

Le Thi, Hoai An; Le, Hoai Minh; Pham Dinh, Tao

Feature selection in machine learning: an exact penalty approach using a difference of convex function algorithm. (English) [Zbl 1343.68201](#)

Mach. Learn. 101, No. 1-3, 163-186 (2015).

Summary: We develop an exact penalty approach for feature selection in machine learning via the zero-norm ℓ_0 -regularization problem. Using a new result on exact penalty techniques we reformulate equivalently the original problem as a Difference of Convex (DC) functions program. This approach permits us to consider all the existing convex and nonconvex approximation approaches to treat the zero-norm in a unified view within DC programming and DCA framework. An efficient DCA scheme is investigated for the resulting DC program. The algorithm is implemented for feature selection in SVM, that requires solving one linear program at each iteration and enjoys interesting convergence properties. We perform an empirical comparison with some nonconvex approximation approaches, and show using several datasets from the UCI database/Challenging NIPS 2003 that the proposed algorithm is efficient in both feature selection and classification.

MSC:

[68T05](#) Learning and adaptive systems in artificial intelligence

[62J07](#) Ridge regression; shrinkage estimators (Lasso)

[90C25](#) Convex programming

Cited in **12** Documents

Keywords:

[zero-norm](#); [feature selection](#); [exact penalty](#); [DC programming](#); [DCA](#)

Full Text: [DOI](#)

References:

- [1] Amaldi, E; Kann, V, On the approximability of minimizing non zero variables or unsatisfied relations in linear systems, *Theoretical Computer Science*, 209, 237-260, (1998) · [Zbl 0915.68072](#) · [doi:10.1016/S0304-3975\(97\)00115-1](#)
- [2] Bach, F., Jenatton, R., Mairal, J., & Obzinski, G. (2012). Optimization with sparsity-inducing penalties foundations and trends. *Foundations and Trends in Machine Learning*, 4(1), 1-106. · [Zbl 06064248](#)
- [3] Bradley, P. S., & Mangasarian, O. L. (1998). Feature selection via concave minimization and support vector machines. In *Proceeding of international conference on machine learning ICML'98*.
- [4] Candes, E; Wakin, M; Boyd, S, Enhancing sparsity by reweighted ℓ_1 minimization, *Journal of Mathematical Analysis and Applications*, 14, 877-905, (2008) · [Zbl 1176.94014](#)
- [5] Chartrand, R; Yin, W, Iteratively reweighted algorithms for compressive sensing, *Acoustics, speech and signal processing, IEEE international conference ICASSP, 2008*, 3869-3872, (2008)
- [6] Chen, X; Xu, FM; Ye, Y, Lower bound theory of nonzero entries in solutions of ℓ_2 - ℓ_1 minimization, *SIAM Journal on Scientific Computing*, 32, 2832-2852, (2010) · [Zbl 1242.90174](#) · [doi:10.1137/090761471](#)
- [7] Chen, Y., Li, Y., Cheng, X.-Q., & Guo, L. (2006). Survey and taxonomy of feature selection algorithms in intrusion detection system. In *Proceedings of inscrypt, 2006*. LNCS (Vol. 4318, 153-167). · [Zbl 1172.94613](#)
- [8] Collober, R., Sinz F., Weston, J., & Bottou, L. (2006). Trading convexity for scalability. In *Proceedings of the 23rd international conference on machine learning ICML 2006* (pp. 201-208). Pittsburgh, PA. ISBN:1-59593-383-2.
- [9] Cristianini, N., & Shawe-Taylor, N. (2000). *Introduction to support vector machines*. Cambridge: Cambridge University Press. · [Zbl 0994.68074](#)
- [10] Dempster, AP; Laird, NM; Rubin, DB, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society: Series B*, 39, 1-38, (1997) · [Zbl 0364.62022](#)
- [11] Fan, J; Li, R, Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association*, 96, 1348-1360, (2001) · [Zbl 1073.62547](#) · [doi:10.1198/016214501753382273](#)
- [12] Fu, WJ, Penalized regression: the bridge versus the lasso, *Journal of Computational and Graphical Statistics*, 7, 397-416, (1998)
- [13] Gasso, G; Rakotomamonjy, A; Canu, S, Recovering sparse signals with a certain family of nonconvex penalties and dc programming, *IEEE Transactions on Signal Processing*, 57, 4686-4698, (2009) · [Zbl 1391.90489](#) · [doi:10.1109/TSP.2009.2026004](#)

- [14] Gorodnitsky, IF; Rao, BD, Sparse signal reconstructions from limited data using FOCUSS: A re-weighted minimum norm algorithm, *IEEE Transactions on Signal Processing*, 45, 600-616, (1997) · doi:10.1109/78.558475
- [15] Guan, W; Gray, A, Sparse high-dimensional fractional-norm support vector machine via DC programming, *Computational Statistics and Data Analysis*, 67, 136-148, (2013) · doi:10.1016/j.csda.2013.01.020
- [16] Gribonval, R; Nielsen, M, Sparse representation in union of bases, *IEEE Transactions on Information Theory*, 49, 3320-3325, (2003) · Zbl 1286.94032 · doi:10.1109/TIT.2003.820031
- [17] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). Heidelberg: Springer. · Zbl 1273.62005
- [18] Huang, J., Horowitz, J., & Ma, S. (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Annals of Statistics*, 36, 587-613. · Zbl 1133.62048
- [19] Kim, Y; Choi, H; Oh, HS, Smoothly clipped absolute deviation on high dimensions, *Journal of the American Statistical Association*, 103, 1665-1673, (2008) · Zbl 1286.62062 · doi:10.1198/016214508000001066
- [20] Knight, K; Fu, W, Asymptotics for lasso-type estimators, *Annals of Statistics*, 28, 1356-1378, (2000) · Zbl 1105.62357 · doi:10.1214/aos/1015957397
- [21] Krause, N., & Singer, Y. (2004). Leveraging the margin more carefully. In *Proceedings of the 21 international conference on Machine learning ICML 2004*. Banff, Alberta, Canada, 63. ISBN:1-58113-828-5.
- [22] Le Thi, H.A. DC Programming and DCA. <http://lita.sciences.univ-metz.fr/~lethi>.
- [23] Le Thi, H. A. (1997). *Contribution à l'optimisation non convexe et l'optimisation globale: Théorie*. Algorithmes et Applications: Habilitation à Diriger des Recherches, Université de Rouen.
- [24] Le Thi, H. A., & Pham Dinh, T. (1997). Solving a class of linearly constrained indefinite quadratic problems by DC algorithms. *Journal of Global Optimization*, 11(3), 253-285. · Zbl 0905.90131
- [25] Le Thi, H. A., & Pham Dinh, T. (2005). 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. · Zbl 1116.90122
- [26] Le Thi, H. A., Belghiti, T., Pham Dinh, T. (2007) A new efficient algorithm based on DC programming and DCA for clustering. *Journal of Global Optimization*, 37, 593-608. · Zbl 1198.90327
- [27] Le Thi, H. A., Le, H. M. & Pham Dinh, T. (2006). Optimization based DC programming and DCA for hierarchical clustering. *European Journal of Operational Research*, 183(3), 1067-1085. · Zbl 1149.90117
- [28] Le Thi, H. A., Le, H. M., Nguyen, V. V., & Pham Dinh, T. (2008). A dc programming approach for feature selection in support vector machines learning. *Journal of Advances in Data Analysis and Classification*, 2, 259-278. · Zbl 1284.90057
- [29] Le Thi, H. A., Nguyen, V. V., & Ouchani, S. (2009). Gene selection for cancer classification using DCA. *Journal of Fonctiers of Computer Science and Technology*, 3(6), 62-72.
- [30] Le Thi, H. A., Huynh, V. N., & Pham Dinh, T. (2012). Exact penalty and error bounds in DC programming. *Journal of Global Optimization dedicated to Reiner Horst*, 52(3), 509-535. · Zbl 1242.49037
- [31] 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) · doi:10.1198/106186005X37238
- [32] Liu, Y; Shen, X, Multicategory ψ -learning, *Journal of the American Statistical Association*, 101, 500-509, (2006) · Zbl 1119.62341 · doi:10.1198/016214505000000781
- [33] Mangasarian, O. L. (1996). Machine learning via polyhedral concave minimization. In H. Fischer, B. Riedmueller, & S. Schaeffler (Eds.), *Applied mathematics and parallel computing—Festschrift for Klaus Ritter* (pp. 175-188). Heidelberg: Physica. · Zbl 0906.68127
- [34] Mallat, S; Zhang, Z, Matching pursuit in a time-frequency dictionary, *IEEE Transactions on Signal Processing*, 41, 3397-3415, (1993) · Zbl 0842.94004 · doi:10.1109/78.258082
- [35] Meinshausen, N, Relaxed lasso, *Computational Statistics and Data Analysis*, 52, 374-393, (2007) · Zbl 1452.62522 · doi:10.1016/j.csda.2006.12.019
- [36] Natarajan, BK, Sparse approximate solutions to linear systems, *SIAM Journal on Computing*, 24, 227-234, (1995) · Zbl 0827.68054 · doi:10.1137/S0097539792240406
- [37] Neumann, J; Schnörr, C; Steidl, G, Combined SVM-based feature selection and classification, *Machine Learning*, 61, 129-150, (2005) · Zbl 1137.90643 · doi:10.1007/s10994-005-1505-9
- [38] Ong, C. S., & Le Thi, H. A. (2013). Learning sparse classifiers with Difference of Convex functions algorithms. *Optimization Methods and Software*, 28(4), 830-854. · Zbl 1282.90181
- [39] Peleg, D; Meir, R, A bilinear formulation for vector sparsity optimization, *Signal Processing*, 8, 375-389, (2008) · Zbl 1186.94273 · doi:10.1016/j.sigpro.2007.08.015
- [40] Pham Dinh, T., & Le Thi, H. A. (1998). DC optimization algorithms for solving the trust region subproblem. *SIAM Journal on Optimization*, 8, 476-505. · Zbl 0913.65054
- [41] Pham Dinh, T., & Le Thi, H. A (2014). Recent advances in DC programming and DCA. *Transactions on Computational Collective Intelligence*, 3(42), 1-37.
- [42] Rakotomamonjy, A; Flamary, R; Gasso, G; Canu, S, $\ell_{1,2}$ penalty for sparse linear and sparse multiple kernel multi-task learning, *IEEE Transactions on Neural Networks*, 22, 13071320, (2011) · doi:10.1109/TNN.2011.2157521
- [43] Rao, BD; Kreutz-Delgado, K, An affine scaling methodology for best basis selection, *IEEE Transactions on Signal Processing*, 47, 187-200, (1999) · Zbl 0984.94010 · doi:10.1109/78.738251

- [44] Rao, BD; Engan, K; Cotter, SF; Palmer, J; KreutzDelgado, K, Subset selection in noise based on diversity measure minimization, *IEEE Transactions on Signal Processing*, 51, 760-770, (2003) · [doi:10.1109/TSP.2002.808076](https://doi.org/10.1109/TSP.2002.808076)
- [45] Rinaldi, F. (2000). *Mathematical Programming Methods for minimizing the zero-norm over polyhedral sets*, PhD thesis, Sapienza, University of Rome (2009)
- [46] Thiao, M., Pham Dinh, T., & Le Thi, H. A. (2010). A DC programming approach for sparse eigenvalue problem. *\textit{Proceeding of ICML}*, \textit{2010}, 1063-1070.
- [47] Tibshirani, R, Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society*, 46, 431-439, (1996) · [Zbl 0850.62538](https://zbmath.org/journals/jrssoc/B/46/431.html)
- [48] Yuille, AL; Rangarajan, A, The convex concave procedure, *Neural Computation*, 15, 915-936, (2003) · [Zbl 1022.68112](https://zbmath.org/journals/nc/15/915.html) · [doi:10.1162/08997660360581958](https://doi.org/10.1162/08997660360581958)
- [49] Wang, L; Zhu, J; Zou, H, The doubly regularized support vector machine, *Statistica Sinica*, 16, 589-615, (2006) · [Zbl 1126.68070](https://zbmath.org/journals/statistica/16/589.html)
- [50] Weston, J; Elisseeff, A; Scholkopf, B; Tipping, M, Use of the zero-norm with linear models and kernel methods, *Journal of Machine Learning Research.*, 3, 1439-1461, (2003) · [Zbl 1102.68605](https://zbmath.org/journals/jmlr/3/1439.html)
- [51] Zhang, HH; Ahn, J; Lin, X; Park, C, Gene selection using support vector machines with non-convex penalty, *Bioinformatics*, 2, 88-95, (2006) · [doi:10.1093/bioinformatics/bti736](https://doi.org/10.1093/bioinformatics/bti736)
- [52] Zou, H; Hastie, T, Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society: Series B*, 67, 301-320, (2005) · [Zbl 1069.62054](https://zbmath.org/journals/jrssoc/B/67/301.html) · [doi:10.1111/j.1467-9868.2005.00503.x](https://doi.org/10.1111/j.1467-9868.2005.00503.x)
- [53] Zou, H, The adaptive lasso and its oracle properties, *Journal of the American Statistical Association*, 101, 1418-1429, (2006) · [Zbl 1171.62326](https://zbmath.org/journals/jasa/101/1418.html) · [doi:10.1198/016214506000000735](https://doi.org/10.1198/016214506000000735)
- [54] Zou, H; Li, R, One-step sparse estimates in nonconcave penalized likelihood models, *Annals of Statistics*, 36, 1509-1533, (2008) · [Zbl 1142.62027](https://zbmath.org/journals/aos/36/1509.html) · [doi:10.1214/009053607000000802](https://doi.org/10.1214/009053607000000802)

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.