zbMATH Open

THE FIRST RESOURCE FOR MATHEMATICS

Le Thi, Hoai An; Le, Hoai Minh; Nguyen, Van Vinh; Dinh, Tao Pham

A DC programming approach for feature selection in support vector machines learning. (English) Zbl 1284.90057

Adv. Data Anal. Classif., ADAC 2, No. 3, 259-278 (2008).

Summary: Feature selection consists of choosing a subset of available features that capture the relevant properties of the data. In supervised pattern classification, a good choice of features is fundamental for building compact and accurate classifiers. In this paper, we develop an efficient feature selection method using the zero-norm l_0 in the context of support vector machines (SVMs). Discontinuity at the origin for l_0 makes the solution of the corresponding optimization problem difficult to solve. To overcome this drawback, we use a robust DC (difference of convex functions) programming approach which is a general framework for non-convex continuous optimisation. We consider an appropriate continuous approximation to l_0 such that the resulting problem can be formulated as a DC program. Our DC algorithm (DCA) has a finite convergence and requires solving one linear program at each iteration. Computational experiments on standard datasets including challenging feature-selection problems of the NIPS 2003 feature selection challenge and gene selection for cancer classification show that the proposed method is promising: while it suppresses up to more than 99% of the features, it can provide a good classification. Moreover, the comparative results illustrate the superiority of the proposed approach over standard methods such as classical SVMs and feature selection concave.

MSC:

90C26 Nonconvex programming, global optimization
62-07 Data analysis (statistics) (MSC2010)
68T05 Learning and adaptive systems in artificial intelligence

Keywords:

feature selection; SVM; nonconvex optimisation; DC programming; DCA

Full Text: DOI

References:

[1] Amaldi E, Kann V (1998) On the approximability of minimizing non zero variables or unsatisfied relations in linear systems. Theor Comput Sci 209: 237–260 · Zbl 0915.68072 · doi:10.1016/S0304-3975(97)00115-1

[2] Bradley PS, Mangasarian OL (1998) Feature selection via concave minimization and support vector machines. In: Proceeding of international conference on machina learning ICML'98

[3] Boser B, Guyon I, Vapnik VN (1992) A training algorithm for optimal margin classifiers. In: Proceedings of the fifth annual workshop on computational learning theory, Kaufmann, San Mateo, pp 144–152

[4] Cortes C, Vapnik VN (1995) Support vector networks. Mach Learn 20(3): 273–297 · Zbl 0831.68098

[5] Cristianini N., Shawe-Taylor J (2000) Introduction to support vector machines. Cambridge University Press, Cambridge · Zbl 0994.68074

[6] Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES (1999) Molecular classifcation of cancer: class discovery and class prediction by gene expression monitoring. Science 286: 531–537· doi:10.1126/science.286.5439.531

[7] Guyon I, Weston J, Barnhill S, Vapnik VN (2002) Gene selection for cancer classification using support vector machines. Mach Learn 46(1–3): 389–422 · Zbl 0998.68111 · doi:10.1023/A:1012487302797

[8] Guyon I, Gunn S, Nikravesh M, Zadeh LA (2006) Feature extraction, foundations and applications. Springer, Berlin · Zbl 1114.68059

[9] Hastie T, Rosset S, Tibshirani R, Zhu J (2004) The entire regularization path for the support vector machine. J Mach Learn Res 5: 1391–1415 · Zbl 1222.68213

[10] Krause N, Singer Y (2004) Leveraging the margin more carefully. In: Proceedings of the twenty-first international conference on machine learning ICML 2004. Banff, Alberta, 63 pp. ISBN:1-58113-828-5

[11] Le Thi HA (1997) Contribution á l'optimisation non convexe et l'optimisation globale: Théorie, Algorithmes et Applications, Habilitation á Diriger des Recherches, Université de Rouen

[12] Le Thi HA, Pham Dinh T (1997) Solving a class of linearly constrained indefinite quadratic problems by DC algorithms. J Global Optim 11(3): 253–285 · Zbl 0905.90131 · doi:10.1023/A:1008288411710

[13] Le Thi HA, Pham Dinh T (2005) The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. Ann Oper Res 133: 23–46 · Zbl 1116.90122 · doi:10.1007/s10479-004-5022-1

[14] Le Thi HA, Belghiti T, Pham Dinh T (2006) A new efficient algorithm based on DC programming and DCA for clustering. J Global Optim 37: 593–608 · Zbl 1198.90327

[15] Le Thi HA, Le Hoai M, Pham Dinh T (2007) Optimization based DC programming and DCA for Hierarchical Clustering. Eur J Oper Res 183: 1067–1085 · Zbl 1149.90117 · doi:10.1016/j.ejor.2005.07.028

[16] Le Thi HA, Mamadou T, Pham Dinh T (2008) DC Programming approach for solving a class of nonconvex programs dealing with zero-norm. In: Modelling, computation and optimization in information systems and management science. CCIS 14. Springer, Heidelberg, pp 348–357

[17] Liu Y, Shen X, Doss H (2005) Multicategory {\(\psi\)}-learning and support vector machine: computational tools. J Computat Graph Stat 14: 219–236· doi:10.1198/106186005X37238

[18] Liu Y, Shen X (2006) Multicategory{\(\psi\)}-learning. J Am Stat Assoc 101: 500–509 · Zbl 1119.62341 · doi:10.1198/016214505000000781

[19] Neumann J, Schnörr C, Steidl G (2005) Combined SVM-based feature selection and classification. Mach Learn 61(1–3): 129–150 · Zbl 1137.90643 · doi:10.1007/s10994-005-1505-9

[20] Pham Dinh T, Le Thi HA (1998) DC optimization algorithms for solving the trust region subproblem. SIAM J Optim 8: 476–505 · Zbl 0913.65054 · doi:10.1137/S1052623494274313

[21] Rakotomamonjy A (2003) Variable selection using SVM-based criteria. J Mach Learn Res 3: 1357–1370 · Zbl 1102.68583

[22] Ronan C, Fabian S, Jason W, Lé B (2006) Trading convexity for scalability. In: Proceedings of the 23rd international conference on machine learning ICML 2006. Pittsburgh, pp 201–208. ISBN:1-59593-383-2

[23] Singh D, Febbo GPG, Ross K, Jackson DG, Manola J, Ladd C, Tamayo P, Renshaw AA, D'Amico AV, Richie JP, Lander ES, Loda M, Kantoff PW, Golub TR, Sellers WR (2002) Gene expression correlates of clinical prostate cancer behavior Copyright {\copyright} 2002 Cell Press, Cancer Cell, vol 1, pp 203–209, March 2002

[24] Shen KQ, Ong CJ, Li XP, Wilder-Smith EPV (2008) Feature selection via sensitivity analysis of SVM probabilistic outputs. Mach Learn 70: 1–20 · Zbl 1470.68175 · doi:10.1007/s10994-007-5025-7

[25] Schmidt M, Fung G, Rosales R (2007) Fast optimization methods for L1 regularization: a comparative study and two new approaches. In: Proceedings of machine learning: ECML 2007, Lecture notes in computer science, vol 4701/2007, pp 286–297

[26] Van't Veer L et al (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature 415: 530–536· doi:10.1038/415530a

[27] Vapnik VN (1995) The nature of statistical learning theory. Springer, New York · Zbl 0833.62008

[28] Weston J, Elisseeff A, Scholkopf B, Tipping M (2003) Use of the zero-norm with linear models and kernel methods. J Mach Learn Res 3: 1439–1461 · Zbl 1102.68605

[29] Yuille AL, Rangarajan A (2003) The convex concave procedure. Neural Computat 15(4):915–936. ISSN:0899-7667 · Zbl 1022.68112