

**Phan, Duy Nhat; Le Thi, Hoai An**

**Group variable selection via  $\ell_{p,0}$  regularization and application to optimal scoring.** (English)

Zbl 1441.62073

Neural Netw. 118, 220-234 (2019).

Summary: The need to select groups of variables arises in many statistical modeling problems and applications. In this paper, we consider the  $\ell_{p,0}$ -norm regularization for enforcing group sparsity and investigate a DC (Difference of Convex functions) approximation approach for solving the  $\ell_{p,0}$ -norm regularization problem. We show that, with suitable parameters, the original and approximate problems are equivalent. Considering two equivalent formulations of the approximate problem we develop DC programming and DCA (DC Algorithm) for solving them. As an application, we implement the proposed algorithms for group variable selection in the optimal scoring problem. The sparsity is obtained by using the  $\ell_{p,0}$ -regularization that selects the same features in all discriminant vectors. The resulting sparse discriminant vectors provide a more interpretable low-dimensional representation of data. The experimental results on both simulated datasets and real datasets indicate the efficiency of the proposed algorithms.

**MSC:**

62F07 Statistical ranking and selection procedures

93E20 Optimal stochastic control

62P10 Applications of statistics to biology and medical sciences; meta analysis

Cited in 2 Documents

**Keywords:**

$\ell_{p,0}$  regularization; DC (Difference of Convex functions) approximation; DC programming; DCA; optimal scoring

**Full Text:** DOI

**References:**

- [1] Argyriou, A.; Evgeniou, T.; Pontil, M., Convex multi-task feature learning, *Machine Learning*, 73, 2, 243-272 (2008)
- [2] Bi, J.; Xiong, T.; Yu, S.; Dundar, M.; Rao, R. B., An improved multi-task learning approach with applications in medical diagnosis, (*ECML PKDD (Vol. 5211)* (2008)), 117-132
- [3] Blodet, M.; Seki, K.; Uehara, K., Block coordinate descent algorithms for large-scale sparse multiclass classification, *Machine Learning*, 93, 31-52 (2013) · Zbl 1293.68216
- [4] Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J., Distributed optimization and statistical learning via the alternating direction method of multipliers, *Foundations and Trends in Machine Learning*, 3, 1, 1-124 (2011)
- [5] Calandriello, D.; Lazaric, A.; Restelli, M., Sparse multi-task reinforcement learning, (*NIPS* (2014))
- [6] Chen, J.; Huo, X., Theoretical results on sparse representations of multiple-measurement vectors, *IEEE Transactions on Signal Processing*, 54, 4634-4643 (2006) · Zbl 1375.94051
- [7] Clemmensen, L.; Hansen, M.; Ersboll, B.; Frisvad, J., A method for comparison of growth media in objective identification of penicillium based on multi-spectral imaging, *Journal of Microbiological Methods*, 69, 249-255 (2007)
- [8] Clemmensen, L.; Hastie, T.; Witten, D.; Ersbøll, B., Sparse discriminant analysis, *Technometrics*, 53, 4, 406-413 (2011)
- [9] Cotter, S. F.; Rao, B. D.; Egan, K.; Kreutz-Delgado, K., Sparse solutions to linear inverse problems with multiple measurement vectors, *IEEE Transactions on Signal Processing*, 53, 2477-2488 (2005) · Zbl 1372.65123
- [10] Danaher, P.; Wang, P.; Witten, D. M., The joint graphical lasso for inverse covariance estimation across multiple classes, *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 76, 373-397 (2014)
- [11] Eksioğlu, E. M., Group sparse RLS algorithms, *International Journal of Adaptive Control and Signal Processing*, 28, 12, 1398-1412 (2014) · Zbl 1334.93188
- [12] Fan, J.; Li, R., Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association*, 96, 456, 1348-1360 (2001) · Zbl 1073.62547
- [13] Fisher, R. A., The use of multiple measurements in taxonomic problems, *Annals of Eugenics*, 7, 179-188 (1936)
- [14] Friedman, J.; Hastie, T.; Hoefling, H.; Tibshirani, R., Pathwise coordinate optimization, *The Annals of Applied Statistics*, 1, 302-332 (2007) · Zbl 1378.90064

- [15] Golub, T. R.; Slonim, D. K.; Tamayo, P.; Huard, C.; Gasenbeek, M.; Mesirov, J. P., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, *Science*, 286, 531-537 (1999)
- [16] Gu, Q.; Li, Z.; Han, J., Linear discriminant dimensionality reduction, (ECML PKDD (Vol. 6911) (2011)), 549-564
- [17] Hastie, T.; Buja, A.; Tibshirani, R., Penalized discriminant analysis, *The Annals of Statistics*, 23, 1, 73-102 (1995) · [Zbl 0821.62031](#)
- [18] Hastie, T.; Tibshirani, R.; Buja, A., Flexible discriminant analysis by optimal scoring, *Journal of the American Statistical Association*, 89, 1255-1270 (1994) · [Zbl 0812.62067](#)
- [19] Hu, Y.; Li, C.; Meng, K.; Qin, J.; Yang, X., Group sparse optimization via lp,q regularization, *Journal of Machine Learning Research*, 18, 1, 960-1011 (2017)
- [20] Huang, J.; Wei, S., Semiparametric regression pursuit, *Statistica Sinica*, 22, 1403-1426 (2012) · [Zbl 1253.62024](#)
- [21] Kha, Z.; Shafait, F.; Mian, A., Joint group sparse PCA for compressed hyperspectral imaging, *IEEE Transactions on Image Processing*, 24, 12, 4934-4942 (2015) · [Zbl 1408.94302](#)
- [22] Khan, J.; Wei, J. S.; Ringner, M.; Saal, L. H.; Ladanyi, M.; Westermann, F., Classification and diagnostic prediction of cancers using expression profiling and artificial neural networks, *Nature Medicine*, 7, 673-679 (2001)
- [23] Lan, X.; Ma, A.; Yuen, P.; Chellappa, R., Joint sparse representation robust feature-level fusion for multi-cue visual tracking, *IEEE Transactions on Image Processing*, 24, 12, 5826-5841 (2015) · [Zbl 1408.94339](#)
- [24] Le Thi, H. A., DC programming and DCA (2005), (Homepage), <http://www.lita.univ-lorraine.fr/lethi/index.php/en/research/dc-programming-and-dca.html>
- [25] 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, *Journal of Advances in Data Analysis and Classification*, 2, 3, 259-278 (2008) · [Zbl 1284.90057](#)
- [26] Le Thi, H. A.; Le Hoai, M.; Pham Dinh, T., Feature selection in machine learning: An exact penalty approach using a difference of convex function algorithm, *Machine Learning* (2014)
- [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.; Pham Dinh, T., DC programming and DCA: thirty years of developments, *Mathematical Programming*, 169, 1, 5-68 (2018) · [Zbl 1387.90197](#)
- [29] Le Thi, H. A.; Pham Dinh, T.; Le Hoai, M.; Vo Xuan, T., DC approximation approaches for sparse optimization, *European Journal of Operational Research*, 244, 26-44 (2015) · [Zbl 1346.90819](#)
- [30] Le Thi, H. A.; Phan, D. N., DC programming and DCA for sparse optimal scoring problem, *Neurocomputing*, 186, 170-181 (2016)
- [31] Le Thi, H. A.; Phan, D. N., DC programming and DCA for sparse Fisher linear discriminant analysis, *Neural Computing and Applications*, 28, 2809-2822 (2017)
- [32] Le Thi, H. A.; T. Vo Xuan, T.; T. Pham Dinh, T., Feature selection for linear SVMs under uncertain data: robust optimization based on difference of convex functions algorithms, *Neural Networks*, 59, 36-50 (2014) · [Zbl 1327.90236](#)
- [33] Lee, S.; Oh, M.; Kim, Y., Sparse optimization for nonconvex group penalized estimation, *Journal of Statistical Computation and Simulation*, 86, 597-610 (2016)
- [34] Leng, C., Sparse optimal scoring for multiclass cancer diagnosis and biomarker detection using microarray data, *Computational Biology and Chemistry*, 32, 417-425 (2008) · [Zbl 1158.92316](#)
- [35] Liu, J.; Ji, S.; Ye, J., Multi-task feature learning via efficient  $\ell_{2,1}$ -norm minimization, (UAI (2009))
- [36] Merchante, L. F.S.; Grandvalet, Y.; Govaert, G., An efficient approach to sparse linear discriminant analysis, (ICML (2012))
- [37] Nie, F.; Huang, H.; Cai, X.; Ding, C., Efficient and robust feature selection via joint  $\ell_{2,1}$ -norms minimization, (NIPS (2010))
- [38] Obozinski, G.; Taskar, B.; Jordan, M., Multi-task feature selection, Technical Report (2006), Department of Statistics, University of California: Department of Statistics, University of California Berkeley
- [39] Obozinski, G.; Taskar, B.; Jordan, M., Joint covariate selection and joint subspace selection for multiple classification problems, *Statistics and Computing*, 20, 2, 231-252 (2010)
- [40] Ong, C.; Le Thi, H. A., Learning sparse classifiers with difference of convex functions algorithms, *Optimization Methods & Software*, 28, 4, 830-854 (2013) · [Zbl 1282.90181](#)
- [41] Ong, C. S.; Le Thi, H. A., Learning sparse classifiers with difference of convex functions algorithms, *Optimization Methods & Software*, 28, 4, 830-854 (2013) · [Zbl 1282.90181](#)
- [42] Peleg, D.; Meir, R., A bilinear formulation for vector sparsity optimization, *Signal Processing*, 88, 2, 375-389 (2008) · [Zbl 1186.94273](#)
- [43] Pham Dinh, T.; Le Thi, H. A., Convex analysis approach to D.C. programming: Theory, algorithms and applications, *Acta Mathematica Vietnamica*, 22, 1, 289-355 (1997) · [Zbl 0895.90152](#)
- [44] Pham Dinh, T.; Le Thi, H. A., A DC optimization algorithm for solving the trust-region subproblem, *SIAM Journal of Optimization*, 8, 2, 476-505 (1998) · [Zbl 0913.65054](#)
- [45] Pham Dinh, T.; Le Thi, H. A., Recent advances in DC programming and DCA, *Transactions on Computational Collective Intelligence*, 8342, 1-37 (2014)
- [46] Quattoni, A.; Carreras, X.; Collins, M.; Darrell, T., An efficient projection for  $\ell_{\infty,1}$ -regularization, (ICML

(2009))

- [47] Sun, L.; Liu, J.; Chen, J.; Ye, J., Efficient recovery of jointly sparse vectors, (NIPS (2009))
- [48] Turlach, B. A.; Venables, W. N.; Wright, S. J., Simultaneous variable selection, *Technometrics*, 47, 3, 349-363 (2005)
- [49] Wang, L.; Chen, G.; Li, H., Group SCAD regression analysis for microarray time course gene expression data, *Bioinformatics*, 23, 1486-1494 (2007)
- [50] Wei, F.; Huang, J., Consistent group selection in high-dimensional linear regression, *Bernoulli*, 16, 1369-1384 (2010) · [Zbl 1207.62146](#)
- [51] Wei, F.; Zhu, H., Group coordinate descent algorithms for nonconvex penalized regression, *Computational Statistics & Data Analysis*, 56, 316-326 (2012) · [Zbl 1239.62082](#)
- [52] Yeoh, E. J.; Ross, M. E.; Shurtleff, S. A.; Williams, W. K.; Patel, D.; Mahfouz, R., Classification, subtype discovery, and prediction of outcome in pediatric lymphoblastic leukemia by gene expression profiling, *Cancer Cell*, 1, 133-143 (2002)
- [53] Yuan, M.; Lin, Y., Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 68, 49-67 (2006) · [Zbl 1141.62030](#)
- [54] Zhang, C.-H., Nearly unbiased variable selection under minimax concave penalty, *The Annals of Statistics*, 38, 2, 894-942 (2010) · [Zbl 1183.62120](#)
- [55] Zhang, H. H.; Liu, Y.; Wu, Y.; Zhu, J., Variable selection for the multiclass SVM via adaptive sup-norm regularization, *Electronic Journal of Statistics*, 2, 149-167 (2008) · [Zbl 1135.62056](#)
- [56] Zhang, Y.; Yeung, D. Y.; Xu, Q., Probabilistic multi-task feature selection, (NIPS (2010))
- [57] Zou, H.; Hastie, T.; Tibshirani, R., Sparse principal component analysis, *Journal of Computational and Graphical statistics*, 15, 265-286 (2006)

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.