

ABSTRACT

Linear Discriminant Analysis

Robert H. Riffenburgh^{avry}

Linear discriminant analysis is the classification of an individual as having arisen from one or the other of two populations on the basis of a scalar linear function of measurements of the individual. This paper is a population and large sample study of linear discriminant analysis. The population study is carried out on three levels:

- (1.1) (a) with loss functions and prior probabilities,
 (b) without loss functions but with prior probabilities,
 (c) with neither.

The first level leads to consideration of risks which may be split into two components, one for each type of misclassification, i.e. classification of an individual into population I given it arose from population II, and classification of it into II given it arose from I. Similarly, the second level leads to consideration of expected errors and the third level leads to consideration of conditional probabilities of misclassification, both again which may be divided into the same two components. At each level the "optimum" discriminator should jointly minimize the two probability components. These quantities are all positive for all hyperplanes. Either one of any pair may be made equal to zero by classifying all individuals of a sample into the appropriate population; but

this maximizes the other one. Consequently, joint minimization must be obtained by some compromise, e.g. by selecting a single criterion to be minimized. Two types of criteria for judging discriminators are considered at each level:

- (1.4) (i) Total risk (a)
- (1.5) Total expected errors (b)
- (1.6) Sum of conditional probabilities of misclassification (c)
- (1.7) (ii) Larger risk (a)
- (1.8) Larger expected error (b)
- (1.9) Larger conditional probability of misclassification (c).

These criteria are not particularly new, but have not been applied to linear discrimination and not been all used jointly.

If A is a k -dimensional row vector of direction numbers, X a k -dimensional row vector of variables, and c a constant, a linear discriminator is

$$(1.10) \quad AX' = c,$$

which also represents a hyperplane in k -space. An individual is classified as being from one or the other population on the basis of its position relative to the hyperplane.

The parameters A and c of (1.10) were investigated to find those sets of values which minimize each of the two criteria at various levels. Exact results were found for A under some circumstances and approximate results in others. At the levels (b) and (c), when exact results were obtained, they were the same for both

criteria and were independent of c . Investigation of the c 's showed the c 's to be exact functions of A and the parameters and yielded one c for each criterion.

At level (c), the c 's for criteria (i) and (ii), $c(\min)$ and $c(\sigma)$ respectively, were compared to $c(m)$, a population analog of the c suggested by other authors, to discover the conditions under which it was better (i.e. having lesser criteria) than both $c(\min)$, $c(\sigma)$ on criterion (ii), (i) respectively.

In the large sample study, variances and covariances were found (in many cases approximately) for all estimates of the parameters entering into the conditional probabilities of misclassification (level (c)). Extension of results to level (b) and to special cases of level (a) were given. From these variances and covariances were derived the expectations of these probabilities for both criteria, at level (c), and comparisons were made where feasible. Results were tabulated.