Journal of the Korean Statistical Society (1998), **27**: 4, pp 471–483

Local Influence Assessment of the Misclassification Probability in Multiple Discriminant Analysis

Kang-Mo Jung [1]

ABSTRACT

The influence of observations on the misclassification probability in multiple discriminant analysis under the equal covariance assumption is investigated by the local influence method. Under an appropriate perturbation we can get information about influential observations and outliers by studying the curvatures and the associated direction vectors of the perturbation-formed surface of the misclassification probability. We show that the influence function method gives essentially the same information as the direction vector of the maximum slope. An illustrative example is given for the effectiveness of the local influence method.

Keywords: Influence function; Influential observations; Local influence; Multiple discriminant analysis; Perturbation

1. INTRODUCTION

In linear discriminant analysis, diagnostic methods based on the influence functions and omission approaches have been suggested for detecting influential observations and outliers. Campbell (1978) used the influence function method. Critchley and Vitiello (1991), and Fung (1992) independently proposed two fundamental statistics, on which many influence measures in two-group discriminant analysis depend. Recently, Fung (1996) considered the influence function method on the misclassification probability in multiple discriminant analysis.

The local influence method was introduced by Cook (1986) as a general method of assessing the influence of minor perturbations of the model and was adapted to two-group discriminant analysis by Kim (1996). Kim (1996) considered the local influence method based on the direction vector of the maximum slope of a path on the surface formed by the perturbed maximum likelihood estimator of the misclassification probability. Wu and Luo (1993) called Cook's

[1]Department of Computer Science and Statistics, Kunsan National University, 68 Miryong-Dong, Kunsan, Chollapuk-Do 573-701, Korea.

likelihood displacement approach the first order local influence method and they extended that to the case of a variable, for example the maximum likelihood estimator of a parameter. It is so-called the second order local influence method based on the direction vectors corresponding to relatively large local maximum curvatures. Jung et al. (1997) studied the second order local influence method in two-group discriminant analysis, and also showed that the local influence measures include the two statistics proposed by Fung (1992).

In this work the second order local influence method is adapted to multiple discriminant analysis for the purpose of investigating the influence of observations on the misclassification probability in multiple discriminant analysis. The misclassification probability in multiple discriminant analysis is not easily obtained, and so the local influence method in two-group discriminant analysis can not be directly extended to multiple discriminant analysis. In Section 2 we describe the local influence procedure in multiple discriminant analysis. More detailed computations for the local influence method are included in Section 3. It will be shown that the first order local influence measure gives the same influence information as the empirical influence function. Also, we can see that the results of this paper include those of Jung et al. (1997) when the number of populations is two. In Section 4 a numerical example is given to show the effectiveness of the local influence method. This example shows that the local influence method gives useful information about influential observations and outliers, even when the influence function method does not detect significantly influential observations and outliers.

2. LOCAL INFLUENCE PROCEDURE

Suppose that there are m populations $(\pi_1, \ldots, \pi_m)$ with multinormal probability distribution $\mathbf{N}_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ $(i = 1, \ldots, m)$ having common nonsingular covariance matrix $\boldsymbol{\Sigma}$. Assume that the prior probabilities for each population π_i are equal and the cost function is constant. Then the minimum expected cost of misclassification rule allocates $\mathbf{x}$ from unknown source to π_i if $\mathbf{x} \in R_i$, where

$$R_i = \bigcap_{j \neq i} \{\mathbf{x} | (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - (\boldsymbol{\mu}_i + \boldsymbol{\mu}_j)/2) > 0\}. \tag{2.1}$$

Let the squared Mahalanobis distance Δ_{ij}^2 between populations π_i and π_j be $(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$. Then the misclassification probability, for $\mathbf{x} \in \pi_i$, under the allocation rule (2.1) is

$$MP_i = 1 - G(\mathbf{b}_{-i}), \tag{2.2}$$

where $\mathbf{b}_{-i} = (b_{i1}, \dots, b_{i,i-1}, b_{i,i+1}, \dots, b_{im})^T$ and $b_{ij} = \Delta_{ij}/2$. Here $G(\mathbf{b}_{-i})$ is the cumulative probability under multinormality, that is, $P[\cap_{j \neq i}(z_j < b_{ij})]$, where z_j's are standard normal random variables with $\rho_{jk} = cov(z_j, z_k) = (\Delta_{ij}^2 + \Delta_{ik}^2 - \Delta_{jk}^2)/(2\Delta_{ij}\Delta_{ik})$. See Proposition 2.1 of Fung (1996) for details. The overall misclassification probability is defined as $EMP = \sum_{i=1}^m MP_i/m$.

Let $\mathbf{x}_t$ $(t = 1, \dots, n)$ be the random sample from p-variate normal distribution $N_p(\boldsymbol{\mu}_k, \boldsymbol{\Sigma})$ if $\mathbf{x}_t$ is coming from π_k. Let I_k be the index set of observations coming from π_k and the indicator function $I_k(t)$ be one if $t \in I_k$ and zero otherwise. Then the maximum likelihood estimator of $\boldsymbol{\Sigma}$ is

$$\widehat{\boldsymbol{\Sigma}} = \sum_{k=1}^{m} \sum_{t=1}^{n} I_k(t)(\mathbf{x}_t - \overline{\mathbf{x}}_k)(\mathbf{x}_t - \overline{\mathbf{x}}_k)^T/n,$$

where $\overline{\mathbf{x}}_k$ is the sample mean vector of π_k.

Let $\mathbf{w} = (w_1, \dots, w_n)^T$ be a given n by 1 vector of perturbations. We consider the perturbation model in which the tth observation $\mathbf{x}_t$ coming from π_k is perturbed according to

$$\mathbf{x}_t \sim \mathbf{N}_p(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}/w_t), \tag{2.3}$$

for $t = 1, \dots, n$ (Kim, 1996). The perturbed model reduces to the unperturbed model when $\mathbf{w} = \mathbf{1}_n \equiv (1, \dots, 1)^T$ of order n. The perturbation vector can be expressed as $w_t = 1 + al_t$ $(t = 1, \dots, n)$, where scalar a indicates the magnitude of the perturbation in the direction $\boldsymbol{l} = (l_1, \dots, l_n)^T$.

Let θ be a parameter of interest, for example MP_i or EMP. The maximum likelihood estimator of θ under the perturbed model is denoted by $\widehat{\theta}(\mathbf{w})$. Then the $(n+1)$ by 1 vector $\tau(\mathbf{w}) = (\mathbf{w}^T, \widehat{\theta}(\mathbf{w}))^T$ forms a surface in the $(n+1)$-dimensional space as $\mathbf{w}$ varies over a certain space. The direction vector of the maximum slope of a path on the surface $\tau(\mathbf{w})$ at $a = 0$ is considered for investigating the local behaviour of observations for the estimator of θ. However, it will be shown in Section 3 that for the misclassification probability in multiple discriminant analysis the direction vector of the maximum slope is essentially the same as the empirical influence function. The direction vectors corresponding to relatively large local curvatures of the surface at $a = 0$ provide information about influential observations and outliers. It is main diagnostics that observations corresponding to significantly large direction cosines of the direction vectors in its absolute value can be influential (Wu and Luo (1993)).

The influence of observations on $\widehat{\theta}$ can be investigated as follows. The first and second order partial derivatives of $\widehat{\theta}(\mathbf{w})$ with respect to a evaluated at $a = 0$ are

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

The direction vector of the maximum slope of a path on the surface $\tau(\mathbf{w})$ at $a = 0$ becomes $\boldsymbol{l}_{slope} \equiv \dot{\boldsymbol{\theta}}/(\dot{\boldsymbol{\theta}}^T\dot{\boldsymbol{\theta}})^{1/2}$ by the Cauchy-Schwarz inequality, where $\dot{\boldsymbol{\theta}}$ is the n by 1 vector whose rth component is $\partial\widehat{\theta}(\mathbf{w})/\partial w_r \mid_{\mathbf{w}=\mathbf{1}_n}$. The curvature and its associated direction vector of the surface at $\mathbf{w} = \mathbf{1}_n$ (refer to equations (2.2) to (2.5) in Wu and Luo (1993)) are obtained by solving the generalized eigenvalue problem

$$|\mathbf{U} - \beta\mathbf{V}| = 0, \tag{2.4}$$

where $\mathbf{U}$ is the n by n matrix having $\partial^2\widehat{\theta}(\mathbf{w})/\partial w_s \partial w_r \mid_{\mathbf{w}=\mathbf{1}_n}$ as its (r, s)th element, $\mathbf{V} = (1 + \dot{\boldsymbol{\theta}}^T\dot{\boldsymbol{\theta}})^{1/2}(\mathbf{I}_n + \dot{\boldsymbol{\theta}}\dot{\boldsymbol{\theta}}^T)$, and $\mathbf{I}_n$ is the identity matrix of order n. The curvature of the surface is given by the eigenvalue in (2.4) 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 $\boldsymbol{l}^T\mathbf{U}\boldsymbol{l}/\boldsymbol{l}^T\mathbf{V}\boldsymbol{l}$.

Equation (2.4) provides a way of classifying influential observations by eigenvector directions. Let $\boldsymbol{l}_{max}$ and $\boldsymbol{l}_{sec}$ be the eigenvectors corresponding to the largest and the second largest absolute eigenvalue in (2.4). Then the scatter plot of $\boldsymbol{l}_{max}$ versus $\boldsymbol{l}_{sec}$ may be helpful in finding large direction cosines of direction vectors corresponding to relatively large local curvatures.

3. DERIVATION

In multiple discriminant analysis, the parameters of interest are the squared Mahalanobis distance Δ_{ij}^2 between π_i and π_j, and the misclassification probability MP_i for an observation coming from π_i. They all involve $\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\mu}_k$ for $k = 1, \ldots, m$. Hence to get the local influence measures obtained by solving the generalized eigenvalue problem (2.4), we need the first and second order partial derivatives of the perturbed maximum likelihood estimators $\widehat{\boldsymbol{\mu}}_k(\mathbf{w}), \mathbf{S}(\mathbf{w})$ and $\mathbf{S}(\mathbf{w})^{-1}$ evaluated at $\mathbf{w} = \mathbf{1}_n$.

The maximum likelihood estimators of $\boldsymbol{\mu}_k, \boldsymbol{\Sigma}$ under the perturbed model (2.3) are obtained as

$$\widehat{\boldsymbol{\mu}}_k(\mathbf{w}) = \sum_{t\in I_k} w_t \mathbf{x}_t / \sum_{t\in I_k} w_t, \tag{3.1}$$

$$\mathbf{S}(\mathbf{w}) = \sum_{t=1}^{n}\sum_{k=1}^{m} w_t I_k(t)(\mathbf{x}_t - \widehat{\boldsymbol{\mu}}_k(\mathbf{w}))(\mathbf{x}_t - \widehat{\boldsymbol{\mu}}_k(\mathbf{w}))^T/n. \tag{3.2}$$

By differentiating equations (3.1) to (3.2) with respect to w_r and putting $\mathbf{w} = \mathbf{1}_n$, the first order partial derivatives of $\widehat{\boldsymbol{\mu}}_k(\mathbf{w})$ and $\mathbf{S}(\mathbf{w})$ evaluated at $\mathbf{w} = \mathbf{1}_n$ are given by

$$\widehat{\boldsymbol{\mu}}_{k,r} \equiv \left.\frac{\partial \widehat{\boldsymbol{\mu}}_k(\mathbf{w})}{\partial w_r}\right|_{\mathbf{w}=\mathbf{1}_n} = I_k(r)(\mathbf{x}_r - \overline{\mathbf{x}}_k)/n_k, \tag{3.3}$$

$$\mathbf{S}_r \equiv \left.\frac{\partial \mathbf{S}(\mathbf{w})}{\partial w_r}\right|_{\mathbf{w}=\mathbf{1}_n} = I_k(r)(\mathbf{x}_r - \overline{\mathbf{x}}_k)(\mathbf{x}_r - \overline{\mathbf{x}}_k)^T/n, \tag{3.4}$$

where n_k is the number of observations coming from π_k. Further differentiation gives the second order partial derivatives evaluated at $\mathbf{w} = \mathbf{1}_n$ as

$$\widehat{\boldsymbol{\mu}}_{k,rs} \equiv \left.\frac{\partial^2 \widehat{\boldsymbol{\mu}}_k(\mathbf{w})}{\partial w_r w_s}\right|_{\mathbf{w}=\mathbf{1}_n} = -I_k(r)I_k(s)(\mathbf{x}_r + \mathbf{x}_s - 2\overline{\mathbf{x}}_k)/n_k^2, \tag{3.5}$$

$$\mathbf{S}_{rs} = -\frac{I_k(r)I_k(s)}{nn_k}[(\mathbf{x}_r - \overline{\mathbf{x}}_k)(\mathbf{x}_s - \overline{\mathbf{x}}_k)^T + (\mathbf{x}_s - \overline{\mathbf{x}}_k)(\mathbf{x}_r - \overline{\mathbf{x}}_k)^T]. \tag{3.6}$$

In what follows, such expressions as the subscripts r and rs of an estimator are interpreted as the first and second order partial derivatives of the perturbed maximum likelihood estimator evaluated at $\mathbf{w} = \mathbf{1}_n$, respectively.

Furthermore, the first and second order partial derivatives of $\mathbf{S}(\mathbf{w})^{-1}$ are easily found by $\partial\mathbf{S}(\mathbf{w})^{-1}/\partial w_r = -\mathbf{S}(\mathbf{w})^{-1}(\partial\mathbf{S}(\mathbf{w})/\partial w_r)\mathbf{S}(\mathbf{w})^{-1}$. From (3.4) and (3.6), we obtain

$$\mathbf{S}_r^{-1} = -\mathbf{S}^{-1}\mathbf{S}_r\mathbf{S}^{-1}, \tag{3.7}$$

$$\begin{aligned}\mathbf{S}_{rs}^{-1} = {} & \frac{1}{n^2}\varphi_{kk',rs}\mathbf{S}^{-1}[(\mathbf{x}_r - \overline{\mathbf{x}}_k)(\mathbf{x}_s - \overline{\mathbf{x}}_{k'})^T + (\mathbf{x}_s - \overline{\mathbf{x}}_{k'})(\mathbf{x}_r - \overline{\mathbf{x}}_k)^T]\mathbf{S}^{-1} \\ & -\mathbf{S}^{-1}\mathbf{S}_{rs}\mathbf{S}^{-1},\end{aligned} \tag{3.8}$$

where $\varphi_{kk',rs} = (\mathbf{x}_r - \overline{\mathbf{x}}_k)^T\mathbf{S}^{-1}(\mathbf{x}_s - \overline{\mathbf{x}}_{k'})I_k(r)I_{k'}(s)$.

For getting the local influence measures for the misclassification probability $\widehat{MP}_i$, we begin to derive the first and second order partial derivatives of the perturbed maximum likelihood estimator $\widehat{\Delta}^2_{ij}(\mathbf{w})$. By the invariance of the maximum likelihood estimator, we get

$$\widehat{\Delta}^2_{ij}(\mathbf{w}) = (\widehat{\boldsymbol{\mu}}_i(\mathbf{w}) - \widehat{\boldsymbol{\mu}}_j(\mathbf{w}))^T \mathbf{S}(\mathbf{w})^{-1}(\widehat{\boldsymbol{\mu}}_i(\mathbf{w}) - \widehat{\boldsymbol{\mu}}_j(\mathbf{w})). \tag{3.9}$$

The chain rule of differentiation and equations (3.3) to (3.8) yield

$$\widehat{\Delta}^2_{ij,r} = \psi_{ijk,r}\left[\frac{2}{n_k}(I_i(r) - I_j(r)) - \frac{1}{n}\psi_{ijk,r}\right], \tag{3.10}$$

$$\begin{aligned}\widehat{\Delta}^2_{ij,rs} = &-\frac{2}{n_k^2}(\psi_{ijk,r} + \psi_{ijk,s})(I_i(r) - I_j(r))I_k(r)I_k(s) + \frac{2}{nn_k}\psi_{ijk,r}\psi_{ijk,s} \\ &+2\varphi_{kk',rs}\left[\frac{1}{n_k}(I_i(r) - I_j(r)) - \frac{1}{n}\psi_{ijk,r}\right]\left[\frac{1}{n_{k'}}(I_i(s) - I_j(s)) - \frac{1}{n}\psi_{ijk',s}\right]\end{aligned} \tag{3.11}$$

where $\psi_{ijk,r} = (\mathbf{x}_r - \overline{\mathbf{x}}_k)^T\mathbf{S}^{-1}(\overline{\mathbf{x}}_i - \overline{\mathbf{x}}_j)I_k(r)$.

Under the perturbation scheme (2.3), the misclassification probability (2.2) can be written as

$$\widehat{MP}_i(\mathbf{w}) = 1 - G(\widehat{\mathbf{b}}_{-i}(\mathbf{w})), \tag{3.12}$$

where $\widehat{\mathbf{b}}_{-i}(\mathbf{w}) = (\widehat{b}_{i1}(\mathbf{w}), \ldots, \widehat{b}_{i,i-1}(\mathbf{w}), \widehat{b}_{i,i+1}(\mathbf{w}), \ldots, \widehat{b}_{im}(\mathbf{w}))^T$, and $G(\cdot)$ is the estimated cumulative probability. Hereafter the notation $\widehat{G}_i$ means $G(\widehat{\mathbf{b}}_{-i}(\mathbf{w}))$. To get the local influence measure for $\widehat{MP}_i$, we need the first and second order partial derivatives of $\widehat{MP}_i(\mathbf{w})$ evaluated at $\mathbf{w} = \mathbf{1}_n$. By the invariance of the maximum likelihood estimator and the chain rule of differentiation for (3.12), we get

$$\widehat{MP}_{i,r} = -\frac{1}{4}\sum_{j\neq i}^{m}\frac{\widehat{\Delta}^2_{ij,r}}{\widehat{\Delta}_{ij}}\widehat{G}_{i,j}, \tag{3.13}$$

where $\widehat{G}_{i,j} = \partial\widehat{G}_i/\partial\widehat{b}_{ij}(\mathbf{w})|_{\mathbf{w}=\mathbf{1}_n} = \int\cdots\int_{\mathcal{D}_{-ij}} f(\mathbf{z}_{-ij}, \widehat{b}_{ij})d\mathbf{z}_{-ij}$, and $\mathcal{D}_{-ij} = \{\mathbf{y} = (y_k) \in \mathbb{R}^{m-2} | y_k < \widehat{b}_{ik}, k = 1, \ldots, m, k \neq i, k \neq j\}$. Here $f(\cdot)$ is the joint density function of the $(m-1)$th order multinormal random variable with zero mean vector and the (j,k)th covariance element $\widehat{\rho}_{jk}$, and $\mathbf{z}_{-ij}$ denotes the $(m-2)$ by 1 vector such as $(z_1, \ldots, z_{i-1}, z_{i+1}, \cdots, z_{j-1}, z_{j+1}, \cdots, z_m)^T$.

Furthermore, the second order partial derivatives of $\widehat{MP}_i(\mathbf{w})$ with respect to w_r, w_s evaluated at $\mathbf{w} = \mathbf{1}_n$ is

$$\widehat{MP}_{i,rs} = -\sum_{j\neq i}^{m}\sum_{k\neq i}^{m} \frac{\widehat{\Delta}^2_{ij,r}\widehat{\Delta}^2_{ik,s}}{16\widehat{\Delta}_{ij}\widehat{\Delta}_{ik}}\widehat{G}_{i,jk} - \sum_{j\neq i}^{m}\widehat{G}_{i,j}\left(\frac{\widehat{\Delta}^2_{ij,rs}}{4\widehat{\Delta}_{ij}} - \frac{\widehat{\Delta}^2_{ij,r}\widehat{\Delta}^2_{ij,s}}{8\widehat{\Delta}^3_{ij}}\right), \quad (3.14)$$

where $\widehat{G}_{i,jk} = \partial^2\widehat{G}_i/\partial\widehat{b}_{ik}(\mathbf{w})\partial\widehat{b}_{ij}(\mathbf{w})|_{\mathbf{w}=\mathbf{1}_n}$. Here $\widehat{G}_{i,jk}$ becomes the $(m-3)$ dimensional integral of the function $f(\mathbf{z}_{-ijk}, \widehat{b}_{ij}, \widehat{b}_{ik})$ over $\mathcal{D}_{-ijk}$, if $j \neq k$ and $\partial\widehat{G}_{i,j}/\partial\widehat{b}_{ij}$, if $j = k$, where $\mathcal{D}_{-ijk}$ is similarly defined as $\mathcal{D}_{-ij}$.

The first and second order partial derivatives for the perturbed overall misclassification probability at $\mathbf{w} = \mathbf{1}_n$ are given by $\widehat{EMP}_r = \sum_{i=1}^m \widehat{MP}_{i,r}/m$ and $\widehat{EMP}_{rs} = \sum_{i=1}^m \widehat{MP}_{i,rs}/m$, respectively.

In particular, when $m = 3$ we can obtain the explicit formulas for $\widehat{G}_{i,j}$ and $\widehat{G}_{i,jk}$ as following

$$\begin{aligned} \widehat{G}_{i,j} &= \int_{-\infty}^{\widehat{b}_{ik'}} f_{k'}(z, \widehat{b}_{ij})dz \\ &= (2\pi)^{-1/2}\exp(-\widehat{b}^2_{ij}/2)\Phi\left((\widehat{b}_{ik'} - \widehat{b}_{ij}\widehat{\rho}_{jk'})/\sqrt{1-\widehat{\rho}^2_{jk'}}\right), \\ \widehat{G}_{i,jk} &= \begin{cases} f_k(\widehat{b}_{ik}, \widehat{b}_{ij}), & \text{if } j \neq k, \\ \int_{-\infty}^{\widehat{b}_{ik'}} f_{k'}(z, \widehat{b}_{ij})\frac{\widehat{\rho}_{jk'}z - \widehat{b}_{ij}}{1-\widehat{\rho}^2_{jk'}}dz, & \text{if } j = k, \end{cases} \end{aligned}$$

where $f_{k'}(z, \widehat{b}_{ij}) = (|1-\widehat{\rho}^2_{jk'}|4\pi^2)^{-1/2}\exp[-(z^2 - 2\widehat{b}_{ij}\widehat{\rho}_{jk'}z + \widehat{b}^2_{ij})/(2(1-\widehat{\rho}^2_{jk'}))]$ and $\Phi(\cdot)$ is the cumulative probability function of the standard normal variate. Here, $j \neq i, k \neq i$ and k' is neither i nor j. Furthermore, we have for $j = k$,

$$\widehat{G}_{i,jk} = -\widehat{b}_{ij}\widehat{G}_{i,j} - \frac{\widehat{\rho}_{jk'}}{2\pi(1-\widehat{\rho}^2_{jk'})^{1/2}}\exp\left[-\frac{\widehat{b}^2_{ij}}{2} - \frac{(\widehat{b}_{ik'} - \widehat{b}_{ij}\widehat{\rho}_{jk'})^2}{2(1-\widehat{\rho}^2_{jk'})}\right].$$

In case $m = 2$, since $\widehat{G}_{i,j} = \phi(\widehat{b}_{ij})$ and $\widehat{G}_{i,jj} = -\phi(\widehat{b}_{ij})\widehat{b}_{ij}$, where $\phi(\cdot)$ is the probability density function of the standard normal variate, the equations (3.13) and (3.14) become (20) and (21) of Jung et al. (1997), respectively.

Fung (1996) calculated the empirical influence functions for $\widehat{\Delta}^2_{ij}$ and $\widehat{MP}_i$ as following. For $\mathbf{x}_r \in \pi_k$,

$$I(\mathbf{x}_r, \widehat{\Delta}^2_{ij}) = \begin{cases} n_k(\widehat{\Delta}^2_{ij} - \psi^2_{ijk,r})/n, & \text{if } i \neq k, j \neq k, \\ n_k\widehat{\Delta}^2_{ij}/n + 2\psi_{ijk,r} - n_k\psi^2_{ijk,r}/n, & \text{if } i = k, \\ n_k\widehat{\Delta}^2_{ij}/n - 2\psi_{ijk,r} - n_k\psi^2_{ijk,r}/n, & \text{if } j = k, \end{cases} \quad (3.15)$$

$$I(\mathbf{x}_r, \widehat{MP}_i) = -\sum_{j\neq i}\frac{\partial G(\widehat{b}_{i1}, \ldots, \widehat{b}_{i,i-1}, \widehat{b}_{i,i+1}, \ldots, \widehat{b}_{im})}{\partial\widehat{b}_{ij}}\frac{I(\mathbf{x}_r, \widehat{\Delta}^2_{ij})}{4\widehat{\Delta}_{ij}}. \quad (3.16)$$

Comparing (3.10) with (3.15) gives that $\widehat{\Delta}^2_{ij,r} = I(\mathbf{x}_r, \widehat{\Delta}^2_{ij})/n_k - \widehat{\Delta}^2_{ij}/n$ for $\mathbf{x}_r$ coming from π_k. That is, $\widehat{\Delta}^2_{ij,r}$ is equivalent to $I(\mathbf{x}_r, \widehat{\Delta}^2_{ij})$, up to constants. It implies that $\boldsymbol{l}_{slope}$ for $\widehat{\Delta}^2_{ij}$ gives the same information as the empirical influence function $I(\mathbf{x}_r, \widehat{\Delta}^2_{ij})$. Furthermore, comparison (3.13) with (3.16) yields that $I(\mathbf{x}_r, \widehat{MP}_i)$ is proportional to $\widehat{MP}_{i,r}$, so we conclude that the empirical influence function on the overall misclassification probability provides the same influence information as the first order local influence method.

4. NUMERICAL EXAMPLE

The local influence method is applied to the famous iris data (Johnson and Wichern, 1992, p. 566) for our illustration. For clarity only two variables, sepal width and petal width, are selected so that we can easily detect the sources of influence from the scatter plot of the data. The observations are labelled as 1 to 50 for iris setosa, 51 to 100 for iris versicolor and 101 to 150 for iris virginica. The local influence method is compared with the influence function method.

First the local influence method was conducted based on the estimates of the misclassification probabilities, $\widehat{MP}_i$ $(i = 1, 2, 3)$ and $\widehat{EMP}$. Here $\widehat{MP}_1$, $\widehat{MP}_2$, $\widehat{MP}_3$ mean the misclassification probability for iris setosa population, iris versicolor, iris virginica, respectively. The scatter plot of $\boldsymbol{l}_{max}$ and $\boldsymbol{l}_{sec}$ gave the same behaviour as the index plot of $\boldsymbol{l}_{slope}$ for all $\widehat{MP}_i$ and $\widehat{EMP}$. As described in the previous section, the index plot of $\boldsymbol{l}_{slope}$ gives the same information as the influence function method. Thus the results from the first and second order local influence method agree with those of the influence function method. We omit the results.

To illustrate the effectiveness of the local influence method, observation 51 such that (3.2, 1.4) is changed to (7.2, 1.4). The modified data together with each sample mean denoted by M and the discriminant lines are given in Fig. 4.1. We find that observations from iris setosa lie far from the other two populations. Fig. 4.1 reveals that observation 51 can be regarded as an outlier from π_2, and observations 51, 71, 78, 120, 130, 134 and 135 are misclassified. Observations 69 and 84 are located near the discriminant plane for π_2 and π_3. These observations may be influential, for observations near the discriminant plane may be misclassified by a minor change.

The first order local influence measures for each $\widehat{MP}_i$ $(i = 1, 2, 3)$ are investigated. The absolute magnitude $|\widehat{MP}_{1,r}|$ $(r = 1, \ldots, 150)$ is much smaller than the other two group measures. So we omit the index plot of $\boldsymbol{l}_{slope}$ for $\widehat{MP}_1$. The first

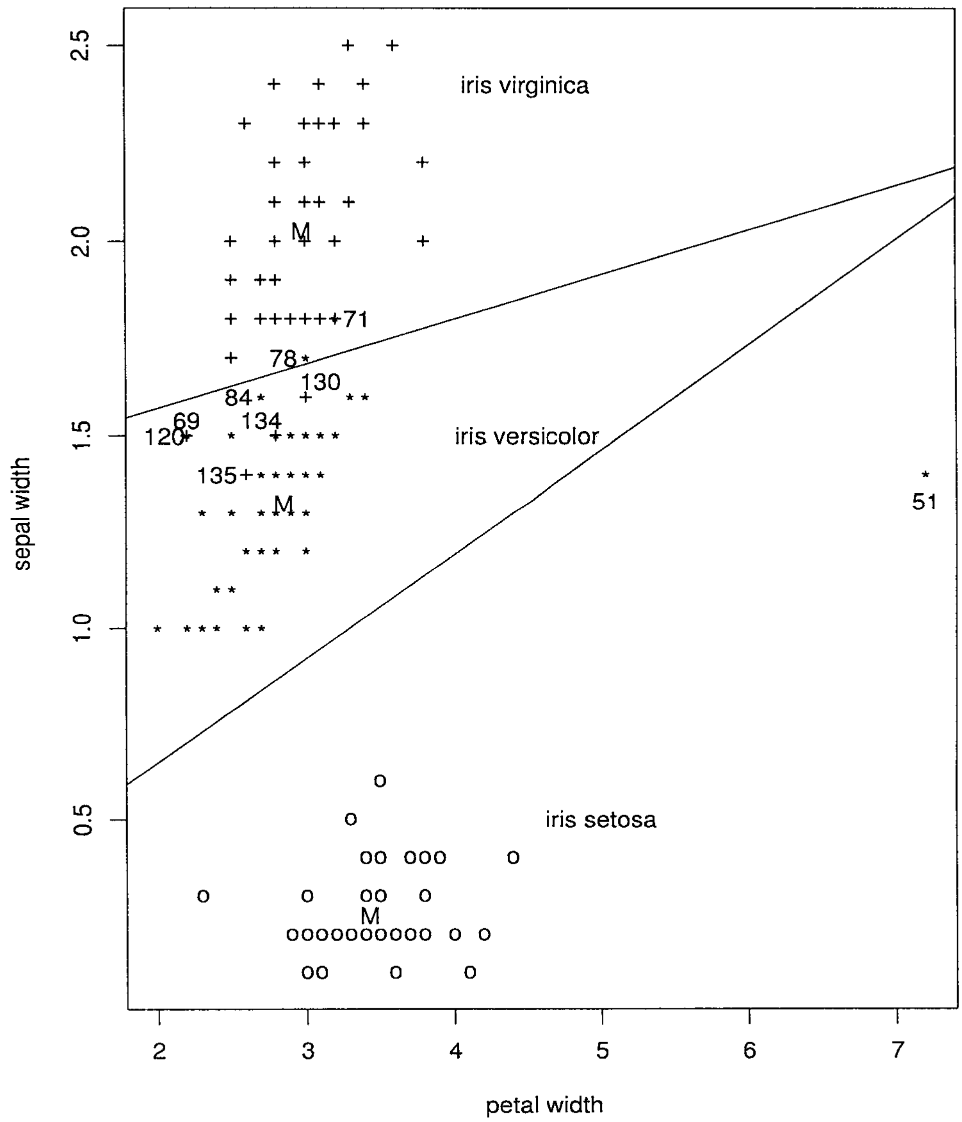
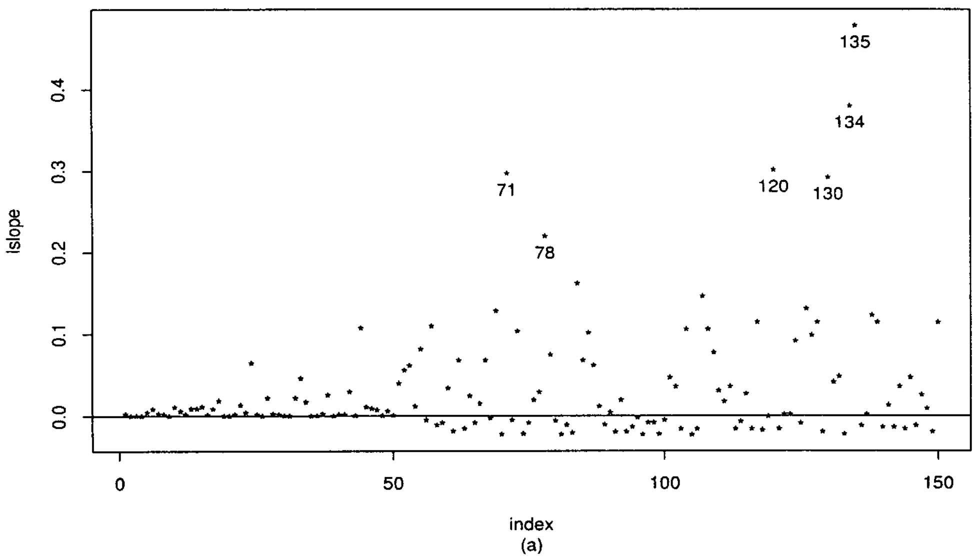


Fig. 4.1: The modified data of sepal and petal width of iris setosa, iris versicolor and iris virginica.



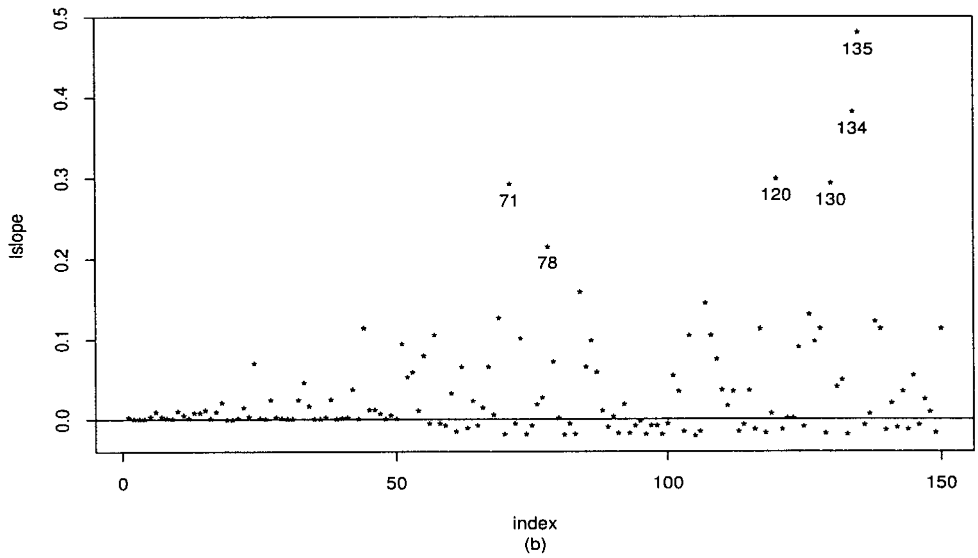


Fig. 4.2: Index plots of $\boldsymbol{l}_{slope}$ for the modified iris data. (a) $\widehat{MP}_3$ (b) $\widehat{EMP}$

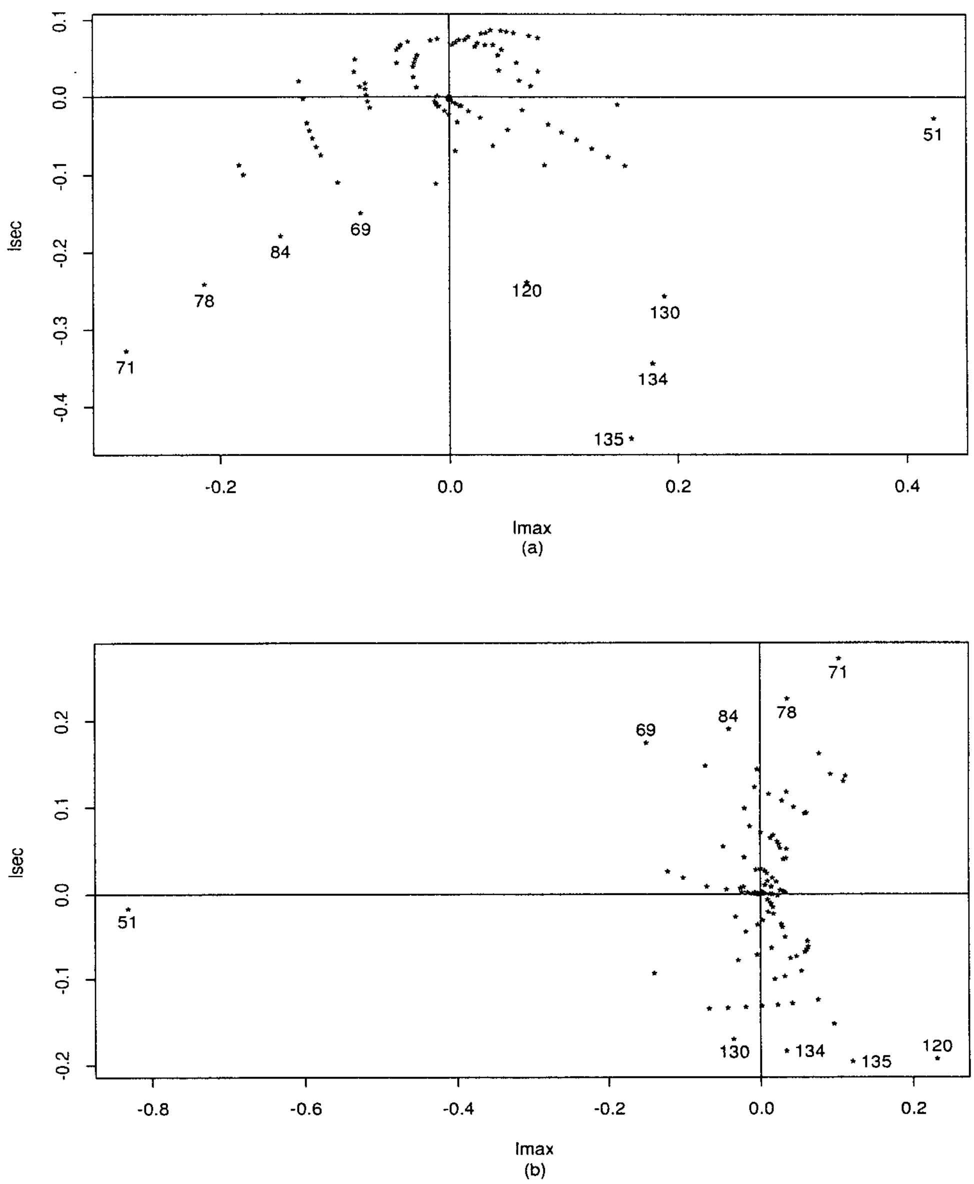


Fig. 4.3: Scatter plots of $\boldsymbol{l}_{max}$ versus $\boldsymbol{l}_{sec}$ for the modified iris data. (a) $\widehat{MP}_3$ (b) $\widehat{EMP}$

order local influence measures $\widehat{MP}_{2,r}$ and $\widehat{MP}_{3,r}$ are nearly identical, and so only the maximum slope vector $\boldsymbol{l}_{slope}$ for $\widehat{MP}_3$ is plotted in Fig. 4.2(a). For the overall misclassification $\widehat{EMP}$, the index plot of $\boldsymbol{l}_{slope}$ is given in Fig. 4.2(b). In these index plots most of misclassified observations are detected, however observations 51, 69, 84 may not be identified. It indicates that the first order local influence measure is not sufficient for detecting influential observations and outliers.

Finally the second order local influence method is performed for $\widehat{MP}_3$ and $\widehat{EMP}$. For $\widehat{MP}_3$ the first four absolutely largest eigenvalues of (2.4) are -0.00171, 0.00159, -0.00157, -0.00075. The fourth value is more or less smaller than the other values. Thus the scatter plots of $\boldsymbol{l}_{max}$ versus $\boldsymbol{l}_{sec}$, $\boldsymbol{l}_{max}$ versus $\boldsymbol{l}_{thr}$, and $\boldsymbol{l}_{sec}$ versus $\boldsymbol{l}_{thr}$ may be helpful for detecting influential observations and outliers. Here the direction vectors $\boldsymbol{l}_{max}, \boldsymbol{l}_{sec}, \boldsymbol{l}_{thr}$ are the eigenvectors corresponding to -0.00171, 0.00159, -0.00157, respectively. Because all three scatter plots have similar results, only the scatter plot of $\boldsymbol{l}_{max}$ versus $\boldsymbol{l}_{sec}$ of $\widehat{MP}_3$ is presented in Fig. 4.3 (a). For $\widehat{EMP}$, the eigenvalues of (2.4) are arranged as -0.00132, -0.00114, 0.00107, -0.00042. By the same reason above, the scatter plot of $\boldsymbol{l}_{max}$ versus $\boldsymbol{l}_{sec}$ is presented in Fig. 4.3 (b). Observation 51 is clearly separated from the others in both scatter plots of Fig. 4.3. Besides observation 51, it can be observed that observations 71, 78, 120, 130, 134, 135 have high local influence, and they are possible candidates for influential observations. Recall that observations 51, 71, 78, 120, 130, 134, 135 are misclassified. And also observations 69, 84 are located in the outer side of Fig. 4.3. The scatter plot of $\boldsymbol{l}_{max}$ versus $\boldsymbol{l}_{sec}$ would provide useful information about the region near the discriminant plane. In summary, the scatter plots of $\boldsymbol{l}_{max}$ versus $\boldsymbol{l}_{sec}$ for $\widehat{MP}_3$ and $\widehat{EMP}$ identify respectively observation 51, 69, 84 (not detected by $\boldsymbol{l}_{slope}$) as well as six misclassified observations.

This example shows that the scatter plot of $\boldsymbol{l}_{max}$ versus $\boldsymbol{l}_{sec}$ in multiple discriminant analysis is very effective, in that it gives influence information about misclassified observations, outliers and observations near the discriminant plane. However, the influence function method is not sufficient for detecting influential observations and outliers.

ACKNOWLEDGEMENTS

The author is grateful to the referees for their thoughtful and invaluable comments and suggestions.

REFERENCES

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

Cook, R. D. (1986). "Assessment of local influence (with discussion)," *Journal of the Royal Statistical Society, Series B*, **48**, 136-169.

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

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

Fung, W.-K. (1996). "The influence of observations on misclassification probability in multiple discriminant analysis," *Communications in Statistics - Theory and Methods*, **25**, 1917-1930.

Johnson, R. A. and Wichern, D. W. (1992). *Applied multivariate statistical analysis, 3 ed.*, Prentice Hall, Englewood Cliffs, New Jersey.

Jung, K.-M. , Kim, M. G. and Kim, B. C. (1997). "Second order local influence in linear discriminant analysis," *Journal of the Japanese Society of Computational Statistics*, **10**, 1-11.

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

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