Privacy-preserving Stacking with Application to Cross-organizational Diabetes Prediction

Quanming Yao[1,2], Xiawei Guo[1], James Kwok[2], Weiwei Tu[1], Yuqiang Chen[1], Wenyuan Dai[1] and Qiang Yang[2]

[1]4Paradigm Inc

[2]Department of Computer Science and Engineering, HKUST

Abstract

To meet the standard of differential privacy, noise is usually added into the original data, which inevitably deteriorates the predicting performance of subsequent learning algorithms. In this paper, motivated by the success of improving predicting performance by ensemble learning, we propose to enhance privacy-preserving logistic regression by stacking. We show that this can be done either by sample-based or feature-based partitioning. However, we prove that when privacy-budgets are the same, feature-based partitioning requires fewer samples than sample-based one, and thus likely has better empirical performance. As transfer learning is difficult to be integrated with a differential privacy guarantee, we further combine the proposed method with hypothesis transfer learning to address the problem of learning across different organizations. Finally, we not only demonstrate the effectiveness of our method on two benchmark data sets, i.e., MNIST and NEWS20, but also apply it into a real application of cross-organizational diabetes prediction from RUIJIN data set, where privacy is of a significant concern[1].

1 Introduction

In recent years, data privacy has become a serious concern in both academia and industry [Dwork et al., 2006; Chaudhuri et al., 2011; Dwork and Roth, 2014; Abadi et al., 2016]. There are now privacy laws, such as Europe's General Data Protection Regulation (GDPR), which regulates the protection of private data and restricts data transmission between organizations. These raise challenges for cross-organizational machine learning [Pathak et al., 2010; Hamm et al., 2016; Papernot et al., 2017; Xie et al., 2017], in which data have to be distributed to different organizations, and the learning model needs to make predictions in private.

A number of approaches have been proposed to ensure privacy protection. In machine learning, differential privacy [Dwork and Roth, 2014] is often used to allow data be exchanged among organizations. To design a differentially private algorithm, carefully designed noise is usually added to the original data to disambiguate the algorithms. Many standard learning algorithms have been extended for differential privacy. These include logistic regression [Chaudhuri et al., 2011], trees [Emekçi et al., 2007; Fong and Weber-Jahnke, 2012], and deep networks [Shokri and Shmatikov, 2015; Abadi et al., 2016]. In particular, linear models are simple and easy to understand, and their differentially private variants (such as privacy-preserving logistic regression (PLR)) [Chaudhuri et al., 2011]) have rigorous theoretical guarantees [Chaudhuri et al., 2011; Bassily et al., 2014; Hamm et al., 2016; Kasiviswanathan and Jin, 2016]. However, the injection of noise often degrades prediction performance.

Ensemble learning can often significantly improve the performance of a single learning model [Zhou, 2012]. Popular examples include bagging [Breiman, 1996a], boosting [Friedman et al., 2000], and stacking [Wolpert, 1992]. These motivate us to develop an ensemble-based method which can benefit from data protection, while enjoying good prediction performance. Bagging and boosting are based on partitioning of training samples, and use pre-defined rules (majority or weighted voting) to combine predictions from models trained on different partitions. Bagging improves learning performance by reducing the variance. Boosting, on the other hand, is useful in converting weak models to a strong one. However, the logistic regression model, which is the focus in this paper, often has good performance in many applications, and is a relatively strong classifier. Besides, it is a convex model and relatively stable.

Thus, in this paper, we focus on stacking. While stacking also partitions the training data, this can be based on either samples [Breiman, 1996b; Smyth and Wolpert, 1999; Ozay and Vural, 2012] or features [Boyd et al., 2011]. Multiple low-level models are then learned on the different data partitions, and a high-level model (typically, a logistic regression model) is used to combine their predictions. By combining with PLR, we show how differential privacy can be ensured in stacking. Besides, when the importance of features is known a priori, they can be easily incorporated in feature-based partitioning. We further analyze the learning guarantee of sample-based and feature-based stacking, and show theoretically that feature-based partitioning can have lower sample complexity (than sample-based partitioning), and thus better performance. By adapting the feature impor-

[1]Correspondence to X. Guo at guoxiawei@4paradigm.com

References

[Abadi et al., 2016] M. Abadi, A. Chu, I. Goodfellow, B. McMahan, I. Mironov, K. Talwar, and L. Zhang. Deep learning with differential privacy. In SIGSAC, pages 308–318. ACM, 2016.

[Bassily et al., 2014] R. Bassily, A. Smith, and A. Thakurta. Private empirical risk minimization: Efficient algorithms and tight error bounds. In FOCS, pages 464–473. IEEE, 2014.

[Boyd et al., 2011] S. Boyd, N. Parikh, and E. Chu. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine learning, 3(1):1–122, 2011.

[Breiman, 1996a] L. Breiman. Bagging predictors. ML, 24(2):123–140, 1996.

[Breiman, 1996b] L. Breiman. Stacked regressions. ML, 24(1):49–64, 1996.

[Chaudhuri et al., 2011] K. Chaudhuri, C. Monteleoni, and A. Sarwate. Differentially private empirical risk minimization. JMLR, 12(Mar):1069–1109, 2011.

[Dwork and Roth, 2014] C. Dwork and A. Roth. The algorithmic foundations of differential privacy. Foundations and Trends® in Theoretical Computer Science, 9(3–4):211–407, 2014.

[Dwork et al., 2006] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In TCC, pages 265–284. Springer, 2006.

[Dzeroski and Zenko, 2004] S. Dzeroski and B. Zenko. Is combining classifiers with stacking better than selecting the best one? ML, 54(3):255–273, 2004.

[Emekçi et al., 2007] F. Emekçi, O. Sahin, D. Agrawal, and A. El Abbadi. Privacy preserving decision tree learning over multiple parties. TKDE, 63(2):348–361, 2007.

[Fong and Weber-Jahnke, 2012] P. Fong and J. Weber-Jahnke. Privacy preserving decision tree learning using unrealized data sets. TKDE, 24(2):353–364, 2012.

[Friedman et al., 2000] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. Annals of Statistics, 28(2):337–407, 2000.

[Friedman et al., 2012] J. Friedman, T. Hastie, and R. Tibshirani. The elements of statistical learning. Springer, 2012.

[Hamm et al., 2016] J. Hamm, Y. Cao, and M. Belkin. Learning privately from multiparty data. In ICML, pages 555–563, 2016.

[Hanley and McNeil, 1983] J. Hanley and B. McNeil. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology, 148(3):839–843, 1983.

[Kasiviswanathan and Jin, 2016] P. Kasiviswanathan and H. Jin. Efficient private empirical risk minimization for high-dimensional learning. In ICML, pages 488–497, 2016.

[Kifer et al., 2012] D. Kifer, A. Smith, and A. Thakurta. Private convex empirical risk minimization and high-dimensional regression. JMLR, 23, 2012.

[Kuzborskij and Orabona, 2013] I. Kuzborskij and F. Orabona. Stability and hypothesis transfer learning. In ICML, pages 942–950, 2013.

[Lang, 1995] Ken Lang. Newsweeder: Learning to filter netnews. In ICML. Citeseer, 1995.

[LeCun et al., 1998] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, 1998.

[Ozay and Vural, 2012] M. Ozay and F. Vural. A new fuzzy stacked generalization technique and analysis of its performance. Technical report, arXiv:1204.0171, 2012.

[Pan and Yang, 2010] J. Pan and Q. Yang. A survey on transfer learning. TKDE, 22(10):1345–1359, 2010.

[Papernot et al., 2017] N. Papernot, M. Abadi, U. Erlingsson, I. Goodfellow, and K. Talwar. Semi-supervised knowledge transfer for deep learning from private training data. In ICLR, 2017.

[Pathak et al., 2010] M. Pathak, S. Rane, and B. Raj. Multiparty differential privacy via aggregation of locally trained classifiers. In NIPS, pages 1876–1884, 2010.

[Rajkumar and Agarwal, 2012] A. Rajkumar and S. Agarwal. A differentially private stochastic gradient descent algorithm for multiparty classification. In AISTAT, pages 933–941, 2012.

[Shalev-Shwartz and Srebro, 2008] S. Shalev-Shwartz and N. Srebro. SVM optimization: inverse dependence on training set size. In ICML, pages 928–935. ACM, 2008.

[Shokri and Shmatikov, 2015] R. Shokri and V. Shmatikov. Privacy-preserving deep learning. In SIGSAC, pages 1310–1321, 2015.

[Smyth and Wolpert, 1999] P. Smyth and D. Wolpert. Linearly combining density estimators via stacking. ML, 36(1-2):59–83, 1999.

[Ting and Witten, 1999] K. Ting and I. Witten. Issues in stacked generalization. JAIR, 10:271–289, 1999.

[Wang et al., 2018] Y. Wang, Q. Gu, and D. Brown. Differentially private hypothesis transfer learning. In ECML, 2018.

[Wolpert, 1992] D. Wolpert. Stacked generalization. Neural Networks, 5(2):241–259, 1992.

[Xie et al., 2017] L. Xie, I. Baytas, K. Lin, and J. Zhou. Privacy-preserving distributed multi-task learning with asynchronous updates. In SIGKDD, pages 1195–1204, 2017.

[Zhou, 2012] Z.-H. Zhou. Ensemble methods: foundations and algorithms. Chapman and Hall/CRC, 2012.