Discrete Support Vector Machines

J.P. Brooks R. Shioda A. Spencer

July, 2007

Abstract

In this paper, we develop a discrete version of the support vector machine (SVM) that minimizes the number of misclassified observations. Traditional SVM approximates the total number of misclassified points with continuous margin errors, which is a measure prone to over-fitting in the presence of outliers. Our approach avoids this by directly minimizing the number of misclassified points by using a mixed-integer quadratic programming model. We also introduce a formulation that can accommodate a variety of choices for the kernel. To improve the solution time for these formulations, we propose a heuristic to find good feasible solutions and a family of cuts for use in a branch and cut framework for optimization. In preliminary computational experimentation, we observe several data sets where our method is significantly less affected by outliers than the traditional continuous SVM.

C&O Research Report : CORR 2007-12
Department of Combinatorics and Optimization
University of Waterloo
Waterloo, ON, Canada