CIMCA 2003 Proceedings/ISBN 1740880684: M. Mohammadian (Ed.)
12-14 February 2003, Vienna - Austria

# Fuzzy Clustered Inverse Regression

Ruck Thawonmas
Intelligent Computer Entertainment Laboratory
Department of Computer Science, Ritsumeikan University
Kusatsu, Shiga, 525-8577, Japan
E-mail: ruck@cs.ritsumei.ac.jp

**Abstract**

*Sliced inverse regression (SIR), introduced by K.C. Li in 1991, is a very general and fast procedure for dimension reduction in regression. For small samples, however, results of SIR are influenced by the choice of slices. In this paper, we propose a fuzzified version of SIR in which slices are replaced by fuzzy clusters. We compare the small sample behavior of the original SIR and the proposed one on simulated data.*

## 1 Introduction

Sliced inverse regression (SIR) is a useful method for extraction of geometric information underlying noisy data of several dimensions. Since being proposed in 1991 [1], due to its computational simplicity, SIR has gained a lot of interests from researchers involved in high dimensional data analysis. Its recent development includes the works in [2, 3].

In theory, SIR has the Fisher consistency property. Namely, it has no estimation bias. However, in practice, e.g., when the sample size is small, the results of SIR are influenced by parameters given by the user, i.e., the number of slices and the slice positions.

In this paper, we address the issue on the positions of slices in SIR, and resolve it by introducing a concept of fuzzy clustering [4] into SIR. Consequently, we propose a method called fuzzy clustered inverse regression (FCIR) that outperforms SIR for small samples.

In the rest of the paper, we review SIR in Section 2. Section 3 describes the proposed FCIR. The small sample behavior of SIR and FCIR are then analyzed on simulated data in Section 4.

## 2 Sliced Inverse Regression

The regression model considered for SIR is as follows:

$$y = g(\beta_1^T \mathbf{x}, \beta_2^T \mathbf{x}, \ldots, \beta_K^T \mathbf{x}, \epsilon), \tag{1}$$

where $y$ is a univariate output variable, $\mathbf{x}$ a p-dimensional variable of interest, $\epsilon$ an error term independent of $\mathbf{x}$ with unknown probability distribution. The superscript $T$ represents transpose. The function $g$ and $K$ p-dimensional vectors $\beta_1, \beta_2, \ldots, \beta_K$ are unknown. The task here is to find space or subspace spanned by $\beta_1, \beta_2, \ldots, \beta_K$.

For the data set $(y_1, \mathbf{x}_1), \ldots, (y_n, \mathbf{x}_n)$ with $(p+1)$ variables and $n$ cases following the above regression model, the algorithm of SIR consists of the following steps.

| sample # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $y_i$ | 0.80 | 0.75 | 0.70 | 0.65 | 0.60 | 0.15 | 0.10 | 0.05 |
| slice 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| slice 2 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |

Table 1: Slicing of samples in SIR. If $y_i \in$ slice $h$, the corresponding element = 1; otherwise, 0.

**Step 1.** Sort the data by $y$ in either increasing order or decreasing one.

**Step 2.** Divide the data set by $y$ into $H$ slices as equally as possible.

**Step 3.** For each slice $h$, compute its sample mean of $\mathbf{x}_i$'s, $\bar{\mathbf{x}}_h = n_h^{-1} \sum_{y_i \in \text{slice } h} \mathbf{x}_i$, where $n_h$ denotes the number of cases in slice $h$.

**Step 4.** Compute the covariance matrix for the slice means of $\mathbf{x}_i$'s, weighted by the slice sizes as follows:

$$\widehat{\sum}_s = n^{-1} \sum_{h=1}^{H} n_h (\bar{\mathbf{x}}_h - \bar{\mathbf{x}})(\bar{\mathbf{x}}_h - \bar{\mathbf{x}})^T,$$

where $\bar{\mathbf{x}}$ denotes the sample mean of $\mathbf{x}_i$'s and hence $\bar{\mathbf{x}} = n^{-1} \sum_{i=1}^{n} \mathbf{x}_i$.

**Step 5.** Compute the sample covariance for $\mathbf{x}_i$'s, $\widehat{\sum}_{\mathbf{x}} = n^{-1} \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$.

**Step 6.** Find the SIR directions by conducting the eigenvalue decomposition of $\widehat{\sum}_s$ with respect to $\widehat{\sum}_{\mathbf{x}}$:

$$\widehat{\sum}_s \hat{\beta}_i = \hat{\lambda}_i \widehat{\sum}_{\mathbf{x}} \hat{\beta}_i, \tag{2}$$

where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_p$, and the $i$-th eigenvector $\hat{\beta}_i$ is called the $i$-th SIR direction.

For Step 2, one can choose slices of equal range or slices of equal (or mostly equal) number of samples. Empirical studies in the literature show that the second approach seems to be better. However, for small samples, this approach poses a problem as illustrated in Table 1, where it is assumed that the output variable $y$ has been already sorted in decreasing order. In this table, we can see that though $y_5$ is nearer to the members of slice 1, $y_1, y_2, \ldots, y_4$, it is assigned to slice 2 in order to make the number of samples in the two slices equal.

## 3 Fuzzy Sliced Inverse Regression

To solve the slicing problem discussed in Section 2, we propose to perform clustering of samples, rather than slicing. A clustering algorithm that we use is Bezdek's fuzzy C-means algorithm [4].

Table 2 gives clustering results for the same samples used in Table 1. The membership degrees of cluster 1 and cluster 2 for $y_5$ represent expected results.

Now, we propose FCIR, a variant of SIR that exploits the fuzzy clustering results. The algorithm of FCIR consists of the following steps.

| sample # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $y_i$ | 0.80 | 0.75 | 0.70 | 0.65 | 0.60 | 0.15 | 0.10 | 0.05 |
| slice 1 | 0.9837 | 0.9962 | 0.9997 | 0.9883 | 0.9883 | 0.0079 | 0.0000 | 0.0057 |
| slice 2 | 0.0163 | 0.0038 | 0.0003 | 0.0117 | 0.0117 | 0.9921 | 1.0000 | 0.9943 |

Table 2: Fuzzy clustering of the same samples in Table 1. Here elements 1 and 0 in Table 1 are replaced by membership degrees of the corresponding clusters.

**Step 1.** Cluster the data by $y$ into $H$ clusters using FCM.

**Step 2.** Let $u_{hi}$ denote the membership degree of cluster $h$ for $y_i$, and $\eta_h = \sum_{i=1}^n u_{hi}$ denote the size of cluster $h$. For each cluster, compute its cluster center in vector space of $\mathbf{x}$, $\bar{\mathbf{v}}_h = \eta_h^{-1} \sum_{i=1}^n u_{hi}\mathbf{x}_i$.

**Step 3.** Compute the covariance matrix for the cluster centers, weighted by the cluster sizes as follows:

$$\widehat{\sum}_f = n^{-1} \sum_{h=1}^H \eta_h(\bar{\mathbf{v}}_h - \bar{\mathbf{x}})(\bar{\mathbf{v}}_h - \bar{\mathbf{x}})^T,$$

where $\bar{\mathbf{x}} = n^{-1} \sum_{i=1}^n \mathbf{x}_i$, and note here that in FCM $n = \sum_{h=1}^H \eta_h$.

**Step 4.** Compute the sample covariance for $\mathbf{x}_i$'s, $\widehat{\sum}_{\mathbf{x}} = n^{-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$.

**Step 5.** Find the FCIR directions by conducting the eigenvalue decomposition of $\widehat{\sum}_f$ with respect to $\widehat{\sum}_{\mathbf{x}}$:

$$\widehat{\sum}_f \hat{\nu}_i = \hat{\theta}_i \widehat{\sum}_{\mathbf{x}} \hat{\nu}_i, \tag{3}$$

where $\hat{\theta}_1 \geq \hat{\theta}_2 \geq \ldots \geq \hat{\theta}_p$, and the $i$-th eigenvector $\hat{\nu}_i$ is called the $i$-th FCIR direction.

## 4 Simulations

In the following two examples, each element of 10-dimensional variable of interest $x_1, x_2, \ldots, x_{10}$ and the error term $\epsilon$ are generated independently from the normal distribution with mean 0 and variance 1.

In the first example, we consider a linear model

$$y = 5 + x_1 + x_2 + x_3 + \epsilon,$$

and in the second example a transformation-inside model

$$y = (5 + x_1 + x_2 + x_3)^2 + \epsilon.$$

It is obvious that $\beta = (1, 1, 1, 0, 0, 0, 0, 0, 0, 0)^T$.

Since the direction of $\beta$ is identifiable, we use the following efficiency measure for the estimated $\hat{b}$ of this direction ($\hat{b} = \hat{\beta}_1$ for SIR and $\hat{b} = \hat{\nu}_1$ for FCIR):

$$e(\hat{b}, \beta) = \frac{(\hat{b}^T \widehat{\sum}_{\mathbf{x}} \beta)^2}{(\hat{b}^T \widehat{\sum}_{\mathbf{x}} \hat{b})(\beta^T \widehat{\sum}_{\mathbf{x}} \beta)}.$$

| $n-H$ | 60-10 | 60-20 | 60-30 | 80-10 | 80-20 | 80-30 | 100-10 | 100-20 | 100-30 |
|---|---|---|---|---|---|---|---|---|---|
| SIR Mean | 0.8445 | 0.7504 | 0.5768 | 0.9211 | 0.8969 | 0.8042 | 0.9605 | 0.9574 | 0.9430 |
| SIR Std. | 0.0846 | 0.1654 | 0.2416 | 0.0400 | 0.0584 | 0.1351 | 0.0191 | 0.0216 | 0.0286 |
| FCIR Mean | 0.8677 | 0.8045 | 0.6524 | 0.9299 | 0.9117 | 0.8732 | 0.9654 | 0.9602 | 0.9510 |
| FCIR Std. | 0.0657 | 0.1175 | 0.2125 | 0.0359 | 0.0481 | 0.0747 | 0.0165 | 0.0198 | 0.0256 |

Table 3: Mean and standard deviation of the efficiency measure of SIR and FCIR for the linear model.

| $n-H$ | 60-10 | 60-20 | 60-30 | 80-10 | 80-20 | 80-30 | 100-10 | 100-20 | 100-30 |
|---|---|---|---|---|---|---|---|---|---|
| SIR Mean | 0.9116 | 0.8592 | 0.7416 | 0.9650 | 0.9565 | 0.9097 | 0.9949 | 0.9971 | 0.9937 |
| SIR Std. | 0.0467 | 0.0937 | 0.1829 | 0.0202 | 0.0264 | 0.0567 | 0.0029 | 0.0019 | 0.0043 |
| FCIR Mean | 0.9225 | 0.8876 | 0.8008 | 0.9688 | 0.9596 | 0.9428 | 0.9967 | 0.9957 | 0.9936 |
| FCIR Std. | 0.0402 | 0.0600 | 0.1270 | 0.0184 | 0.0245 | 0.0358 | 0.0026 | 0.0044 | 0.0067 |

Table 4: Mean and standard deviation of the efficiency measure of SIR and FCIR for the transformation-inside model.

Note that $0 \leq e(\hat{b}, \beta) \leq 1$. The efficiency measure $e(\hat{b}, \beta)$ is 1 when $\hat{b}$ and $\beta$ indicate the same direction. On the contrary, when $\hat{b}$ and $\beta$ are orthogonal to each other, $e(\hat{b}, \beta)$ becomes 0.

Tables 3 and 4 give the mean and standard deviation of the efficiency measure of SIR and FCIR for the above linear model and transformation-inside model, respectively. For each combination of the number of cases $n$ and the number of slices or clusters $H$, 1000 simulated samples were generated.

As we can see from the two tables, FCIR performs better than SIR for small samples, $n = 60$ or 80. Namely, compared to SIR, FCIR gives higher means and lower standard deviations of the efficiency measure. In addition, for small samples, the performance of FCIR drops more gracefully, as $H$ being increased. For larger samples, $n = 100$, their performances become comparable.

## 5 Conclusions

The new method FCIR, a fuzzified version of SIR, that we have proposed appears to perform better than the original SIR for small samples. An issue remained unsolved in the paper is on practical numbers of clusters. For this issue, pooling methods discussed in [3] might be applied to FCIR. How to deal with a case where outliers reside in the data is also a challenging issue. We leave both of them as our future work.

## Acknowledgement

The author was supported in part by the Japan Society for the Promotion of Science (JSPS) under Grant-in-Aid for Encouragement of Young Scientists, Grant Number: 13780309, and by the Ritsumeikan University 's Kyoto Art and Entertainment Innovation Research, a $21^{st}$ Century Center of Excellence Program, funded by JSPS.

## References


[1] K.C. Li, Sliced Inverse Regression for Dimension Reduction, with Discussions, *Journal of the American Statistical Association*, Vol. 86, 316-342, 1991.

[2] K.C. Li, *High Dimensional Data Analysis via the SIR/PHD Approach*, On-line Lecture Notes, `http://www.stat.ucla.edu/~kcli/sir-PHD.pdf` , April 2000.

[3] J. Saracco, Pooled Slicing Methods versus Slicing Methods, *Commun. Statist.- Simula.*, Vol. 30, No. 3, 489-511, 2001.

[4] J.C. Bezdek, *Pattern Recogniton with Fuzzy Objective Function Algorithms*, Plenum, New York, 1981.