Pattern Recognition 47 (2014) 2241–2254

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Influence functions for a linear subspace method

Kuniyoshi Hayashi [a,b,*]

[a] Graduate School of Environmental and Life Science, Okayama University, Japan
[b] CREST, Japan Science and Technology Agency, Japan

ARTICLE INFO

Article history:
Received 29 October 2012
Received in revised form
3 October 2013
Accepted 30 November 2013
Available online 27 December 2013

Keywords:
CLAFIC
Cross-validation
Perturbation analysis
Single-case diagnostics

ABSTRACT

A linear subspace method, which is one of discriminant methods, was proposed as a pattern recognition method and was studied. Because the method and its extensions do not encounter the situation of singular covariance matrix, we need not consider extensions such as generalized ridge discrimination, even when treating a high dimensional and sparse dataset. In addition, classifiers based on a multi-class discrimination method can function faster because of the simple decision procedure. Therefore, they have been widely used for face and speech recognition. However, it seems that sufficient studies have not been conducted about the statistical assessment of training data performance for classifier in terms of prediction accuracy. In statistics, influence functions for statistical discriminant analysis were derived and the assessments for analysis result were performed. These studies indicate that influence functions are useful for detecting large influential observations for analysis results by using discrimination methods and they contribute to enhancing the performance of a target classifier.

In this paper, we propose the statistical diagnostics of a classifier on the basis of an influence function by using the linear subspace method. We first propose the discriminant score for the linear subspace method. Next, we derive the sample and empirical influence functions for the average of the discriminant scores to detect large influential observations for the misclassification rate. Finally, through a simulation study and a real data analysis, we detect the outliers in the training dataset using the derived influence function and develop a highly sophisticated classifier in the linear subspace method.

1. Introduction

The demands for discrimination methods that classify unknown high dimensional observations have accelerated in many fields. Watanabe [20,21] proposed a linear subspace method to perform discrimination for a high dimensional pattern vector. If a vector pattern has a relatively large number of features, this method can efficiently solve a discriminant problem with a fast calculation time, even if the number of target classes is large. Therefore, this method has been used in real time recognition. The linear subspace method and its extensions have been studied by numerous authors, including Watanabe and Pakvasa [22], Kittler [12], Karhunen and Oja [10], Oja [16], Kohonen [13], Laaksonen and Oja [14], Maeda and Murase [15], Gülmezoğlu et al. [6], Gunal and Edizkan [5], Kitamura et al. [11], and Yamashita and Wakahara [23]. In the field of machine learning and pattern recognition, there are many powerful classifiers. In pattern recognition, the extensions of CLAFIC method have been widely used in face and speech recognition. For example, kernel nonlinear subspace method is an extended method of CLAIFC method and it was a powerful method relative to the support vector machine [15]. In the subspace method, there are two notable points relative to the support vector machine. One is fast processing in huge multi-class classification. The other is a possibility of a statistical quantitative assessment based on influence function. Therefore, in this paper, we apply a diagnostics based on influence function to a basic linear subspace method, CLAFIC method.

In statistics, influence function has been used to assess the influence of training sample for many target statistics in a lot of multivariate methods, and a statistical sensitivity analysis and diagnostics based on its function have been developed by many researchers. Particularly, the diagnostics of discrimination based on the influence function was discussed by Campbell [1], Fung [4], and Huang et al. [9]. Using an influence function of the important statistics for a classifier, we can detect an influential observation for analysis results. If influential observations are outliers such as mislabeled data, we can largely improve the performance of a classifier by deleting them from the training data. In the field of pattern recognition, as I mentioned before, many extensions for the linear subspace method have been proposed, but the statistical evaluation of their classifiers on the basis of influence functions has not been established yet. In this paper, we focus on the Class-Featuring Information Compression (CLAFIC) method and propose

* Correspondence address: Graduate School of Environmental and Life Science, Okayama University, Japan.
E-mail address: k-hayashi@ems.okayama-u.ac.jp

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.11.030

a novel discriminant score for CLAFIC and its average. In addition, we define an important statistics and derive the empirical and sample influence functions. Through a simulation study and a real data analysis, we show that the diagnostics based on the influence function is also useful for the evaluation of a classifier in the linear subspace method.

2. CLAFIC

We denote the number of feature quantities or variables in a pattern vector $\boldsymbol{x}_i=(x_{i1},\dots,x_{ip})^{\mathrm{T}}$ as p, and we suppose that there are K classes. In addition, we represent a training sample in the k-th class as $\boldsymbol{x}_i^k$, where i ranges from 1 to n_k. n_k is the number of samples in the k-th class. Then, the autocorrelation matrix of the training data in the k-th class is defined as follows:

$$\hat{G}_k=\frac{1}{n_k}\sum_{i=1}^{n_k}\boldsymbol{x}_i^k\boldsymbol{x}_i^{k^{\mathrm{T}}}\quad (k=1,\dots,K). \tag{1}$$

In the first step, we solve the eigenvalue problem for $\hat{G}_k$. Here we denote the p eigenvalues as $\hat{\lambda}_1^k\ge\cdots\ge\hat{\lambda}_p^k\ge 0$ and represent the normalized eigenvector corresponding to $\hat{\lambda}_s^k$ as $\hat{\boldsymbol{u}}_s^k (s=1,\dots,p)$. Next, we calculate a projection matrix in the k-th class as follows:

$$\hat{P}_k=\sum_{s=1}^{p_k}\hat{\boldsymbol{u}}_s^k\hat{\boldsymbol{u}}_s^{k^{\mathrm{T}}}\quad (1\le p_k\le p), \tag{2}$$

where p_k is determined for the minimum value m satisfying $\tau\le(\sum_{s=1}^m\hat{\lambda}_s^k/\sum_{s=1}^p\hat{\lambda}_s^k)(1\le m\le p)$ for a given τ. In this regard, we suppose that the neighboring eigenvalues of $\hat{\lambda}_{p_k}^k$ are not equal. τ corresponds to a cumulative proportion in principal component analysis. Using a cross-validation method, we can determine the optimal value of τ. Next, we project a test observation $\boldsymbol{x}^*$ into all subspaces based on (2) and calculate the squares of the projected norms

$$\boldsymbol{x}^{*^{\mathrm{T}}}\hat{P}_k\boldsymbol{x}^*\quad (k=1,\dots,K). \tag{3}$$

We finally classify $\boldsymbol{x}^*$ into the class that gives the maximum value in (3). In this study, as a matter of convenience, we normalize all training and test data to one.

3. Influence functions

In general data analysis, we necessarily face to an outlier problem that there is gross error in datasets to be analyzed. Therefore, as described in [8], to describe the structure best fitting the bulk of the data and to identify deviating data points or deviating substructures for further treatment are so important for data analysts. Their two things are the main aims of robust statistics [8]. In particular, by using the influence function that was introduced by [7], we can detect outliers for a target statistics. In the field of statistics, a diagnostics based on the influence function has been developed as sensitivity analysis for many statistical multivariate methods [19]. On the other hand, sensitivity analysis based on the influence function for prediction accuracy on cross-validation in discriminant analysis has been less discussed. In general, in discrimination, if the average of the discriminant scores in each class is large, a magnitude of separation between classes is also large. Therefore, in this situation, we can expect that the prediction accuracy in discrimination is improved. The sign of the influence function for its statistics is also important for improving prediction accuracy. However, it seems that an application for the influence function and its sign has not been actively performed for statistical methods and pattern recognition methods. Then, in this section, we focus on a subspace method, CLAFIC method. We calculate the discriminant score for the CLAFIC method and its average. We derive the influence function for the average.

In this section, we introduce three functions: the theoretical, empirical, and sample influence functions.

3.1. Theoretical influence function

Suppose $\theta_k(F^1,\dots,F^K)$ $(k=1,\dots,K)$ are differentiable functionals of F^g, where F^g $(g=1,\dots,K)$ are the cumulative distribution functions (cdfs).

Then, the g-th perturbed cdf at $\boldsymbol{x}^g$ is defined as follows:

$$\tilde{F}^g=(1-\varepsilon)F^g+\varepsilon\delta_{\boldsymbol{x}^g}, \tag{4}$$

where $\delta_{\boldsymbol{x}^g}$ is the cdf of a unit point mass at $\boldsymbol{x}^g$.

The theoretical influence function (*TIF*) of θ_k at $\boldsymbol{x}^g$ is given by

$$TIF(\boldsymbol{x}^g;\theta_k)=\lim_{\varepsilon\to 0}\frac{\theta_k(F^1,\dots,\tilde{F}^g,\dots,F^K)-\theta_k(F^1,\dots,F^g,\dots,F^K)}{\varepsilon}. \tag{5}$$

In the above definition, we assume that the true distribution functions F^g in all classes are known.

3.2. Empirical influence function

The empirical influence function (*EIF*) is obtained by replacing F^g in *TIF* by $\hat{F}^g$, an empirical cdf [2]. With *EIF*, we can also evaluate the influence of each observation in the g-th class on an analysis result. We define

$$\hat{F}^g=\frac{1}{n_g}\sum_{j=1}^{n_g}\delta_{\boldsymbol{x}_j^g}\quad (g=1,\dots,K). \tag{6}$$

Then, *EIF* at a point of $\boldsymbol{x}_r^g(g=1,\dots,K;\ r=1,\dots,n_g)$ is given by

$$EIF(\boldsymbol{x}_r^g;\hat{\theta}_k)=\lim_{\varepsilon\to 0}\frac{\hat{\theta}_k(\hat{F}^1,\dots,\tilde{\hat{F}}^g,\dots,\hat{F}^K)-\hat{\theta}_k(\hat{F}^1,\dots,\hat{F}^g,\dots,\hat{F}^K)}{\varepsilon}, \tag{7}$$

where $\hat{\theta}_k$ is the empirical version of θ_k and $\tilde{\hat{F}}^g=(1-\varepsilon)\hat{F}^g+\varepsilon\delta_{\boldsymbol{x}_r^g}$.

3.3. Sample influence function

To evaluate the influence of observation for $\hat{\theta}_k$, we can use the sample influence function (*SIF*) instead of (7). *SIF* is calculated as follows:

$$SIF(\boldsymbol{x}_r^g;\hat{\theta}_k)=-(n_g-1)\{\hat{\theta}_k(\hat{F}^1,\dots,\tilde{\hat{F}}^g,\dots,\hat{F}^K)-\hat{\theta}_k(\hat{F}^1,\dots,\hat{F}^g,\dots,\hat{F}^K)\}, \tag{8}$$

where $\tilde{\hat{F}}^g=(1+1/(n_g-1))\hat{F}^g-(1/(n_g-1))\delta_{\boldsymbol{x}_r^g}$. The perturbation term $-1/(n_g-1)$ in the sample influence function corresponds to ε in the empirical influence function.

4. Influence functions for CLAFIC

In this section, we develop a discriminant score and its average for CLAFIC. Subsequently, we derive the influence functions for an essential part of the discriminant score.

4.1. Discriminant score

We define a discriminant score for $\boldsymbol{x}_i^k$ in the linear subspace method as follows: