Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Multivariate Analysis 96 (2005) 384–403

Journal of Multivariate Analysis

www.elsevier.com/locate/jmva

Influence of observations on the misclassification probability in quadratic discriminant analysis

Christophe Croux*, Kristel Joossens

Department of Applied Economics, Katholieke Universiteit Leuven, Naamsestraat 69, B-3000 Leuven, Belgium

Received 12 February 2004
Available online 16 December 2004

Abstract

In this paper it is studied how observations in the training sample affect the misclassification probability of a quadratic discriminant rule. An approach based on partial influence functions is followed. It allows to quantify the effect of observations in the training sample on the performance of the associated classification rule. Focus is on the effect of outliers on the misclassification rate, merely than on the estimates of the parameters of the quadratic discriminant rule. The expression for the partial influence function is then used to construct a diagnostic tool for detecting influential observations. Applications on real data sets are provided.

AMS 2000 subject classification: 62H30; 62F35

Keywords: Classification; Diagnostics; Misclassification probability; Outliers; Partial influence functions; Quadratic discriminant analysis

1. Introduction

In discriminant analysis one observes two groups of multivariate observations, forming together the *training sample*. For the data in this training sample, it is known to which group they belong. On the basis of the training sample a discriminant function Q will be constructed. Such a rule is used afterwards to classify new observations, for which the group membership is unknown, into one of the two groups. Data are generated by two different

* Corresponding author. Fax: +321 632 6732.
E-mail addresses: christophe.croux@econ.kuleuven.ac.be (C. Croux), kristel.joossens@econ.kuleuven.ac.be (K. Joossens).

doi:10.1016/j.jmva.2004.11.001

distributions, having densities $f_1(x)$ and $f_2(x)$. The higher the value of Q the more likely the new observation has been generated by the first distribution. Taking the log-ratio of the densities yields

$$Q(x) = \log \frac{f_1(x)}{f_2(x)}.$$

For f_1 a normal density with mean μ_1 and covariance matrix Σ_2, and for f_2 another normal density with parameters μ_2 and Σ_2, one gets

$$Q(x)=\frac{1}{2}\left\{(x-\mu_2)^t\Sigma_2^{-1}(x-\mu_2)-(x-\mu_1)^t\Sigma_1^{-1}(x-\mu_1)\right\}+\frac{1}{2}\log\left(\frac{|\Sigma_2|}{|\Sigma_1|}\right). \tag{1.1}$$

Here, $|\Sigma|$ stands for the determinant of a square matrix Σ. The above equation can be written as a quadratic form

$$Q(x) = x^t Ax + b^t x + c, \tag{1.2}$$

where

$$A = \tfrac{1}{2}(\Sigma_2^{-1} - \Sigma_1^{-1}), \tag{1.3}$$

$$b = \Sigma_1^{-1}\mu_1 - \Sigma_2^{-1}\mu_2, \tag{1.4}$$

$$c = \frac{1}{2}\log\left(\frac{|\Sigma_2|}{|\Sigma_1|}\right) + \frac{1}{2}(\mu_2^t\Sigma_2^{-1}\mu_2 - \mu_1^t\Sigma_1^{-1}\mu_1). \tag{1.5}$$

The function $Q(x)$ is called the quadratic discriminant function. Although it has been derived from normal densities it can also be applied as such without making distributional assumptions.

Future observations will now be classified according to the following discriminant rule: if $Q(x) > \tau$, where τ is a selected cut-off value, then assign x to the first group. On the other hand if $Q(x) < \tau$, then assign x to the second group. Now let π_1 be the prior probability that an observation to classify will be generated by the first distribution, and set $\pi_2 = 1 - \pi_2$. For normal source distributions it is known that the optimal discriminant rule, in the sense of minimizing the expected probability of misclassification, is given by the above quadratic rule with $\tau = \log(\pi_2/\pi_1)$, e.g. [21, Chapter 11]. In practice, the prior probabilities π_1 and π_2 are often unknown and one uses $\tau = 0$.

The discriminant function (1.1) still depends on the unknown population quantities μ_1, μ_2, Σ_1 and Σ_2, and needs to be estimated from the training sample. So let $x_1, \ldots, x_{n_1}$ be a sample of p-variate observations coming from the first distribution H_1^0 and $x_{n_1+1}, \ldots, x_n$ a second sample drawn from H_2^0. These samples together constitute the training sample. An observation in the training sample will influence the sample estimates of location and covariance, and hence the discriminant rule. In quadratic discriminant analysis (QDA) the primary interest is not in knowing or interpreting the parameter values in (1.2). The aim is to use QDA for classification purposes. Focus in this paper is on how observations belonging to the training sample affect the total probability of misclassification, and this effect will be quantified by the influence function. Influence functions in the multi-sample setting were already considered by several authors, e.g. [11,14]. In this paper, the formalism of partial