A NONPARAMETRIC DISCRIMINANT VARIABLE-ELIMINATION ALGORITHM

FOR CLASSIFICATION TO TWO POPULATIONS

S. PADMANABAN[1] & MARTIN L. WILLIAM[2]

[1]NIRRH Field Unit, Indian Council of Medical Research, KMC Hospital, Chennai, India

[2]Department of Statistics, Loyola College, Chennai, India

ABSTRACT

This paper provides a nonparametric discriminant variable-elimination algorithm to discriminate two multivariate populations and an associated optimal decision rule for membership-prediction. This is an alternative to the 'forward-stepwise' approach recently proposed for the same classification problem by Padmanaban and William (2016). As in the referred work, the present work relaxes the 'equal variance-covariance matrices' condition traditionally imposed and develops a discrimination-classification procedure by excluding variables that do not contribute to the 'discrimination', one-by-one in a backward-stepwise manner. The exclusion of variables in the discriminant is determined on the basis of least 'discriminating ability' as reflected in 'difference' between the distributions of the discriminant in the two populations. A decision-rule for classification or membership-prediction with a view to maximize correct predictions, balancing between 'sensitivity' and 'specificity', is provided. The proposed alogorithm 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 and also with the forward-stepwise discriminant algorithm of the same authors.

KEYWORDS: Classification, Discriminant, Variable-Elimination, Kolmogorov-Smirnov Statistic