression, for which the loss is convex and the convergence analysis was already known.

The second contribution is a private version of ADMM sharing. This is achieved by adding Gaussian noise to the solution of the augmented Lagrangian subproblem (6). The authors provide some bound on the sensitivity which allows to scale the variance of the noise to ensure DP. However the scale relies on some boundedness assumptions: in particular Assumption 2.1 for which it seems tricky to find a tight bound in many practical situations. The authors do not explain how they tackle this in practice. A clear illustration of this problem is the fact that in their experiments the authors do not give any (epsilon,delta)-DP guarantee but only mention the standard deviation of the noise. How can one figure out the corresponding DP guarantee?
Maybe the authors should look into other techniques to ensure DP for which the noise variance does not depend on quantities that are difficult to bound tightly, or investigate the use of clipping. For instance, the ADMM objective perturbation mechanism proposed by

[5] could potentially be extended to the present context.

Another key limitation of this paper is the lack of utility analysis for the private version. There is no result that characterizes the utility loss compared to the non-private version (in fact, it is not shown whether the private algorithm converges at all).

Other comments:
- In related work on learning over distributed features, the authors could also mention Frank-Wolfe algorithms, see e.g.
Bellet et al. A Distributed Frank-Wolfe Algorithm for Communication-Efficient Sparse Learning. SDM 2015
Moharrer et al. Distributing Frank-Wolfe via Map-Reduce. IJCAI 2018
- In the experiments, it would be nice to also see accuracy to measure the impact of the noise
- In Figure 1a, the algorithm seems to diverge?
- In the local features baseline, it would be better to average the loss/accuracy of the local model over each party instead of picking a single one arbitrarily.

3. Please provide an "overall score" for this submission.
3: A clear reject. I vote and argue for rejecting this submission.

4. Please provide a "confidence score" for your assessment of this submission.
5: You are absolutely certain about your assessment. You are very familiar with the related work.

5. Improvements: What would the authors have to do for you to increase your score?
- More practical mechanisms to ensure concrete (eps,delta)-differential privacy guarantees
- Utility analysis for the private algorithm

Reviewer #5

Questions

1. Contributions: Please list three things this paper contributes (e.g., theoretical, methodological, algorithmic, empirical contributions; bridging fields; or providing an important critical analysis). For each contribution, briefly state the level of significance (i.e., how much impact will this work have on researchers and practitioners in the future?). If you cannot think of three things, please explain why. Not all good papers will have three contributions.

1) Authors propose a new ADMM algorithm to solve the empirical risk minimization problem for vertically partitioned data.

2) Theoretical convergence guarantees and iteration complexity results for non-convex loss functions in the non differentially private setting.

3) A privacy preserving version of the algorithm. A privacy analysis is provided.

2. Detailed comments: Please provide a thorough review of the submission, including its originality, quality, clarity, and significance. Hover over the "?" next to this prompt to see a brief description of these metrics.

Strengths

An extensive convergence analysis for the non-DP version of the algorithm based on techniques of reference [22]. The difference to the analysis of [22] seems to be significant (see e.g. Assumption 1 of this article and Assumption A of reference [22]).

A differentially private version of the algorithm via Gaussian mechanism.

Weaknesses

There seems to be some redundancy in the presentation: In comparison to the non-DP version (equations 6 to 8) the private version has simply the Gaussian noise adding (second and third lines of equation (16)).

The proof of Lemma 1 (sensitivity bound) is missing, I could not find the appendix either.

The privacy accounting part seems to be the weakest part of the article. It is even omitted in the experiments. The accuracy is compared to the amount of noise added and there is no accounting of epsilon and/or delta. The privacy analysis (Theorem 3 and Corollary 1) is superficial, the results are not used. There should be a reference for Theorem 3. Moreover, much better bounds exist for the Gaussian mechanism, see e.g.

Balle, Borja, and Yu-Xiang Wang. "Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising." arXiv preprint arXiv:1805.06530 (2018)

My biggest concern is the validity of the experiments. In order to compute the epsilon and delta values, the sensitivity should be determined first. This would require determining explicitly the constants d_m, \lambda, c_1, M, \rho and b_1 given in Lemma 1. If, for example, M (the number of vertical partitions) is large, then by Theorem 3 the required noise parameter sigma for a reasonable epsilon might be much larger than the values 0.3 and 1.0 (Figure 3), where the ADMM with DP noise seems to break down. In other words, it would be important to determine what are the delta,epsilon values at these deviation points.

Originality

It seems to me that the originality of the method lies in the convergence proof of the non-DP version of the algorithm. It remains unclear what is the contribution of the DP version as no privacy accounting is actually carried out.

3. Please provide an "overall score" for this submission.
4: An okay submission, but not good enough; a reject. I vote for rejecting this submission, although I would not be upset if it were accepted.

4. Please provide a "confidence score" for your assessment of this submission.
3: You are fairly confident in your assessment. It is possible that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work. Math/other details were not carefully checked.

5. Improvements: What would the authors have to do for you to increase your score?

A privacy accounting should be added. This means determining the parameters of Lemma 1 explicitly. Add privacy accounting to the experiments.

Add the proof of Lemma 1.

If you could modify the non-DP convergence analysis to the DP version, this would make the article much stronger and clearer.

Consider more recent privacy accounting methods for the Gaussian mechanism than that of Theorem 3.

Small:
Fix Definition 1 (Differential privacy definition), the set S is missing in the inequality.

The fonts for the adjacent data sets should be the same in the appendix (D and mathcal(D)).