

A NONPARAMETRIC DISCRIMINANT VARIABLE-SELECTION ALGORITHM FOR CLASSIFICATION TO TWO POPULATIONS

S. PADMANABAN¹ & MARTIN L. WILLIAM²

¹NIRRH Field Unit, I.C.M.R, KMC Hospital, Chennai, India

²Department of Statistics, Loyola College, Chennai, India

ABSTRACT

This paper proposes a nonparametric discriminant variable-selection algorithm to discriminate two multivariate populations and an associated optimal decision rule for membership-prediction. The present work relaxes the 'equal variance-covariance matrices' condition traditionally imposed and develops a discrimination-classification procedure by including variables that best contribute to the 'discrimination', one-by-one in a forward-stepwise manner. The inclusion of variables in the discriminant is determined on the basis of best 'discriminating ability' as reflected in 'maximal difference' between the distributions of the discriminant in the two populations. A new decision-rule for classification or membership-prediction with a view to maximize correct predictions is provided. The proposed algorithm is applied to develop an optimal discriminant for predicting preterm labour among expecting mothers in the city of Chennai, India, and its performance is compared with logistic regression.

KEYWORDS: Classification, Discriminant, Kolmogorov-Smirnov Statistic