

Le Thi, Hoai An; Vo, Xuan Thanh; Pham Dinh, Tao

Feature selection for linear SVMs under uncertain data: robust optimization based on difference of convex functions algorithms. (English) [Zbl 1327.90236](#)
Neural Netw. 59, 36-50 (2014).

Summary: In this paper, we consider the problem of feature selection for linear SVMs on uncertain data that is inherently prevalent in almost all datasets. Using principles of Robust Optimization, we propose robust schemes to handle data with ellipsoidal model and box model of uncertainty. The difficulty in treating ℓ_0 -norm in feature selection problem is overcome by using appropriate approximations and Difference of Convex functions (DC) programming and DC Algorithms (DCA). The computational results show that the proposed robust optimization approaches are superior than a traditional approach in immunizing perturbation of the data.

MSC:

90C26 Nonconvex programming, global optimization

68T05 Learning and adaptive systems in artificial intelligence

Cited in 14 Documents

Keywords:

feature selection; svm; robust optimization; DC programming; dca

Software:

UCI-ml

Full Text: [DOI](#)

References:

- [1] Ben-Tal, A.; El Ghaoui, L.; Nemirovski, A., Robust optimization, (2009), Princeton University Press Princeton, NJ · [Zbl 1221.90001](#)
- [2] Bhattacharyya, C.; Grate, L. R.; Jordan, M. I.; Ghaoui, L. E.; Mian, I. S., Robust sparse hyperplane classifier: application to uncertain molecular profiling data, *Journal of Computational Biology*, 11, 1073-1089, (2004)
- [3] Bhattacharyya, C.; Pannagadatta, K. S.; Smola, A. J., A second order cone programming formulation for classifying missing data, (Saul, L. K.; Weiss, Y.; Bottou, L., *Advances in neural information processing systems*, Vol. 17, NIPS17, (2004), MIT Press Cambridge, MA)
- [4] Bi, J.; Zhang, T., Support vector classification with input data uncertainty, (Saul, L. K.; Weiss, Y.; Bottou, L., *Advances in neural information processing systems*, Vol. 17, NIPS17, (2004), MIT Press Cambridge, MA)
- [5] Bradley, P. S.; Magasarian, O. L., Feature selection via concave minimization and support vector machines, (Shavlik, J., *Machine learning proceedings of the fifteenth international conference, ICML'98*, (1998), Morgan Kaufmann), 82-90
- [6] Caramanis, C.; Mannor, S.; Xu, H., Optimization for machine learning, (Sra, S.; Nowozin, S.; Wright, S., *Robust optimization in machine learning*, (2011), MIT Press Cambridge, MA)
- [7] Carroll, R.; Stefanski, L. A.; Ruppert, D.; Crainiceanu, C. M., Measurement error in nonlinear models, (2006), Chapman and Hall
- [8] Chambolle, A.; DeVore, R. A.; Lee, N. Y.; Lucier, B. J., Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage, *IEEE Transactions on Image Processing*, 7, 319-335, (1998) · [Zbl 0993.94507](#)
- [9] Collobert, R.; Sinz, F.; Weston, J.; Bottou, L., Large scale transductive svms, *Journal of Machine Learning Research*, 7, 1687-1712, (2006) · [Zbl 1222.68173](#)
- [10] Collobert, R.; Sinz, F.; Weston, J.; Bottou, L., Trading convexity for scalability, (Proceedings of the 23rd international conference on machine learning, ICML'06, (2006), ACM New York, NY, USA), 201-208
- [11] Dempster, A. P.; Laird, N. M.; Rubin, D. B., Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 39, 1-38, (1977) · [Zbl 0364.62022](#)
- [12] Eves, H. W., Elementary matrix theory, (1966), Allyn and Bacon, Inc. Boston · [Zbl 0136.24706](#)
- [13] Fan, J.; Li, R., Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association*, (2001) · [Zbl 1073.62547](#)
- [14] Frank, A., & Asuncion, A. (2010). UCI machine learning repository <http://archive.ics.uci.edu/ml>. Irvine, CA: University of California, School of Information and Computer Science.
- [15] Fu, W. J., Penalized regression: the bridge versus the lasso, *Journal of Computational and Graphical Statistics*, 7, 397-416,

(1998)

- [16] Golub, T. R.; Slonim, D. K.; Tamayo, P.; Huard, C.; Gaasenbeek, M.; Mesirov, J. P., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, *Science*, 286, 531-537, (1999)
- [17] Gordon, G. J.; Jensen, R. V.; Hsiao, L.-L.; Gullans, S. R.; Blumenstock, J. E.; Ramaswamy, S., Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma, *Cancer Research*, 62, (2002)
- [18] Hermes, L., & Buhmann, J.M. (2000). Feature selection for support vector machines. In \textit{Proceedings. 15th international conference on pattern recognition, Vol. 2} (pp. 712-715).
- [19] Krause, N.; Singer, Y., Leveraging the margin more carefully, (Proceedings of the twenty-first international conference on machine learning, ICML'04, (2004), ACM New York, NY, USA), 63-70
- [20] Le Hoai, M.; Le Thi, H. A.; Pham Dinh, T.; Huynh, V. N., Block clustering based on difference of convex functions (DC) programming and DC algorithms, *Neural Computation*, 25, 259-278, (2013)
- [21] Le Thi, H. A.; Belghiti, T.; Pham Dinh, T., A new efficient algorithm based on DC programming and DCA for clustering, *Journal of Global Optimization*, 37, 593-608, (2006) · [Zbl 1198.90327](#)
- [22] Le Thi, H. A.; Le Hoai, M.; Nguyen, V. V.; Pham Dinh, T., A DC programming approach for feature selection in support vector machines learning, *Advances in Data Analysis and Classification*, 2, 259-278, (2008) · [Zbl 1284.90057](#)
- [23] Le Thi, H. A.; Le Hoai, M.; Pham Dinh, T., Optimization based DC programming and DCA for hierarchical clustering, *European Journal of Operational Research*, 183, 1067-1085, (2007) · [Zbl 1149.90117](#)
- [24] Le Thi, H. A.; Le Hoai, M.; Pham Dinh, T., New and efficient DCA based algorithms for minimum sum-of-squares clustering, *Pattern Recognition*, 47, 388-401, (2014) · [Zbl 1326.68225](#)
- [25] Le Thi, H. A.; Pham Dinh, T., Solving a class of linearly constrained indefinite quadratic problems by DC algorithms, *Journal of Global Optimization*, 11, 2776-2807, (1997)
- [26] Le Thi, H. A.; Pham Dinh, T., Large scale global molecular optimization from exact distance matrices by a D.C. optimization approach, *SIAM Journal on Optimization*, 14, 77-114, (2003) · [Zbl 1075.90071](#)
- [27] Le Thi, H. A.; Pham Dinh, T., The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems, *Annals of Operations Research*, 133, 23-46, (2005) · [Zbl 1116.90122](#)
- [28] Le Thi, H. A.; Vo, X. T.; Pham Dinh, T., Robust feature selection for svms under uncertain data, (Perner, P., *Advances in data mining. applications and theoretical aspects*, Lecture Notes in Computer Science, vol. 7987, (2013), Springer Berlin, Heidelberg), 151-165
- [29] Liu, Y.; Shen, X., Multicategory ψ -learning, *Journal of the American Statistical Association*, 101, 500-509, (2006) · [Zbl 1119.62341](#)
- [30] Liu, Y.; Shen, X.; Doss, H., Multicategory ψ -learning and support vector machine: computational tools, *Journal of Computational and Graphical Statistics*, 14, 219-236, (2005)
- [31] Liu, Y.; Wu, Y., Variable selection via a combination of the l_0 and l_1 penalties, *Journal of Computational and Graphical Statistics*, 16, 782-798, (2007)
- [32] Neumann, J.; Schnörr, C.; Steidl, G., Combined SVM-based feature selection and classification, *Machine Learning*, 61, 129-150, (2005) · [Zbl 1137.90643](#)
- [33] Ong, C. S.; Le Thi, H. A., Learning sparse classifiers with difference of convex functions algorithms, *Optimization Methods & Software*, 28, 830-854, (2013) · [Zbl 1282.90181](#)
- [34] Pant, R.; Trafalis, T. B.; Barker, K., Support vector machine classification of uncertain and imbalanced data using robust optimization, (Proceedings of the 15th WSEAS international conference on computers, (2011), World Scientific and Engineering Academy and Society (WSEAS) Stevens Point, Wisconsin, USA), 369-374
- [35] Peleg, D.; Meir, R., A bilinear formulation for vector sparsity optimization, *Signal Processing*, 8, 375-389, (2008) · [Zbl 1186.94273](#)
- [36] Pham Dinh, T.; Le Thi, H. A., Convex analysis approach to DC programming: theory, algorithms and applications, *Acta Mathematica Vietnamica*, 22, 289-357, (1997)
- [37] Pham Dinh, T.; Le Thi, H. A., A DC optimization algorithms for solving the trust-region subproblem, *SIAM Journal on Optimization*, 8, 476-505, (1998) · [Zbl 0913.65054](#)
- [38] Pham Dinh, T.; Le Thi, H. A., Recent advances in DC programming and DCA, *Transactions on Computational Collective Intelligence*, 8342, 1-37, (2014)
- [39] Rakotomamonjy, A., Variable selection using SVM-based criteria, *Journal of Machine Learning Research*, 3, 1357-1370, (2003) · [Zbl 1102.68583](#)
- [40] Rao, B. D.; Kreutz-Delgado, K., An affine scaling methodology for best basis selection, *IEEE Transactions on Signal Processing*, 47, 187-200, (1999) · [Zbl 0984.94010](#)
- [41] Rinaldi, F., Mathematical programming methods for minimizing the zero norm over polyhedral sets, (2009), Operations Research at Sapienza University of Rome, (Ph.D. thesis)
- [42] Shivaswamy, P. K.; Bhattacharyya, C.; Smola, A. J., Second order cone programming approaches for handling missing and uncertain data, *Journal of Machine Learning Research*, 7, 1238-1314, (2006) · [Zbl 1222.68305](#)
- [43] Thiao, M.; Pham Dinh, T.; Le Thi, H. A., DC programming approach for a class of nonconvex programs involving ℓ_0 -norm, (Le Thi, H. A.; Bouvry, P.; Pham Dinh, T., *Modelling, computation and optimization in information systems and management sciences*, Vol. 14, CCIS, (2008), Springer Berlin, Heidelberg), 358-367

- [44] Thiao, M., Pham Dinh, T., & Le Thi, H.A. (2010). A DC programming approach for sparse eigenvalue problem. In `\textit{Proceeding of ICML 2010}` (pp. 1063-1070).
- [45] Tibshirani, R., Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society: Series B*, 58, 267-288, (1996) · [Zbl 0850.62538](#)
- [46] Wang, H.; Li, G.; Jiang, G., Robust regression shrinkage and consistent variable selection via the lad-lasso, *Journal of Business & Economics Statistics*, 25, 347-355, (2007)
- [47] Weston, J.; Elisseeff, A.; Schölkopf, B.; Kaelbling, P., Use of the zero-norm with linear models and kernel methods, *Journal of Machine Learning Research*, 3, 1439-1461, (2003) · [Zbl 1102.68605](#)
- [48] Wu, Y.; Liu, Y., Robust truncated-hinge-loss support vector machines, *Journal of the American Statistical Association*, 102, 974-983, (2007) · [Zbl 05564425](#)
- [49] Yuille, A. L.; Rangarajan, A., The concave-convex procedure (CCCP), *Neural Computation*, 15, 915-936, (2003) · [Zbl 1022.68112](#)
- [50] Zhang, C. H., Nearly unbiased variable selection under minimax concave penalty, *Annals of Statistics*, 38, 894-942, (2010) · [Zbl 1183.62120](#)
- [51] Zou, H., The adaptive lasso and its oracle properties, *Journal of the American Statistical Association*, 101, 1418-1429, (2006) · [Zbl 1171.62326](#)

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.