

Abstract

Principal component analysis or factor analysis are often used to reduce the dimensionality of the original variables. However, the principal component or factor, which has larger variance (i.e eigenvalue) explaining larger proportion of total sample variance, may not retain the most information for other analyses later. For example, using the first few principal components or factors having the largest corresponding eigenvalues as discriminant variables, the discriminant result may not be good or even appropriate.

We first discuss two methods, given by Mardia et al. (1979) and Chang (1983) for choosing discriminant variables when data are randomly obtained from a mixture of two multivariate normal distributions. We then use the discriminant result (or classification error rates) to compare these two methods and the traditional method of using the principal components, which have the larger corresponding eigenvalues, as discriminant variables. We also prove that the both the two methods have the same selection order on principal components and factor (obtained by the principal component method). Furthermore, we use the method of Mardia et al. to select appropriate discriminators when data is from three populations.