Sufficient Dimension Reduction-Based Classification of Nigerian Cities by Crimes against Properties' Safety

K. O. Olorede[1]; W. B. Yahya[2]

[1]Department of Statistics and Mathematical Sciences,
Kwara State Ubniversity, Malete, Nigeria

[2]Department of Statistics,
University of Ilorin, Ilorin, Nigeria.
E-mail: kabir.olorede@kwasu.edu.ng[1]

Abstract **—This study adopts the method of Sufficient Dimension Reduction (SDR) to estimate sufficient predictors for visualizing the data of crimes against pproperties in Nigerian cities and training statical classification models that are capable of efficiently detecting true safety status of such cities without losing information. Modified version of the sliced inverse regression (SIR) methods was adopted by replacing the usual maximum likelihood covariance estimator by the Hybridized Smoothed Maximum Entropy Covariance Estimator (HSMEC) proposed by Olorede and Yahya in the dimension reduction step. All the seven statistical classifiers achieved excellent results based the first sufficient predictor estimated by the modified (SIR-HSMEC) with k-Nearest Neighbour model with one optimal neighbour achiving false positive rate of 0% and 100% classification accuracy, sensitivity, specificity, and area under the curve, respectively.**

Keywords - *Sufficient Dimension Reduction, Sliced Inverse Regression, Crime against Properties, Hybridized Smoothed Covariance Estimator, Maximum Entropy, Sufficient Predictors*

I. Introduction

Certainly, we live in a time when moving is second nature to us. This movement covers many facets of our lives. We move because living in one place for the rest of our lives is not a sentence we wish to serve. One good question, therefore, is that: do we move to a state where safety is not an issue or to a state where safety is a major concern? Are there some crime types we could observe more closely to know if a state is safe or dangerous base on the actual figure of crimes committed therein? What are the states with significant frequencies of such crime types in Nigeria? These are some of the questions that can be answered using efficient classification techniques on suitable data on crime for the safety of everyone.

Consider a classification problem involving a discrete univariate binary response Y and p-dimensional predictor vector $\boldsymbol{X} = (\boldsymbol{X_1},\ \boldsymbol{X_2}, \dots, \boldsymbol{X_p})^{\boldsymbol{T}}$ as in the case while whether a city has hihe or low level of crime safety according to figure of reported cases of such crime type therein. Such crime data may be considered to have high dimension even with fewwer covariates, p, than number of samples, n, as long as p>2. This is because with $\boldsymbol{p > 2}$, the data can no longer be visualized in a 2- or 3-dimension. To enhance visualization and mitigate this high-dimensionaly, tt may be helpful to seek a reduction R(X) that retains most of the information in the original data for predicting Y. This is called dimension reduction. To avoid this obvious loss of information, a reduction R(X) that retains all information in the original data for the regression of Y on X is preferable. This is called sufficient dimension reduction (SDR, Cook, 1998; Li, 1991). The SDR approach assumes that the response variable relates to only a few linear combinations of the many covariates (Ma & Zhu 2013) and seeks a low-dimensional subspace $\mathbb{S}$ of of the original predictor space with $\boldsymbol{d} < \boldsymbol{p}$ such that

$$\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{X} | \mathbb{B}^{\boldsymbol{T}} \boldsymbol{X}, \qquad (1)$$

where $\perp\!\!\!\perp$ denotes statistical independence (Dawid, 1979), and $\mathbb{B}^{\boldsymbol{T}}\boldsymbol{X}$ denotes the orthogonal projection on to $\mathbb{S}$. The dimension reduction subspace $\mathbb{S}(\mathbb{B})$ is the linear space spanned by columns of the loading matrix $\mathbb{B}$. The goal of SDR to identify the central subspace $\mathbb{S}_{Y|X}$ (Cook 1996, 1998, and 1994b), definedned as the column space of $\mathbb{B}$ which satisfies (1) with the smallest number of columns d. Following Cook 2018, the central subspace, $\mathbb{S}_{Y|X}$, is the intersection of all the dimension reduction subspaces when it is itself a dimension reduction subspace. Consequently, if

$\mathbb{B} \in \mathbb{R}^{p\times d}$ **is** a basis of $\mathbb{S}_{Y|X}$, then $\mathbb{B}^T\boldsymbol{X}$ contains all regression information that **X** has about Y.

The sliced Inverse Regression (Li, 1991) method is one of the several inverse regression methods existing in the SDR literature. These methods do not make any assumptions on the distribution of the predictors. Other inverse based SDR methods include the Sliced Average Variance Estimation (SAVE, Cook 2000; Cook & Weisberg 1991), Principal Hessian Directions (PHD, Li ,1992; Cook 1998), the Iterative Hessian Transformation (Cook and Li, 2002), Directional Regression (DR, Li and Wang, 2007), the Inverse Regression Estimation (IRE, Cook and Ni, 2005), and the Minimum Average Variance Estimator (MAVE, Xia, Li, and Zhu, 2002).

Other cousins of of the inverse based SDR methods include the likelihood-based methods such as the Likelihood Acquired Directions (LAD, Cook & Forzani 2009), the Principal Fitted Components (PFC, Cook, 2007; Cook & Forzani 2008b), and the Envelope model (Cook, Li, & Chiaromonte 2007, 2010).

The focus of this study is to demonstrate the effectiveness of the newly proposed Maximum Entropy Covariance estimator (MEC, Olored and Yahya, 2019) as a more prudent way in forming convex covariance mixture to eradicate loss of covariance information to achieve highest classification accuracy in statistical covariance based methods, even oin low-dimensional aaplications without the need for time-consuming optimization. The proposal in this work replaces the maximum likelihood covariance estimate in the dimension reduction step of the SIR with the HSMEC estimator to obtain SIR-HSMEC to achieve improved prediction accuracy and model interpretation through covariance hybridization and entropy maximization.

II. Material And Method

2.1 Data Source and Description

Data set spanning twelve crime variables relating to offences against properties in the year 2013 across the Nigerian cities, as obtained from data base of the Nigeria Police Force (NPF) were employed in this study. The original data set was unsupervised in nature since it contained on number of recorded cases of armed robbery (Arbr), demanding with menace (DWM), theft, burglary, house breaking (HBkg), store breaking (StBrkg), false pretence (FlsPrtc), forgery, receiving stolen property, (RcvgStlnPpt), unlawful possession (UnlfPoss), arson (Arson), and other offences (Other.Off) without explicit supervising outcome (response) variable, Y. The response was manually curated from the recorded cased by finding othe overall average number of reported cased for all the twelve crime variables (137.4257) and the average number of recorded cased per state for the thirty-six states and the federal capital territory (FCT). Any state whose average recorded number of cases for all the variables is larger than the overall average number of cases is considered to have high rate (low safety level) of crimes against properties and hence coded 0 while those with mean recorded cases less than the overall mean recorded cased were considered having low rate (high safety level) and hence coded 1. The dummy codes assigned were due the fact that high safety was the target class. The low safety level was set as baseline. Table 1 further summarized the data.

Table 1: Date Summary

Class Label	Dummy Code	# Sample Per Class	Sample Ratio	# Features
High	0	32	32/37	12 Offences
Low	1	5	5/37	
Total	2 classes	37		

2.2 Classifier Performance Evaluation Metrics

Classification results of the classifiers were tabulated using the binary confusion matrix summarized in Table 2. The number correct claasifications including the true positive (TP), and the true negative (TN) as well as the misclassification results including the false positive (FP), and the false negative (FN) were utilized for estimating eleven performance evaluation metrics in equations 2 through 13.

Table 3: Binary Confusion Matrix

Predicted Crime	Actual Crime	
	High	**Low**
High	TP	FP
Low	FN	TN

$$\boldsymbol{N = TP + TN + FP + FN}$$

$$Correct\ Classfification\ Rate\ (CCR) = \frac{TP + TN}{N} \quad (2)$$

$$\text{Misclassification Error Rate (MER)} = 1 - CCR \quad (3)$$

$$\text{Sensitivity (SEN)} = \frac{TP}{TP{=}FN} \quad (4)$$

$$\text{Speificity (SPEC)} = \frac{TN}{TN+FP} \quad (5)$$

$$\text{Balance Accuracy (BA)} = \frac{SEN+SPEC}{2} \quad (6)$$

$$\text{FPR} = \frac{FP}{FP+TN} \quad (7)$$

$$\text{G-Mean} = \sqrt{SEN \times SPEC} \quad (8)$$

$$\text{Positive Predictive Value (PPV)} = \frac{TP}{TP+FP} \quad (9)$$

Negative Predictive Value (NPV) $= \frac{TN}{TN+FN}$ (9)

Except equation (3) for which lower value is desired, the higher the value of each of the metrics in equations (2) and (4) through (9), the better the classifier.

2.3 Method

The core idea of the SIR as an inverse based SDR method is that: for any vector $b \in \mathbb{R}^p, E(b^T\boldsymbol{X}|\mathbb{B}^T\boldsymbol{X})$ is a linear function of $\mathbb{B}^T\boldsymbol{X}$ (Hilafu, 2017 § 2.1 p. 3520). If this is true, Hilafu (2017) emphasized that the centred inverse moment $E(X|Y) - E(X)$ belongs to a subspace of R^p spanned by $\Sigma\mathbb{B}$, where Σ is MLE covariance of X. The implication of this is tha the covariance matrix $\mathbb{M} := Cov[E(X|Y)]$ is degenerate in any $\Sigma-$ orthogonal to Span($\mathbb{B}$). Consequently, the eigenvectors of the d nonzero eigenvalues of $\Sigma^{-1}\mathbb{M}$ span the subspace spanned the columns of the loading matrix $\mathbb{B}$ and serves a s the SIR estimates. This procedure is summarized in algorithm I.

Algorithm I: SIR

1. Let $\widehat{\boldsymbol{\Sigma}}, \overline{\boldsymbol{X}}_y$, and $\overline{\boldsymbol{X}}$ be the sample versions of $\boldsymbol{\Sigma}, \boldsymbol{E}(\boldsymbol{X}|\boldsymbol{Y})$, and $\boldsymbol{E}(\boldsymbol{X})$, respectively.
2. Estimate $\widehat{\mathbb{M}}$ for $\widehat{\boldsymbol{M}} = \sum_{h=1}^{H} \frac{n_h}{n} (\overline{\boldsymbol{X}}_h - \overline{\boldsymbol{X}}) (\overline{\boldsymbol{X}}_h - \overline{\boldsymbol{X}})^T$, where H and h denote the number of slices and the number observations in slice h, respectively.
3. The eigenvectors $\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbb{M}}$ corresponding to the $\boldsymbol{d}$ largest eigenvalues serve as the directions of SIR.

For a qualitative response like in this study, number slices, h, is set as the natural number of response categories. Li (1991) suggested ordering the values of quantitative response and slicing it into nonoverlapping ranges, and turning these into categories to estimate the conditional mean $E(\boldsymbol{X}|\boldsymbol{Y})$. To achieve SIR-HSMEC, HSMEC was used to replace usual MLE covariance stimator in steps of 2 and 3 of Algorithm I.

III. Applications

To investigate the performance of the SIR-HSMEC method, only the first sufficient predictor was utilized as the single predictor in the subsequent classification example to evaluate the effectiveness of five standard statistical classifiers in discriminating Nigerian cities according to their level crimes against property safety level. These classifiers including: the Classification Trees (CTree, Brieman et al., 1984; Ripley, 1996), The k-Nearest Neighbors (k-NN,Ripley, 1996, Venbles and Ripley, 2002) with only one optimal number of neighbour (i.e. 1-NN), the Linear Discriminant Analysis (LDA, James et al., 2013; Ripley, 1996, Venables abd Ripley, 2002), , The quadratic Discriminant Analysis (QDA, James et al., 2013; Ripley, 1996, Venables abd Ripley, 2002), the Random Forest (RF, Brieman, 2001; Breiman, 2002), the Neive Bayes classifier, and the Logistic Classifier (James et al., 2013) were trained and validated using 80:20 holdout scheme. All computations were done using the version 4.1.0 of the R software for statistical computing and graphics (R Core Team, 2021).

Table 4: SIR-HSMEC Based Classifier Performance Evaluation (%)

Metrics	LR	RF	QDA	LDA	NB	1-NN	CTree
Accuracy	0.71	0.86	0.57	0.71	0.57	1.00	0.86
Sensitivity	0.80	0.80	0.40	0.60	0.40	1.00	0.80
Specificity	0.50	1.00	1.00	1.00	1.00	1.00	1.00
BA	0.65	0.90	0.70	0.80	0.70	1.00	0.90
G-Mean	0.80	0.89	0.63	0.77	0.63	1.00	0.89
FPR	0.50	0.00	0.00	0.00	0.00	0.00	0.00
PPV	0.80	1.00	1.00	1.00	1.00	1.00	1.00
NPV	0.50	0.67	0.40	0.50	0.40	1.00	0.67
AUC	0.65	0.83	0.70	0.75	0.70	1.00	0.83

Based on results presented in Table 4, the SSIR-HSMEC estimator is successful with all classifiers with false positive rates of 0% except the Logistic classifer which had 50% false positive rate. Except the QDA and the NB classifiers which achieved sensitivity of 40% each, all classifiers achieved at least 60% sensitifity. This implies that these classifiers are correctly detect cities with high safety level at least 60% of the time. Except the logistic classifier which can only detect cities with low safety levels 50% of the times, all other classifiers can detect cities with low level of crimes against property safety 100% of the time. In terms of balance of classification ability between sensitivity and specificity, all the classifiers achieved good balance accuracy and geometric mean. The least balance accuracy (BA) and geometric means achieved by the classifiers are 65% (Logistic BA) and 63% (QDA G-Mean), respectively.

All the classifiers achieved 100% positive predictive value except the logistic classifier which achieved PPV of 80%. This implies that If these classifiers detect high safety level, a low rate of crimes against property is detected at least 80% of the time. The least negative predictive value achieved by the classifiers is 40% (QDA and NB). With least area under curve value of approximately, 65%, it is evident that all the classifiers are better than a random classifier.

Overall, the k-Nearest Neighbour with 1 optimal neighbour is the best classifier for the classification task in this study. It consitently ranked best among all the classifiers evaluated across all the performance metrics.

IV. Conclusion

The proposed SIR-HSMEC fully addresses loss of covariance information in classification problems. The developed is recommended whenever analysis goal is to extract core relevant information hidden in the city crime safety discrimination problem. This work has demonstrated that the proposed MEC estimator efficiently deals with singularity and instability of sample covariance estimate in SDR applications without requirement for time-consuming covariance estimation procedures. The proposed MEC estimator can also be used in applications other than SDR to circumvent covariance instability. The eigenvalues of the HSMEC are well-conditioned and are positive definite thereby providing a positive and invertible plug-in covariance matrix needed in the SDR step. It is, therefore, evident that consistent classifiers with just one SIR predictor can be developed. With this single predictor, many classifiers whose performance would have been otherwise disrupted by noise in the data can now achieve almost perfect classification results.

References

Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. sJ. (1984), Classification and regression trees, Wadsworth International Group, Belmont, California.

Breiman, L. (2001), "Random Forests", The Journal of Machine Learning, 45(1), 5-32.

Breiman, L. (2002), Manual On Setting Up, Using, and Understanding Random Forests V3.1, edu/~breiman/Using_random_forests_V3.1.pdf, https://www.stat.berkeley.

Cook, R., & Li, B. (2002). Dimension reduction for conditional mean in regression. Ann. Statist., 30, 455-474.

Cook, R. D., & Ni, L. (2005). Sufficient dimension reduction via inverse regression: A minimum discrepancy approach. Journal of the American Statistical Association, 100 470, 410-428. doi: 10.2307/27590565

Cook, R. D. (1994b). Using dimension-reduction subspaces to identify important inputs in models of physical systems. In 1994 proceedings of the section on physical and engineering sciences, alexandria, va: American statistical association (p. 18-25).

Cook, R. (1996). Graphics for regression with a binary response. Journal of the American Statistical Association, 91, 983-992.

Cook, R. D. (2000). Save: a method for dimension reduction and graphics in regression. Communications in Statistics - Theory and Methods, 29(9-10), 2109-2121. Retrieved from https://doi.org/10.1080/03610920008832598 doi: 10.1080/03610920008832598

Cook, R. D., & Weisberg, S. (1991). Sliced inverse regression for dimension reduction: Comment. Journal of the American Statistical Association, 86(414), 328-332.

Cook, R. D., & Forzani, L. (2008b). Principal Fitted components for dimension reduction in regression. Statistical Science, 23(4), 485-501.

Cook, R. D., & Forzani, L. (2009). Likelihood-based Sufficient dimension reduction. Journal of American Statistical Association, 104(485), 197-208.

Cook, R. D. (2007). Fisher lecture: Dimension reduction in regression (with discussion). Statistical Science, 22, 1-26. doi: 10.1214/088342306000000682

Cook, R. D., Li, B., & Chiaromonte, F. (2007). Dimension reduction in egression without matrix inversion. Biometrika, 94(3), 569-584.

Cook, R. D., Li, B., & Chiaromonte, F. (2010). Envelopes models for parsimonious and efficient multivariate linear regression. Statistica Sinica, 20(3), 927-960.

Cook, R. (1998). Regression graphics: Ideas for studying regressions through graphics. John Wiley & Sons.

Dawid, A. P. (1979). Conditional independence in statistical theory. Journal of the Royal Statistical Society. Series B (Methodological), 41(1), 1-31. Retrieved from http://www.jstor.org/stable/2984718

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R, 1st edition, New York, Springer Texts in Statistics, DOI:10.1007/978-1-4614-7138-7_1.

Li, K.-C. (1991). Sliced inverse regression for dimension Reduction (with discussion). Journal of the American Statistical Association, 86, 316-342.

Li, K.-C. (1992). On principal hessian directions for data Visualization and dimension reduction: Another application of stein's lemma. Journal of The American Statistical Association – J AMER STATIST ASSN, 87, 1025-1039. doi:

10.1080/01621459.1992.10476258

Ma, Y., & Zhu, L. (2013). A review on dimension reduction.
Int. Statist. Rev., 81(1), 134-150

R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/, 2021.

Olorede, K. O., & Yahya, W. B. (2019). A new covariance estimator for sufficient dimension reduction in high-dimensional and undersized sample problems (Vol. arXiv:1909.13017v1 [stat.ME]). preprint available at https://arxiv.org/pdf/1909.13017.pdf.

Ripley, B. D. (1996), Pattern Recognition and Neural Networks, Cambridge.

Venables, W. N., and Ripley, B. D. (2002), Modern Applied Statistics with S. 4th edition, Springer.

Xia, Y., Li, W. K., & Zhu, L.-X. (2002). An adaptive estimation of dimension reduction space. Journal of the Royal Statistical Society B, 64, 363-410.