

Chapter 16

Linear Programming Approaches for Multiple– Class Discriminant and Classification Analysis

Minghe Sun

University of Texas at San Antonio, USA

ABSTRACT

New linear programming approaches are proposed as nonparametric procedures for multiple-class discriminant and classification analysis. A new MSD model minimizing the sum of the classification errors is formulated to construct discriminant functions. This model has desirable properties because it is versatile and is immune to the pathologies of some of the earlier mathematical programming models for two-class classification. It is also purely systematic and algorithmic and no user ad hoc and trial judgment is required. Furthermore, it can be used as the basis to develop other models, such as a multiple-class support vector machine and a mixed integer programming model, for discrimination and classification. A MMD model minimizing the maximum of the classification errors, although with very limited use, is also studied. These models may also be considered as generalizations of mathematical programming formulations for two-class classification. By the same approach, other mathematical programming formulations for two-class classification can be easily generalized to multiple-class formulations. Results on standard as well as randomly generated test datasets show that the MSD model is very effective in generating powerful discriminant functions.

DOI: 10.4018/978-1-4666-1589-2.ch016

1. INTRODUCTION

In this study, new linear programming (LP) approaches are developed as nonparametric procedures for multiple-class discrimination and classification. Specifically, a LP model minimizing the sum of the classification errors, or the L_1 -norm, and a LP model minimizing the maximum of the classification errors, or the L_∞ -norm, are formulated. These new models are more straightforward and simpler than any of the previous LP formulations for multiple-class discrimination and classification. Although having their own rights, these new models may also be considered as generalizations from two-class to multiple-class techniques. The LP model minimizing the L_1 -norm is similar to the MSD (minimizes the sum of deviations) models and that minimizing the L_∞ -norm is similar to the MMD (minimizes the maximum of deviations) models for two-class classification. Properties of these new models are studied. The MSD model is immune to the difficulties caused by pathologies of earlier mathematical programming (MP) models for two-class classification although the MMD model is very limited in use.

Discriminant and classification analysis has been fundamental scientific research and practical applications. Discriminant analysis involves the study of the differences between two or more classes of objects represented by measurements, or variables, of different characteristics or attributes. Classification analysis involves the study of assigning new observations into one of the classes based on the measurements of the characteristics. Applications of discriminant and classification analysis are diverse. For example, reported applications in business include financial management, human resource management, and marketing; applications in biology and medicine include patient classification, disease diagnosis and species classification; and applications in environment and geography include remote sensing image pattern classification and pollution control. In fact, this

study was motivated by the need to identify a few from many thousands of genes that can be used to classify tissue samples into two or more classes, such as normal and tumor tissues, and to identify genes responsible for certain diseases (Ambroise & McLachlan, 2002; Dudoit, Fridlyand, & Speed, 2002; Nguyen & Rocke, 2002; Sun, 2003; Sun & Xiong, 2003). The availability of large datasets collected using current information technology, such as the Internet, imaging and remote sensing, made the traditional discriminant and classification techniques inadequate. Discriminant and classification analysis techniques will play a more important role in data analysis as information technology advances and as huge datasets need to be analyzed with data mining tools.

Based on known values of the attributes or variables and known class memberships of the observations in a sample, discriminant functions are constructed. The attribute values of an observation can be evaluated by these discriminant functions to obtain discriminant scores based on which the observation is assigned to a class. For many decades, statistical techniques, such as Fisher's linear discriminant function (LDF) (Fisher, 1936), Smith's quadratic discriminant function (Smith, 1947) and logistic regression (Hand, 1981), have been standard tools for this purpose. Statistical methods perform well when the data analyzed satisfy the underlying assumptions, such as multivariate normality and equal covariance matrices of the independent variables, although minor deviations from these assumptions do not severely affect the performance of these techniques. More recently, other techniques, such as MP (Freed & Glover, 1981a, 1981b; Hand, 1981), neural networks (Stern, 1996) and classification trees (Breiman *et al.*, 1984), have become alternative tools to rival statistical techniques. A revolutionary development in classification is the support vector machine (SVM) (Vapnik, 1995, 1998). A spectrum of techniques is needed because no single technique always outperforms others under all situations (Johnson & Wichern, 1988).

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/linear-programming-approaches-multiple-class/65968

Related Content

Top Management's Role in Promoting Decision Support Systems Efficiency: An Exploratory Study in Government Sector in Saudi Arabia

Abdullah Ibrahim Alkrajji (2021). *Research Anthology on Decision Support Systems and Decision Management in Healthcare, Business, and Engineering* (pp. 1409-1429).

www.irma-international.org/chapter/top-managements-role-in-promoting-decision-support-systems-efficiency/282648

Implementation of an Intelligent Model for Decision Making Based on CBR for Supply Chain Solution in Retail for a Cluster of Supermarkets

Adrian F. Loera-Castro, Jaime Sanchez, Jorge Restrepo, Angel Fabián Campoya Morales and Julian I. Aguilar-Duque (2021). *Research Anthology on Decision Support Systems and Decision Management in Healthcare, Business, and Engineering* (pp. 1224-1240).

www.irma-international.org/chapter/implementation-of-an-intelligent-model-for-decision-making-based-on-cbr-for-supply-chain-solution-in-retail-for-a-cluster-of-supermarkets/282637

Best Practices for Culturally Sensitive Data Visualizations

Michael Gendron, Christopher Hutwelker and Krzysztof Kisz (2017). *Decision Management: Concepts, Methodologies, Tools, and Applications* (pp. 1777-1798).

www.irma-international.org/chapter/best-practices-for-culturally-sensitive-data-visualizations/176832

A Bi-Objective Vehicle Routing Problem with Time Window by Considering Customer Satisfaction

Masoud Rabbani, Mahyar Taheri and Mohammad Ravanbakhsh (2016). *International Journal of Strategic Decision Sciences* (pp. 16-39).

www.irma-international.org/article/a-bi-objective-vehicle-routing-problem-with-time-window-by-considering-customer-satisfaction/163959

An Investigation Tool for Mounting Sustainable Practice: Modeling Using GIVTFNs in an Indian Context

Atul Kumar Sahu, Nitin Kumar Sahu, Anoop Kumar Sahu, Mridul Singh Rajput and Harendra Kumar Narang (2020). *International Journal of Decision Support System Technology* (pp. 25-49).

www.irma-international.org/article/an-investigation-tool-for-mounting-sustainable-practice/249132