58e A Mixed-Integer Programming Approach to Multi-Class Data Classification Problem

Metin Turkay and Fadime Uney

This paper presents a new data classification method based on mixed-integer programming. Traditional approaches that are based on partitioning the data sets into two groups perform poorly for multi-class data classification problems. The proposed approach is based on the use of hyper-boxes for defining boundaries of the classes that include all or some of the points in that set. A mixed-integer programming model is developed for representing existence of hyper-boxes and their boundaries. In addition, the relationships among the discrete decisions in the model are represented using propositional logic and then converted to their equivalent integer constraints using Boolean algebra. The proposed approach for multi-class data classification is illustrated on an example problem. The efficiency of the proposed method is tested on the prediction of folding types in proteins. The computational results show that the proposed method is very accurate and efficient on very challenging multi-class data classification problems.