Performance of Nonparametric Discriminant Analysis on Educational Indicator-Profiles

Lin Miao-Hsiang (1st Author)
Institute of Statistical Science, Academia Sinica
Taipei 115 Taiwan
Taipei, Taiwan, Republic of China
miao@stat.sinica.edu.tw.

Huang Su-Yun (2nd Author)
Institute of Statistical Science, Academia Sinica
Taipei 115 Taiwan
Taipei, Taiwan, Republic of China
syhuang@stat.sinica.edu.tw.

1. Purposes of This Study

In the fields of education and applied psychology, the parametric discriminant method (PDM) has proven useful for identifying structure dimensions along which the groups differ. However, the classification rules of PDM based on the centour method performed badly when the profile variables as a set all measuring cognitive attributes.

The use of multiple measurements becomes inevitable in relation to the context of building educational indicator systems for monitoring schooling performances. For example, to assess students' learning progress in science education, it is required that the educational indicators consist of measurements on science achievement (or ability), plus participations, interests, attitudes or aspirations toward science. Such set of the educational indicators, containing scores on both cognitive and non-cognitive attributes, is considered crucial for the central educational problems: diagnosing student learning difficulties and facilitating instructional guidance.

On the other hand, profile vectors of educational indicators are of mixed type—continuous, discrete, and categorical. This endemic measurement aspect results in the observed data vectors failing to comply with the two basic assumptions underlying the classical PDM: a multivariate normal distribution on the profile variables and equal group covariance matrices. Nonparametric versions of discriminant methods (NDM) can be generalized to deal with discrete and mixed data vectors. The purpose of this paper is to evaluate the performance goodness of NDM when applied to the central educational problems-- instructional groupings/diagnosis.

2. Samples and Data Structure

This article analyzed the data collected from a large-scale project for developing educational indicators used to monitor and upgrade Taiwan's elementary and secondary science education. The set of eight educational indicators developed for assessing sixth grade students were four subtests of biology, physics, chemistry and earth science, along with four scales of interests / attitudes toward science, homework involvement, teaching quality, and parents' education levels. The sample for this study consists of 1288 sixth- graders from 30 classrooms distributed in six public schools in Taipei City and County. The information on this sample included students' past GPA records and their scores on the eight educational indicators.

This study uses students' past GPA records to designate the membership of three groups: remedial-, normal-, and advanced-instruction groups. The procedure for identifying the group membership was based on "district norm" instead of "school norms". This study thus obtained a data matrix of 1288 by 9, which contains a classification variable with three levels and eight discriminating variables. The eight discriminating variables are categorical and correlated, and do not have the same measurement units and standard deviations.

3. Initial Analyses and Results

Tentative results showed the followings: 1) when applied to the training data sets, NDM relative to PDM yield much higher percentages of correct classification, regardless of the pooled or individual within-group covariance matrices being used to derive the classification criterion; 2) when applied to the test data sets, the kernel methods perform better than PDM in terms of reducing misclassification rates; however, when the prior probabilities of the groups taken into account, the kernel methods fail to outperform the parametric discriminant method. (rigid analyses to be done).

REFERENCE

Cooley, William W., and Paul R. Lohnes (1971). *Multivariate Data Analysis*, New York: Wiley.

Hand, D. J. (1982). *Kernel Discriminant Analysis*, New York: Research Studies Press.

Nunnally, J. C. (1978). *Psychometric Theory*, New York: McGraw-Hill Book Company.

ODDen, A. (1990). Educational Indicators in the United States：The Need for Analysis. Educational Research, June-July, 24-29.

Sam E. (1999). *Nonparametric Curve Estimation*. New York: Springer.

Shavelson, R., McDonnell, L., Oakes, J., Carey, N.,& Picus, L., (1987). Indicator Systems for Monitoring Math and Science Education, The RAND Corporation, Santa Monica, CA.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*, New York: Chapman and Hall.

RESUME (Lin Miao-Hsiang)

EDUCATION

Cornell University, U.S.A. 9/1981 - 1/1985 Ph.D. Psychometrics/Educational Measurement
University of Minnesota, U.S.A. 9/1979 - 6/1981 M.S Educational Measurement/Statistics
National Taiwan Normal University 9/1970 - 6/1974 B.S Educational Psychology

POSITIONS

2/1995 - Research Fellow, Institute of Statistical Science, Academia Sinica, R.O.C.
2/1996 - Adjunct Full Professor, Institute of Sociology, National Taipei University, Taiwan,
4/1987 – 2/1995 Associate Research Fellow, Institute of Statistical Science, Academia Sinica, R.O.C.
1/1985 – 12/1985 Measurement Specialist/lead Analyst, National Evaluation System, Inc. U.S.A.

HONORS

A. ISI Member: Ordinary Member of the International Statistical Institute
B. Excellence Award from Divisions of Science Education and Humanities/Social Science, N.S.C, R.O.C.