

Huang, Lingxiao; Jin, Yifei; Li, Jian

SVM via saddle point optimization: new bounds and distributed algorithms. (English)

Zbl 07238980

Eppstein, David (ed.), 16th Scandinavian symposium and workshops on algorithm theory. SWAT 2018, June 18–20, 2018, Malmö University, Malmö, Sweden. Wadern: Schloss Dagstuhl – Leibniz Zentrum für Informatik (ISBN 978-3-95977-068-2). LIPIcs – Leibniz International Proceedings in Informatics 101, Article 25, 13 p. (2018).

Summary: We study two important SVM variants: hard-margin SVM (for linearly separable cases) and ν -SVM (for linearly non-separable cases). We propose new algorithms from the perspective of saddle point optimization. Our algorithms achieve $(1 - \epsilon)$ -approximations with running time $\tilde{O}(nd + n\sqrt{d/\epsilon})$ for both variants, where n is the number of points and d is the dimensionality. To the best of our knowledge, the current best algorithm for ν -SVM is based on quadratic programming approach which requires $\Omega(n^2d)$ time in worst case [T. Joachims, Making large-scale SVM learning practical. Techn. Rep., Technische Universität Dortmund. (1998; doi:10.17877/DE290R-5098); J.C. Platt, “Fast training of support vector machines using sequential minimal optimization”, in: Advances in kernel methods: support vector learning. Cambridge: MIT Press. 185–208 (1999)]. In the paper, we provide the first nearly linear time algorithm for ν -SVM. The current best algorithm for hard margin SVM achieved by Gilbert algorithm [B. Gärtner and M. Jaggi, in: Proceedings of the 25th annual symposium on computational geometry, SCG 2009, Aarhus, Denmark, June 8–10, 2009. New York, NY: Association for Computing Machinery (ACM). 33–42 (2009; Zbl 1380.68396)] requires $O(nd/\epsilon)$ time. Our algorithm improves the running time by a factor of $\sqrt{d}/\sqrt{\epsilon}$. Moreover, our algorithms can be implemented in the distributed settings naturally. We prove that our algorithms require $\tilde{O}(k(d + \sqrt{d/\epsilon}))$ communication cost, where k is the number of clients, which almost matches the theoretical lower bound. Numerical experiments support our theory and show that our algorithms converge faster on high dimensional, large and dense data sets, as compared to previous methods.

For the entire collection see [Zbl 1390.68019].

MSC:

68Wxx Algorithms in computer science

Keywords:

ν -SVM; hard-margin SVM; saddle point optimization; distributed algorithm

Software:

Pegasos; Scikit; SVMlight

Full Text: DOI

References:

- [1] Nir Ailon and Bernard Chazelle. Faster dimension reduction. *Communications of the ACM*, 53(2):97-104, 2010.
- [2] Zeyuan Allen-Zhu. Katyusha: The first direct acceleration of stochastic gradient methods. *STOC 16*, 2016.
- [3] Zeyuan Allen-Zhu, Zhenyu Liao, and Yang Yuan. Optimization algorithms for faster computational geometry. In *LIPIcs*, volume 55, 2016.
- [4] Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(1):121-164, 2012.
- [5] Kristin P Bennett and Erin J Bredensteiner. Duality and geometry in svm classifiers. In *ICML*, pages 57-64, 2000.
- [6] Marshall W. Bern and David Eppstein. Optimization over zonotopes and training support vector machines. In *WADS*, pages 111-121, 2001.
- [7] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144-152. ACM, 1992.

- [8] Kenneth L Clarkson, Elad Hazan, and David P Woodruff. Sublinear optimization for machine learning. *{\it Journal of the ACM (JACM)}*, 59(5):23, 2012.
- [9] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *{\it Machine learning}*, 20(3):273-297, 1995.
- [10] DJ Crisp and CJC Burges. A geometry interpretation of μ -svm classifiers. *{\it NIPS}*, pages 244-251, 2000.
- [11] John Duchi and Yoram Singer. Efficient online and batch learning using forward backward splitting. *{\it JMLR}*, 10(Dec):2899-2934, 2009.
- [12] Pedro A Forero, Alfonso Cano, and Georgios B Giannakis. Consensus-based distributed support vector machines. *{\it JMLR}*, 11(May):1663-1707, 2010.
- [13] Vojtěch Franc and Soeren Sonnenburg. Optimized cutting plane algorithm for support vector machines. In *{\it ICML 08}*, pages 320-327. ACM, 2008.
- [14] Alan Frieze, Ravi Kannan, and Santosh Vempala. Fast monte-carlo algorithms for finding low-rank approximations. *{\it Journal of the ACM (JACM)}*, 51(6):1025-1041, 2004.
- [15] Bernd Gärtner and Martin Jaggi. Coresets for polytope distance. In *{\it SOCG 09}*, pages 33-42. ACM, 2009.
- [16] Elmer G Gilbert. An iterative procedure for computing the minimum of a quadratic form on a convex set. *{\it SIAM Journal on Control}*, 4(1):61-80, 1966.
- [17] Hans Peter Graf, Eric Cosatto, Leon Bottou, Igor Durdanovic, and Vladimir Vapnik. Parallel support vector machines: The cascade svm. In *{\it NIPS}*, volume 17, 2004.
- [18] Sarel Har-Peled, Dan Roth, and Dav Zimak. Maximum margin coresets for active and noise tolerant learning. In *{\it IJCAI}*, pages 836-841, 2007.
- [19] Elad Hazan, Tomer Koren, and Nati Srebro. Beating sgd: Learning svms in sublinear time. In *{\it Advances in Neural Information Processing Systems}*, pages 1233-1241, 2011.
- [20] Cho-Jui Hsieh, Kai-Wei Chang, Chih-Jen Lin, S Sathiy Keerthi, and Sellamanickam Sundararajan. A dual coordinate descent method for large-scale linear svm. In *{\it ICML}* *{\it 08}*, pages 408-415. ACM, 2008.
- [21] Thorsten Joachims. Making large-scale svm learning practical. Technical report, Technical Report, SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund, 1998.
- [22] Thorsten Joachims. Training linear svms in linear time. In *{\it Proceedings of the 12th ACM}* *{\it SIGKDD international conference on Knowledge discovery and data mining}*, pages 217-226. ACM, 2006.
- [23] Anatoli Juditsky, Fatma Kılınç Karzan, and Arkadi Nemirovski. Randomized first order algorithms with applications to ℓ_1 -minimization. *{\it Mathematical Programming}*, 142(1-2):269-310, 2013.
- [24] Jyrki Kivinen, Alexander J Smola, and Robert C Williamson. Online learning with kernels. *{\it IEEE transactions on signal processing}*, 52(8):2165-2176, 2004.
- [25] Eyal Kushilevitz. Communication complexity. *{\it Advances in Computers}*, 44:331-360, 1997.
- [26] :12
- [27] :13
- [28] Yangwei Liu, Hu Ding, Ziyun Huang, and Jinhui Xu. Distributed and robust support vector machine. In *{\it LIPIcs}*, volume 64, 2016.
- [29] Jorge López and José R Dorronsoro. Linear convergence rate for the mdm algorithm for the nearest point problem. *{\it Pattern Recognition}*, 48(4):1510-1522, 2015.
- [30] Yumao Lu, Vwani Roychowdhury, and Lieven Vandenberghe. Distributed parallel support vector machines in strongly connected networks. *{\it IEEE Transactions on Neural Networks}*, 19(7):1167-1178, 2008.
- [31] BF Mitchell, Vladimir Fedorovich Dem'yanov, and VN Malozemov. Finding the point of a polyhedron closest to the origin. *{\it SIAM Journal on Control}*, 12(1):19-26, 1974.
- [32] A Navia-Vazquez, D Gutierrez-Gonzalez, Emilio Parrado-Hernández, and JJ NavarroAbellan. Distributed support vector machines. *{\it IEEE Trans. Neural Networks}*, 17(4):1091-1097, 2006.
- [33] Yu Nesterov. Excessive gap technique in nonsmooth convex minimization. *{\it SIAM Journal}* *{\it on Optimization}*, 16(1):235-249, 2005.
- [34] Alon Orlitsky and Abbas El Gamal. Average and randomized communication complexity. *{\it IEEE Transactions on Information Theory}*, 36(1):3-16, 1990.
- [35] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *{\it JMLR}*, 12(Oct):2825-2830, 2011.
- [36] John C Platt. 12 fast training of support vector machines using sequential minimal optimization. *{\it Advances in kernel methods}*, pages 185-208, 1999.
- [37] Bernhard Schölkopf, Alex J Smola, Robert C Williamson, and Peter L Bartlett. New support vector algorithms. *{\it Neural computation}*, 12(5):1207-1245, 2000.
- [38] Shai Shalev-Shwartz, Yoram Singer, Nathan Srebro, and Andrew Cotter. Pegasos: primal estimated sub-gradient solver for SVM. *{\it Math. Program.}*, 127(1):3-30, 2011.
- [39] Alexander J Smola, SVN Vishwanathan, Quoc V Le, et al. Bundle methods for machine learning. In *{\it NIPS}*, volume 20, pages 1377-1384, 2007.

- [40] Ivor W Tsang, Andras Kocsor, and James T Kwok. Simpler core vector machines with enclosing balls. In *{\it ICML 07}*, pages 911-918. ACM, 2007.
- [41] Ivor W Tsang, James T Kwok, and Pak-Ming Cheung. Core vector machines: Fast svm training on very large data sets. *{\it JMLR}*, 6(Apr):363-392, 2005.
- [42] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *{\it STOC 79}*, pages 209-213. ACM, 1979.
- [43] Caoxie Zhang, Honglak Lee, and Kang G Shin. Efficient distributed linear classification algorithms via the alternating direction method of multipliers. In *{\it Artificial Intelligence and} {\it Statistics}*, pages 1398-1406, 2012.
- [44] Yuchen Zhang and Xiao Lin. Stochastic primal-dual coordinate method for regularized empirical risk minimization. In *{\it ICML}*, pages 353-361, 2015.
- [45] Ji Zhu, Saharon Rosset, Robert Tibshirani, and Trevor J Hastie. 1-norm support vector machines. In *{\it NIPS}*, pages 49-56, 2004.

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.