Statistics 135

Chapter 11

Discriminant Analysis

Chris Drake

Department of Statistics

University of California, Davis

Discrimination and Classification

1 Discrimination refers to separating distinct groups of objects into separate groups of clusters. Discriminant analysis is more exploratory in nature and is often employed when causal mechanisms are not well understood. The goal is to describe differential features of objects from several known groups.

2 Classification refers to allocating new objects into previously defined groups. Classification procedures lead to well defined rules for assigning new objects. The goal is to derive rules that can be used to optimally assign new objects to established groups.

Example:

Admissions offices of colleges might desire to

a Investigate factors that discriminate between applicants who are likely to succeed in a degree program and graduate

b Classify new applicants as likely to graduate or likely to drop out.

Classification for two populations

1. Two populations or classes with labels π_1 and π_2.
2. A vector $\mathbf{X} = X_1, ..., X_p$ of observations which differ in values (have different distributions) between the two populations.
3. The measured characteristics $X_1, ..., X_p$ are examined for differences between the two populations. This is done with the help of a learning sample where population membership is known.
4. The set of all possible sample outcomes is divided into two regions R_1 and R_2. If an observation falls into R_1 it is classified as belonging to population π_1. The learning sample is the basis for determining regions R_1 and R_2.
5. A "good" classification rule minimizes the chances for misclassifying objects.
6. Classification rules should take into account prior probabilities of belonging to π_1 rather than π_2 and vice versa.
7. Cost of misclassifying an object from π_1 into π_2 and vice versa should be taken into account.

Mathematical Model

1 $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are the probability density functions associated with $\mathbf{X}$ for populations π_1 and π_2.

2 Ω is the sample space and R_1 the set of values $\mathbf{x}$ for which objects are classified as π_1; $R_2 = \Omega - R_1$ is the set of values for which objects are classified as π_2. Note, that R_1 and R_2 are mutually exclusive and exhaustive.

3 The probability of classifying and object as π_2 when it belongs to π_1 is given by

$$P(2 \mid 1) = P(\mathbf{X} \in R_2 \mid \pi_1) = \int_{R_2} f_1(\mathbf{x}) d\mathbf{x}$$

$P(1 \mid 2)$ is similarly defined.

4 p_1 is the prior probability of π_1 and p_2 is the prior probability of π_2. Then the probability an observation is classified correctly as π_1 equals

$$P(\mathbf{X} \in R_1 \mid \pi_1) P(\pi_1) = P(1 \mid 1) p_1$$

5 The probability an object is misclassified as π_1 when it belongs to π_2 is given by

$$P(\mathbf{X} \in R_1 \mid \pi_2)P(\pi_2) = P(1 \mid 2)p_2$$

6 Similarly,

$$P(\mathbf{X} \in R_2 \mid \pi_2)P(\pi_2) = P(2 \mid 2)p_2$$

and

$$P(\mathbf{X} \in R_2 \mid \pi_1)P(\pi_1) = P(2 \mid 1)p_1$$

are the probabilities of correctly classifying an observation from π_2 and misclassifying an observation from π_1 into π_2.

7 The costs of classifying an observation correctly are zero and the cost of misclassifying an object from π_1 into π_2 is $c(2 \mid 1)$ and $c(1 \mid 2)$ for misclassifying an object from π_2 into π_1. The expected cost of misclassification ECM is

$$ECM = c(2 \mid 1)P(2 \mid 1)p_1 + c(1 \mid 2)P(1 \mid 2)p_2$$

smallskip

8 A reasonable rule should keep ECM as small as possible.

9 The regions that minimize ECM are given by

$$R_1: \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right)\left(\frac{p_2}{p_1}\right) \qquad R_2: \quad \frac{f_2(\mathbf{x})}{f_1(\mathbf{x})} < \left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right)\left(\frac{p_2}{p_1}\right)$$

10 If $p_1 = p_1$ then

$$R_1: \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right) \quad and \quad R_2: \quad \frac{f_2(\mathbf{x})}{f_1(\mathbf{x})} < \left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right)$$

11 while for equal costs

$$R_1: \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{p_2}{p_1} \quad and \quad R_2: \quad \frac{f_2(\mathbf{x})}{f_1(\mathbf{x})} < \frac{p_2}{p_1}$$

12 and for equal costs and prior probabilities

$$R_1: \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq 1 \quad and \quad R_2: \quad \frac{f_2(\mathbf{x})}{f_1(\mathbf{x})} < 1$$

13 When the prior probabilities are unknown, they are often taken to be equal.

14 If the misclassification cost ratio is unknown, it i soften taken to be one.

15 Criterion for determining R_1, R_2 that does not consider cost is the total probability of misclassification:

$$\begin{aligned}TPM =& P((\mathbf{X} \in R_2)\text{and comes from } \pi_1) \\ &+ P((\mathbf{X} \in R_1)\text{and comes from } \pi_2) \\ =& p_1 \int_{R_2} f_1(\mathbf{x})d\mathbf{x} + p_2 \int_{R_1} f_2(\mathbf{x})d\mathbf{x}\end{aligned}$$

16 Allocation by the posterior probability $P(\pi_i \mid \mathbf{x}_0)$ where

$$P(\pi_1 \mid \mathbf{x}_0) = \frac{p_1 f_1(\mathbf{x}_0)}{p_1 f_1(\mathbf{x}_0) + p_2 f_2(\mathbf{x}_0)}$$

$$P(\pi_2 \mid \mathbf{x}_0) = 1 - P(\pi_1 \mid \mathbf{x}_0)$$

Two multivariate populations

1 Two populations:

  a π_1 with $\mathbf{X} \sim N(\mu_1, \mathbf{\Sigma})$

  b π_2 with $\mathbf{X} \sim N(\mu_2, \mathbf{\Sigma})$.

2 The regions with minimum ECM for known μ_1, μ_2 and $\mathbf{\Sigma}$ are

$$(a) \qquad R_1 : \; exp\left[-\frac{1}{2}\big(\mathbf{x}-\mu_1\big)'\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu_1\big) + \frac{1}{2}\big(\mathbf{x}-\mu_2\big)'\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu_2\big)\right] \geq \left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right)\left(\frac{p_2}{p_1}\right)$$

$$(b) \qquad R_2 : \; exp\left[-\frac{1}{2}\big(\mathbf{x}-\mu_1\big)'\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu_1\big) + \frac{1}{2}\big(\mathbf{x}-\mu_2\big)'\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu_2\big)\right] < \left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right)\left(\frac{p_2}{p_1}\right)$$

3 Note, the left and right hand sides of (2a) and (2b) are positive, therefore, we can work with logarithms instead. This leads to the classification rule: allocate $\mathbf{x}_0$ to π_1 if

$$(\mu_1 - \mu_2)'\mathbf{\Sigma}^{-1}\mathbf{x}_0 - \frac{1}{2}(\mu_1 - \mu_2)'\mathbf{\Sigma}^{-1}(\mu_1 + \mu_2) \geq ln\left[\left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$

otherwise, allocate $\mathbf{x}_0$ to π_2.

4 μ_1, μ_2 and $\mathbf{\Sigma}$ are usually unknown; if we have samples $\mathbf{x}_{11}, .., \mathbf{x}_{1n_1}$ and $\mathbf{x}_{21}, .., \mathbf{x}_{2n_2}$, we replace μ_i by $\bar{\mathbf{x}}_i$ for $i = 1, 2$ and $\mathbf{\Sigma}$ by $\mathbf{S}_{pooled}$ to get the estimated ECM rule: allocate $\mathbf{x}_0$ to π_1 if

$$(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_{pooled}^{-1}\mathbf{x}_0 - \frac{1}{2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_{pooled}^{-1}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \geq ln\left[\left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$

otherwise, allocate $\mathbf{x}_0$ to π_2.