The Korean Communications
in Statistics Vol. 6, No. 1, 1999
pp. 43-51

Local Influence in Quadratic Discriminant Analysis

Kang-Mo Jung[1)]

Abstract

The local influence method is adapted to quadratic discriminant analysis for the identification of influential observations affecting the estimation of probability density function, probabilities and log odds. The method allows a simultaneous perturbation on all observations so that it can identify multiple influential observations. The proposed method is applied to a real data set, and satisfactory result is obtained.

1. Introduction

Discriminant analysis is a multivariate technique concerned with separating distinct sets of observations and allocating new observations to previously defined group by a discriminant function. However, the performance of most common discriminant rules is sensitive to outliers or influential observations. Recently, the detection methods of such observations are proposed.

In linear discriminant analysis, Campbell (1978) used the influence function. Critchley and Vitiello (1991) and Fung (1992) independently proposed influence measures based on the case deletion method. Kim (1996b) suggested the local influence method on the misclassification probability and Jung *et al.* (1997b) extended that to the second order local influence method.

In quadratic discriminant analysis, Fung (1996) proposed several measures for detecting influential observations based on the single case deletion. As mentioned in Fung (1996), a single case deletion diagnostic may not identify multiple influential observations due to masking. So further influence analysis is necessary.

Cook (1986) proposed the local influence method to assess the influence of local departures from the assumptions in statistical methods or from the observed data. The curvature of the likelihood displacement was used for measuring the influence of local departures. Wu and Luo (1993) extended Cook's approach to the second order local influence method. The local influence method allows a simultaneous perturbation on all observations. Our experience with numerical examples reveals that the local influence method is very effective in detecting jointly multiple influential observations and that it avoids the masking effect. In this work, the second order local influence method is adapted to quadratic discriminant analysis for the purpose of investigating the influence of observations on the estimation of probability density function, probabilities and log odds.

---

1) Full-time lecturer, Department of Computer Science and Statistics, Kunsan National University, 68 Miryong-Dong, Kunsan, Chollapuk-Do, 573-701, Korea.

Section 2 contains a brief review of quadratic discriminant analysis and some perturbed statistics for investigating observations that affect significantly the performance of the quadratic discriminant rule. In Section 3, under a simultaneous perturbation the local influence method is explicitly derived. A numerical example is illustrated in Section 4.

2. Local Influence Diagnostics

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n_1}$ and $\boldsymbol{x}_{n_1+1}, \ldots, \boldsymbol{x}_{n_1+n_2}$ $(n = n_1 + n_2)$ be two independent random samples from p-variate normal density distributions $N_p(\boldsymbol{\mu}_1, \Sigma_1)$ and $N_p(\boldsymbol{\mu}_2, \Sigma_2)$, respectively. The quadratic discriminant rule assigns an observation $\boldsymbol{x}$ to population 1 if

$$(\boldsymbol{x} - \boldsymbol{\mu}_1)^T \Sigma_1^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_1) + \log|\Sigma_1| < (\boldsymbol{x} - \boldsymbol{\mu}_2)^T \Sigma_2^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_2) + \log|\Sigma_2|, \quad (1)$$

and to population 2 otherwise. If $\boldsymbol{\mu}_i, \Sigma_i, i = 1,2$ are unknown, the parameters in (1) are estimated by the usual sample mean vector $\overline{\boldsymbol{x}}_i$ and the sample covariance matrix S_i with divisor n_i.

We consider a simultaneous perturbation on all observations coming from two populations. It is totally different from the perturbation of the influence function method and the case deletion method, which can perturb a single observation. The perturbed model is specified by a perturbation vector $\boldsymbol{w} = (w_1, \ldots, w_n)^T$ in which

$$\boldsymbol{x}_r \sim N_p(\boldsymbol{\mu}_i, \Sigma_i / w_r), \quad r = 1, \ldots, n, \quad (2)$$

where the group index i is changed according to the observation index r. The perturbation vector $\boldsymbol{w}$ can be written as $\boldsymbol{w} = \mathbf{1}_n + a\boldsymbol{l}$, where $\boldsymbol{l} = (l_1, \ldots, l_n)^T$ denotes the direction vector of unit length, scalar a indicates the magnitude of perturbation, and $\mathbf{1}_n \equiv (1, \ldots, 1)^T$ of order n. When $a = 0$, that is $\boldsymbol{w} = \mathbf{1}_n$, the perturbed model reduces to the unperturbed model.

A serious weakness of the quadratic rule is that it is sensitive to departures from normality (Johnson and Wichern, 1992). For a measure of atypicality Fung (1996) used the Mahalanobis squared distance of an observation from its group mean, which is defined by $d_j^2 = (\boldsymbol{x}_j - \tilde{\boldsymbol{x}}_j)^T \tilde{S}_j^{-1} (\boldsymbol{x}_j - \tilde{\boldsymbol{x}}_j)$. Here $\tilde{\boldsymbol{x}}_j = \overline{\boldsymbol{x}}_1$ if $j \in I_1 \equiv \{1, \ldots, n_1\}$ and $\tilde{\boldsymbol{x}}_j = \overline{\boldsymbol{x}}_2$ if $j \in I_2 \equiv \{n_1 + 1, \ldots, n_1 + n_2\}$, and $\tilde{S}_j$ is similarly defined. Kim (1996a) proposed the local influence method on normality using the likelihood displacement approach. To assess the local influence on normality, we consider the perturbed statistic defined as

$$D_i(\boldsymbol{w}) = \frac{1}{n_i} \sum_{j \in I_i} (d_j^2 - d_j^2(\boldsymbol{w}))^2, \quad i = 1, 2, \quad (3)$$

where $d_j^2(\boldsymbol{w})$ is the Mahalanobis squared distance under the perturbed model (2). That is,

$d_j^2(w)=(x_j-\widehat{\mu}_1(w))^T S_1(w)^{-1}(x_j-\widehat{\mu}_1(w))$ if $j \in I_1$, where $\widehat{\mu}_1(w)$ and $S_1(w)$ are the maximum likelihood estimators of μ_1 and Σ_1 under (2), respectively. For $j \in I_2$, $d_j^2(w)$ is similarly defined. This statistic, however, is defined for individual populations only and does not consider the specific structure of discriminant analysis.

We extend the diagnostic measures proposed by Fung (1996) to generally perturbed statistics such as the square of relative probability (RPSQ), the square of relative log odds (RLOSQ), the square of relative log odds for individual group (IRLOSQ) defined below.

$$RPSQ(w) = \frac{1}{n}\sum_{j=1}^{n}[\hat{p}_1(x_j)-\hat{p}_1(x_j, w)]^2,$$

$$RLOSQ(w) = \frac{1}{n}\sum_{j=1}^{n}\left[\log\left\{\frac{\hat{p}_1(x_j)}{\hat{p}_2(x_j)}\right\}-\log\left\{\frac{\hat{p}_1(x_j, w)}{\hat{p}_2(x_j, w)}\right\}\right]^2,$$

$$IRLOSQ_i(w) = \frac{1}{n_i}\sum_{j\in I_i}[\log \hat{f}_i(x_j)-\log \hat{f}_i(x_j, w)]^2, \quad i=1,2,$$

where $\hat{f}_i(x)=(2\pi)^{-p/2}|S_i|^{-1/2}\exp\{-(x-\bar{x}_i)^T S_i^{-1}(x-\bar{x}_i)/2\}$ is the plug-in estimate of the normal density, $\hat{p}_i(x)=\hat{f}_i(x)/(\hat{f}_1(x)+\hat{f}_2(x))$ is the estimated probability that observation x belongs to group i, $\hat{f}_i(x, w)$ and $\hat{p}_i(x, w)$ are the estimates corresponding to $\hat{f}_i(x)$ and $\hat{p}_i(x)$ under the perturbed model (2), respectively. More precisely, $\hat{f}_i(x, w)$ is obtained by substituting $\widehat{\mu}_i(w)$, $S_i(w)^{-1}$ into $\bar{x}_i$, S_i^{-1} in $\hat{f}_i(x)$, and thus $\hat{p}_i(x, w)=\hat{f}_i(x, w)/(\hat{f}_1(x, w)+\hat{f}_2(x, w))$.

Let $Q(w)$ be the perturbed statistic under the perturbation (2), for example, $D_i(w)$ or $RPSQ(w)$, etc. Then the $(n+1)$ by 1 vector $a(w)=(w^T, Q(w))^T$ forms a surface in the $(n+1)$-dimensional space as w varies over a certain space. Since $D_i(w)$ and $IRLOSQ_i(w)$ are not affected by the perturbation of the other population, the dimension n reduces to n_i. The direction vector of the maximum slope of a path on the surface $a(w)$ at $a=0$ is considered for investigating the local behaviour of observations in quadratic discriminant analysis. Further influence information can be obtained through curvatures on the surface. The first and second direction vectors, namely $l_{\max}$ and $l_{\sec}$, respectively, of the surface at $w = 1_n$ corresponding to the largest and second largest absolute curvatures, yield information about influential observations. Observations corresponding to significantly large direction cosines of the two direction vectors above in its absolute value can be influential. Here the absolute value of the curvature is required since the curvature can be negative.

In the local influence method the influence of observations can be investigated as follows. The first and second order partial derivatives of $Q(w)$ with respect to a evaluated at $a=0$

are

$$\frac{\partial Q(\boldsymbol{w})}{\partial a}\Big|_{a=0} = \sum_{r=1}^{n}\left(\frac{\partial Q(\boldsymbol{w})}{\partial w_r}\Big|_{\boldsymbol{w}=\mathbf{1}_n}\right)l_r ,$$

$$\frac{\partial^2 Q(\boldsymbol{w})}{\partial a^2}\Big|_{a=0} = \sum_{s=1}^{n}\sum_{r=1}^{n}\left(\frac{\partial^2 Q(\boldsymbol{w})}{\partial w_s \partial w_r}\Big|_{\boldsymbol{w}=\mathbf{1}_n}\right)l_r l_s .$$

The curvature and its associated direction vector of the surface at $\boldsymbol{w} = \mathbf{1}_n$ (refer to equations (2.2) to (2.5) in Wu and Luo (1993)) are obtained by solving the generalized eigenvalue problem

$$|\boldsymbol{F} - \beta \boldsymbol{G}| = 0 , \tag{4}$$

where $\boldsymbol{F}$ is the $n \times n$ matrix having $\partial^2 Q(\boldsymbol{w})/\partial w_s \partial w_r|_{\boldsymbol{w}=\mathbf{1}_n}$ as its (r, s)th element, $\boldsymbol{G} = (1 + \dot{\boldsymbol{Q}}^T \dot{\boldsymbol{Q}})^{1/2}(\boldsymbol{I}_n + \dot{\boldsymbol{Q}} \dot{\boldsymbol{Q}}^T)$, and $\dot{\boldsymbol{Q}}$ is the $n \times 1$ column vector whose rth element is $\partial Q(\boldsymbol{w})/\partial w_r|_{\boldsymbol{w}=\mathbf{1}_n}$. The curvature of the surface is given by the eigenvalue in (4) and the direction vector $\boldsymbol{l}$ is its associated eigenvector of unit length. This comes from the fact that the curvature is equivalent to the value of $\boldsymbol{l}^T \boldsymbol{F} \boldsymbol{l} / \boldsymbol{l}^T \boldsymbol{G} \boldsymbol{l}$. For the perturbed statistics $D_i(\boldsymbol{w})$, $RPSQ(\boldsymbol{w})$, $RLOSQ(\boldsymbol{w})$ and $IRLOSQ(\boldsymbol{w})$, $\boldsymbol{G}$ becomes $\boldsymbol{I}_n$, because $\dot{\boldsymbol{Q}} = \mathbf{0}$ which will be shown in Section 3. Hence $\boldsymbol{l}_{max}$ and $\boldsymbol{l}_{sec}$ are the eigenvectors corresponding to the largest and second largest absolute eigenvalue of $\boldsymbol{F}$, respectively.

3. Derivation

Consider the quadratic discriminant function for two populations in (1). The perturbed statistics proposed in Section 2 involve $\widehat{\boldsymbol{\mu}}_i(\boldsymbol{w})$, $\boldsymbol{S}_i(\boldsymbol{w})$, $\boldsymbol{S}_i(\boldsymbol{w})^{-1}$ and $|\boldsymbol{S}_i(\boldsymbol{w})|$, $i = 1, 2$. To solve the generalized eigenvalue problem (4), we need the first and second order partial derivatives of the perturbed maximum likelihood estimators described above evaluated at $\boldsymbol{w} = \mathbf{1}_n$. However, it is not necessary to get the second order partial derivatives of the perturbed maximum likelihood estimators, which are not involved in the second order partial derivatives of $D_i(\boldsymbol{w})$, $RPSQ(\boldsymbol{w})$, $RLOSQ(\boldsymbol{w})$ and $IRLOSQ(\boldsymbol{w})$ evaluated at $\boldsymbol{w} = \mathbf{1}_n$. It can be seen in Subsection 3.1 and 3.2.

The maximum likelihood estimators $\widehat{\boldsymbol{\mu}}_i(\boldsymbol{w})$, $\boldsymbol{S}_i(\boldsymbol{w})$ under the perturbation scheme (2) are obtained by

$$\widehat{\boldsymbol{\mu}}_i(\boldsymbol{w}) = \sum_{j \in I_i} w_j \boldsymbol{x}_j / \sum_{j \in I_i} w_j, \tag{5}$$

$$\boldsymbol{S}_i(\boldsymbol{w}) = \frac{1}{n_i} \sum_{j \in I_i} w_j (\boldsymbol{x}_j - \widehat{\boldsymbol{\mu}}_i(\boldsymbol{w}))(\boldsymbol{x}_j - \widehat{\boldsymbol{\mu}}_i(\boldsymbol{w}))^T. \tag{6}$$

By differentiating (5) and (6) with respect to w_r and putting $\mathbf{w} = \mathbf{1}_n$, the first order partial derivatives of $\widehat{\boldsymbol{\mu}}_i(\mathbf{w})$, $S_i(\mathbf{w})$ evaluated at $\mathbf{w} = \mathbf{1}_n$ become

$$\widehat{\boldsymbol{\mu}}_{i,r} \equiv \frac{\partial}{\partial w_r} \widehat{\boldsymbol{\mu}}_i(\mathbf{w})\Big|_{\mathbf{w}=\mathbf{1}_n} = \frac{I_i(r)}{n_i}(\mathbf{x}_r - \bar{\mathbf{x}}_i), \tag{7}$$

$$S_{i,r} \equiv \frac{\partial}{\partial w_r} S_i(\mathbf{w})\Big|_{\mathbf{w}=\mathbf{1}_n} = \frac{I_i(r)}{n_i}(\mathbf{x}_r - \bar{\mathbf{x}}_i)(\mathbf{x}_r - \bar{\mathbf{x}}_i)^T. \tag{8}$$

where $I_i(r)$ be one if $r \in I_i$ and zero otherwise. Hereafter, such notation as $\widehat{\boldsymbol{\mu}}_{i,r}$ denotes the first order partial derivative of a perturbed statistic with respect to w_r evaluated at $\mathbf{w} = \mathbf{1}_n$. We have known that $\partial S_i(\mathbf{w})/\partial w_r = - S_i(\mathbf{w})^{-1}(\partial S_i(\mathbf{w})/\partial w_r) S_i(\mathbf{w})^{-1}$, and (8) gives immediately

$$S_{i,r}^{-1} = -\frac{I_i(r)}{n_i} S_i^{-1}(\mathbf{x}_r - \bar{\mathbf{x}}_i)(\mathbf{x}_r - \bar{\mathbf{x}}_i)^T S_i^{-1}. \tag{9}$$

Let $C(\mathbf{w})$ be the perturbed matrix of a $p \times p$ symmetric matrix C with $C(\mathbf{1}_n) = C$. Then the determinant of $C(\mathbf{w})$ can be rewritten as $|C(\mathbf{w})| = \prod_{j=1}^{p} \lambda_j(\mathbf{w})$, where $\lambda_j(\mathbf{w})$ is the perturbed eigenvalue of jth largest eigenvalue λ_j for the matrix $C(\mathbf{1}_n)$. Thus $\partial |C(\mathbf{w})|/\partial w_r|_{\mathbf{w}=\mathbf{1}_n} = |C| \sum_{j=1}^{p} \lambda_{j,r}/\lambda_j$, where $\lambda_{j,r} = \partial \lambda_j(\mathbf{w})/\partial w_r|_{\mathbf{w}=\mathbf{1}_n}$. Following (6) in Jung *et al.* (1997a) leads to

$$\frac{\partial}{\partial w_r} |C(\mathbf{w})|\Big|_{\mathbf{w}=\mathbf{1}_n} = |C| \sum_{j=1}^{p} \sum_{k=1}^{p} C_{jk,r} C_{jk}^{-1}, \tag{10}$$

where $C_{jk,r}$ is the (j,k)th element of $\partial C(\mathbf{w})/\partial w_r|_{\mathbf{w}=\mathbf{1}_n}$, C_{jk}^{-1} is the (j,k)th element of C^{-1}. Consequently, (8) and (10) give

$$|S_i|_r = I_i(r) \frac{|S_i|}{n_i} (\mathbf{x}_r - \bar{\mathbf{x}}_i)^T S_i^{-1} (\mathbf{x}_r - \bar{\mathbf{x}}_i). \tag{11}$$

3.1. Mahalanobis Squared Distance

To get the curvature of the perturbed surface $(\mathbf{w}^T, D_i(\mathbf{w}))^T$ at $\mathbf{w} = \mathbf{1}_n$ we need the first and second order partial derivatives of $D_i(\mathbf{w})$ evaluated at $\mathbf{w} = \mathbf{1}_n$.

The first order partial derivative of $D_i(\mathbf{w})$ with respect to w_r can be written as

$$D_{i,r}(\mathbf{w}) \equiv \frac{\partial}{\partial w_r} D_i(\mathbf{w}) = -\frac{2}{n_i} \sum_{j \in I_i} (d_j^2 - d_j^2(\mathbf{w})) \frac{\partial d_j^2(\mathbf{w})}{\partial w_r}.$$

Then $D_{i,r}(\mathbf{w})|_{\mathbf{w}=\mathbf{1}_n} = 0$, for $d_j^2(\mathbf{w}) = d_j^2$ when $\mathbf{w} = \mathbf{1}_n$. The chain rule of the differentiation yields the second order partial derivatives of $D_i(\boldsymbol{w})$ with respect to w_r, w_s evaluated at $\boldsymbol{w} = \mathbf{1}_n$ such as

$$D_{i,rs} \equiv \frac{\partial^2}{\partial w_s \partial w_r} D_i(\boldsymbol{w})\bigg|_{\boldsymbol{w}=\mathbf{1}_n} = \frac{2}{n_i} \sum_{j \in I_i} d^2_{j,r} d^2_{j,s}.$$

Since $d^2_{j,r} = -2(\boldsymbol{x}_j - \bar{\boldsymbol{x}}_i)^T S_i^{-1} \hat{\boldsymbol{\mu}}_{i,r} + (\boldsymbol{x}_j - \bar{\boldsymbol{x}}_i)^T S_{i,r}^{-1}(\boldsymbol{x}_j - \bar{\boldsymbol{x}}_i)$ for $j \in I_i$, (7) and (9) imply

$$D_{i,rs} = \frac{2}{n_i^3} I_i(r) I_i(s) \sum_{j \in I_i} (2e_{ijr} + e^2_{ijr})(2e_{ijs} + e^2_{ijs}),$$

where $e_{ijr} = (\boldsymbol{x}_r - \bar{\boldsymbol{x}}_i)^T S_i^{-1}(\boldsymbol{x}_j - \bar{\boldsymbol{x}}_i)$. In what follows, such expressions as $D_{i,rs}$ are interpreted as the second order partial derivatives of the perturbed statistics with respect to w_r, w_s evaluated at $\boldsymbol{w} = \mathbf{1}_n$.

Since $D_{i,r}$ is always zero, the generalized eigenvalue problem (4) reduces to the eigenvalue problem. The eigenvectors corresponding to the largest and the second largest absolute eigenvalue for the $n_i \times n_i$ matrix having the (r,s)th element $D_{i,rs}$ form local influence diagnostics.

3.2. RPSQ, RLOSQ, IRLOSQ

As described in Subsection 3.1, it can be easily shown that the first order partial derivatives of $RPSQ(\boldsymbol{w})$, $RLOSQ(\boldsymbol{w})$, $IRLOSQ(\boldsymbol{w})$ evaluated at $\boldsymbol{w} = \mathbf{1}_n$ are zero.

The second order partial derivative of the perturbed statistic $RPSQ(\boldsymbol{w})$ becomes

$$\begin{aligned} RPSQ_{rs} &= \frac{2}{n} \sum_{j=1}^{n} \hat{p}_{1,r}(\boldsymbol{x}_j)\, \hat{p}_{1,s}(\boldsymbol{x}_j) \\ &= \frac{2}{n}(-1)^{k+k'} I_k(r) I_{k'}(s) \sum_{j=1}^{n} \frac{(\hat{f}_1(\boldsymbol{x}_j)\hat{f}_2(\boldsymbol{x}_j))^2}{(\hat{f}_1(\boldsymbol{x}_j) + \hat{f}_2(\boldsymbol{x}_j))^4} \frac{\hat{f}_{k,r}(\boldsymbol{x}_j)}{\hat{f}_k(\boldsymbol{x}_j)} \frac{\hat{f}_{k',s}(\boldsymbol{x}_j)}{\hat{f}_{k'}(\boldsymbol{x}_j)}, \end{aligned}$$

for $k, k' = 1,2$. Since

$$\begin{aligned} 2\frac{\hat{f}_{k,r}(\boldsymbol{x})}{\hat{f}_k(\boldsymbol{x})} &= 2\frac{\partial}{\partial w_r} \log \hat{f}_k(\boldsymbol{x}, \boldsymbol{w})\bigg|_{\boldsymbol{w}=\mathbf{1}_n} \\ &= -\frac{|S_k|_r}{|S_k|} + 2(\boldsymbol{x} - \bar{\boldsymbol{x}}_k)^T S_k^{-1} \hat{\boldsymbol{\mu}}_{k,r} - (\boldsymbol{x} - \bar{\boldsymbol{x}}_k)^T S_{k,r}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}_k), \end{aligned}$$

therefore from (7), (9) and (11) we get

$$RPSQ_{rs} = \frac{1}{2n^3}(-1)^{k+k'} I_k(r) I_{k'}(s) \sum_{j=1}^{n} \frac{(\hat{f}_1(\boldsymbol{x}_j)\hat{f}_2(\boldsymbol{x}_j))^2}{(\hat{f}_1(\boldsymbol{x}_j) + \hat{f}_2(\boldsymbol{x}_j))^4} e'_{kjr} e'_{k'js},$$

where $e'_{kjr} = e_{krr} - 2e_{kjr} - e^2_{kjr}$.

In the same manner, we can show that

$$RLOSQ_{rs} = \frac{1}{2n^3}(-1)^{k+k'} I_k(r) I_{k'}(s) \sum_{j=1}^{n} e'_{kjr}\, e'_{k'js},$$
$$IRLOSQ_{i,rs} = \frac{1}{2n_i^3} I_i(r) I_i(s) \sum_{j \in I_i} e'_{ijr} e'_{ijs}.$$

Note that the terms $D_{i,rs}$, $RPSQ_{rs}$, $RLOSQ_{rs}$ and $IRLOSQ_{i,rs}$ do not require the second order partial derivatives of the perturbed maximum likelihood estimators under the perturbed model (2) evaluated at $\boldsymbol{w} = \mathbf{1}_n$. The maximum slope direction vector in the surface by the perturbed model is always zero, and so diagnostics through curvature is desired.

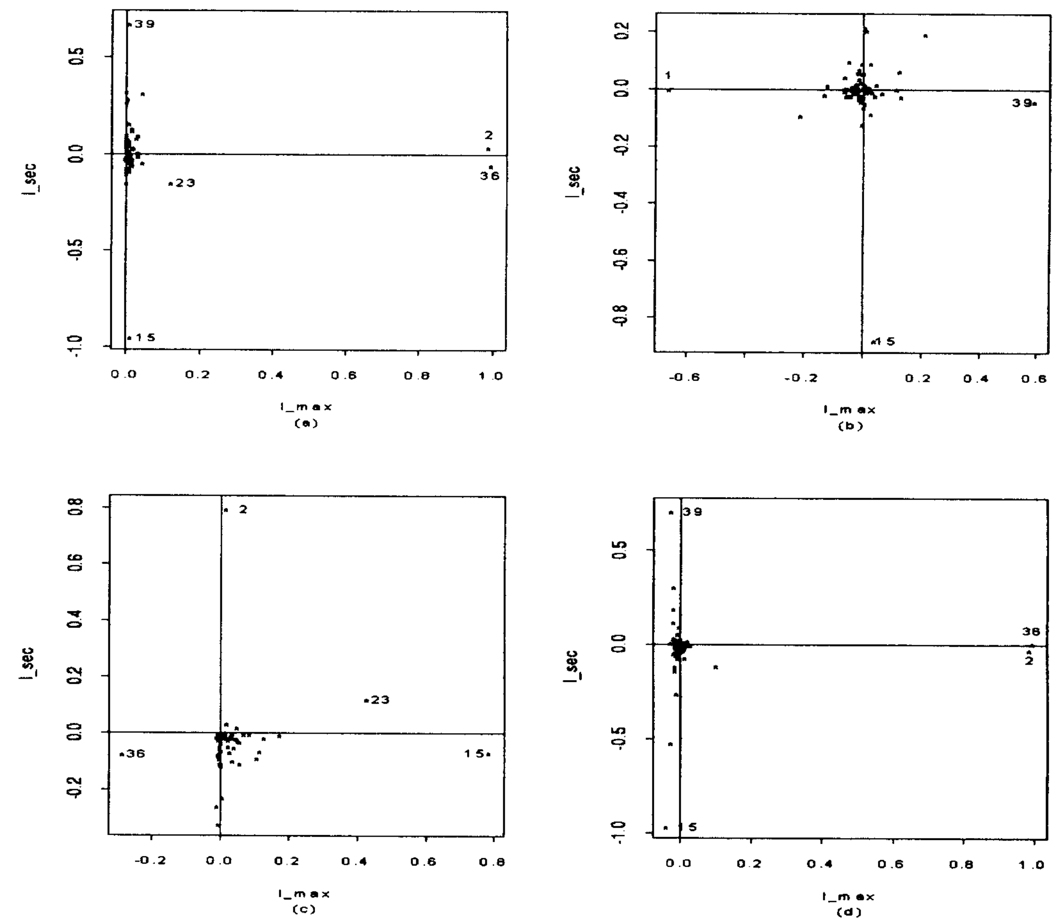

Fig. 1. The scatter plot of $\boldsymbol{l}_{\max}$ versus $\boldsymbol{l}_{\sec}$ for the surface (a) $(\boldsymbol{w}^T, D_i(\boldsymbol{w}))^T$ (b) $(\boldsymbol{w}^T, RPSQ(\boldsymbol{w}))^T$ (c) $(\boldsymbol{w}^T, RLOSQ(\boldsymbol{w}))^T$ (d) $(\boldsymbol{w}^T, IRLOSQ_i(\boldsymbol{w}))^T$.

4. Numerical Example

The local influence method is applied to two species of biting flies (Johnson and Wichern, 1992, p. 282), Leptoconops cartei and Leptoconops torrens. The two species were thought for many years to be the same because they are very similar morphologically. The number of observations for each group is 35 which have measurements on seven variables. The observations are labelled as 1 to 35 for Leptoconops cartei and 36 to 70 for Leptoconops torrens. This data set has also been studied by Fung (1996). His conclusion is that it is reasonable to view observations 1, 2, 15, 23, 36 as influential.

The scatter plot of $\boldsymbol{l}_{\max}$ versus $\boldsymbol{l}_{\sec}$ for the surface $(\boldsymbol{w}^T, D_i(\boldsymbol{w}))^T$ for $i=1,2$ is shown in Fig. 1 (a) overlaid by the direction vectors for each population. From Fig. 1 (a), we can observe that observations 2, 15, 36, 39 are influential. The result of another local influence method on normality (Kim (1996a)) is similar to Fig. 1 (a), so we omit it.

We have also get the results for *RPSQ*, *RLOSQ* and *IRLOSQ*, which are presented in Fig. 1 (b) to (d), respectively. From these figures we may conclude that observations 1, 2, 15, 23, 36, 39 are influential in quadratic discriminant rule (1). Observation 39 seems to be influential, for $\hat{p}_2(\boldsymbol{x}_{39})$ is changed from 0.222 to 0.873 after removing observation 1 (See Table 2 of Fung (1996)). It reveals that observation 39 may be masked by observation 1. The local influence method can identify observation 39 as well as observations 1, 2, 15, 23, 36.

References

[1] Campbell, N. A. (1978). The influence function as an aid in outlier detection in discriminant analysis, *Applied Statistics*, Vol. 27, 251-258.

[2] Cook, R. D. (1986). Assessment of local influence (with discussions), *Journal of Royal Statistical Society, Series B*, Vol. 48, 133-169.

[3] Critchley, F. and Vitiello, C. (1991). The influence of observations on misclassification probability estimates in linear discriminant analysis, *Biometrika*, Vol. 78, 677-690.

[4] Fung, W.-K. (1992). Some diagnostic measures in discriminant analysis, *Statistics & Probability Letters*, Vol. 13, 279-285.

[5] Fung, W.-K. (1996). Diagnosing influential observations in quadratic discriminant analysis, *Biometrics*, Vol. 52, 1235-1241.

[6] Johnson, R. A. and Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis, 3rd ed.*, Prentice Hall, Englewood Cliffs, New Jersey.

[7] Jung, K.-M., Kim, M. G. and Kim, B. C. (1997a). Local influence in maximum likelihood factor analysis, *Computational Statistics & Data Analysis*, Vol. 24, 483-491.

[8] Jung, K.-M., Kim, M. G. and Kim, B. C. (1997b). Second order local influence in linear discriminant analysis, *Journal of the Japanese Society of Computational Statistics*,

Vol. 10, 1-11.

[9] Kim, M. G. (1996a). Local influence in multivariate normal data, *Journal of Applied Statistics*, Vol. 23, 479-485.

[10] Kim, M. G. (1996b). Local influence on misclassification probability, *Journal of the Korean Statistical Society*, Vol. 25, 145-151.

[11] Wu, X. and Luo, Z. (1993). Second-order approach to local influence, *Journal of Royal Statistical Society, Series B*, Vol. 55, 929-936.