Solving Mixed Integer Formulation of the KS Maximization Problem – Dual Based Methods & Results from Practical Problems

Debashish Sarkar
Management Sciences Group
CIT, New Jersey, USA
Deb.Sarkar@cit.com

Abstract

Typically credit-scoring formulations model the response variable (good or bad) as a logistic function and the parameter estimates are obtained by maximizing the log likelihood function. One of the most common metrics to measure the effectiveness of the credit scoring function is the KS (Kolmogorov-Smirnov) statistic. We present a mixed integer programming formulation for the credit-scoring problem by directly maximizing the KS statistic. Generally, it is believed that MLE estimators produce "good" KS values; however, no results are known as to how far are these values from the "maximum" KS possible for large practical problems.

For a typical application, the KS maximization problem is a mixed integer program with several thousand binary variables whose solution require significant computer resources and is hard to solve. We show that the solution to a specific Lagrangian dual problem is computationally tractable and the quality of the solution is very attractive (the optimum dual variables of the Lagrangian dual define the credit scoring model weights). We compare the KS values from the dual and the MLE methodologies for several practical problems.