Proceedings of 11th International Conference on Computer and Information Technology (ICCIT 2008)
25-27 December, 2008, Khulna, Bangladesh

Sensitivity Analysis in Robust and Kernel Canonical Correlation Analysis

Md. Ashad Alam †, Mohammed Nasser‡ and Kenji Fukumizu♯

†Hajee Mohammad Danesh Science and Technology University, Dinajpur-5200, Bangladesh
‡Rajshahi University,Rajshshi-6205, Bangladesh
♯Institute of Statistical Mathematics,Tokyo 106-8569, Japan
Emails: kakon_bud@yahoo.com; mnasser.ru@gmail.com; fukumizu@ism.ac.jp

Abstract—**A number of measures of canonical correlation coefficient are now used in pattern recognition in the different literature. Some robust forms of classical canonical correlation coefficient are introduced recently to address the robustness issue of the canonical coefficient in the presence of outliers and departure from normality. Also a few number of kernels are used in canonical analysis to capture nonlinear relationship in data space, which is linear in some higher dimensional feature space. But not much work has been done to investigate their relative performances through simulation and also from the view point of sensitivity. In this paper an attempt has been made to compare performances of kernel canonical correlation coefficients (Gaussian, Laplacian and Polynomial) with that of classical and robust canonical correlation coefficient measures using simulation and influence function. We investigate the bias, standard error, MSE, qualitative robustness index, sensitivity curve of each estimator under a variety of situations and also employ boxplots and scatter plots of canonical variates to judge their performances. We observe that the class of kernel estimators perform better than the class of classical and robust estimators in general and the kernel estimator with Laplacian function has shown the best performance for large sample size.**

Index Terms—**Pattern Recognition, Robust CCA, KCCA, Monte Carlo Simulation, Influence Function and Sensitivity Curves.**

I. Introduction

The main goal of pattern recognition is supervised or unsupervised classification. It is the study of how machines can observe the environment, learn to distinguish patterns of interest from their background, and make a sound and reasonable decisions about the categories of the patterns. The statistical approach has been most intensively studied to meet up this goal properly [1]. In this approach each pattern is represented in terms a number (such as s) of features or measurements and is viewed as a point in a s-dimensional space. The aim is to choose those features that allow pattern vectors belonging to different categories to occupy compact and disjoint regions in a s-dimensional feature space. The effectiveness of the representation space (feature set) is determined by how well patterns from different classes can be separated. To do this, the measure of the relation between distinct features is essential. Classical CCA is a multivariate procedure for assessing the linear relationship between two sets of variables(or features). Canonical correlation was developed by Hotelling [2]. His basic elegant paper on the topic, "Relations between two sets of variables" appeared in 1936. In this paper, Hotelling described an example where one set of variables consists of mental tests and the other of physical measurements on a group of people. Then it acquired popularity and many statisticians, biometricians, economists, social scientists, and many other researchers applied it in diverse field of knowledge. During the last one decade the use of CCA is utmost popular method in the area of computer science such as generic object recognition, image analysis,image retrieval, machine translation, image segmentation, computer vision, pattern recognition, and facial expression recognition. Now-a-days KCCA are used in different fields such as computer science, information technology and so on. There are some limitations of KCCA as complexity of transformation, dimensionality (dimension is larger than sample size) and time complexity to calculate. Most of the KCCA researchers, namely [3], [4], [5] etc. tried to overcome limitations of KCCA. In a recent work [6], the influence function of kernel PCA and a robust KPCA has been theoretically derived. One observation of their analysis is that KPCA with a bounded kernel such as Gaussian is robust in that sense that the influence function does not diverge, while for KPCA with unbounded kernels such as polynomial the IF goes to infinity. This can be understood by the boundedness of the transformed data in the feature space by a bounded kernel. While this is not a result for CCA but for PCA, it is reasonable to expect that KCCA with a bounded kernel is also robust. This consideration motivates us to do some empirical studies on the robustness of KCCA. It is very important to know how KCCA is effected by outliers and to develop measures of accuracy. Therefore, we do intend to study a number of conventional robust estimates and KCCA with different functions.

In this article we consider five canonical correlation coefficients and investigate their performances by simulation, influence function and sensitivity curve. We

1-4244-2136-7/08/$20.00 ©2008 IEEE

briefly discuss the classical, robust and kernel canonical correlation coefficients that we consider in our study in section II and III. Section IV offers the simulation as well as influence function results of the canonical correlation coefficients and finds the overall best performer. The final section gives conclusion.

II. Classical and Robust Canonical Correlation Analysis

Classical CCA can be seen as the problem of finding basis vectors for two sets of variables such that the correlation between the projections of the variables onto these basis vectors are mutually maximized. The classical covariance and correlation matrices as well as eigenvector and values are highly sensitive to outlying observations as was shown in the context of CCCA. An obvious approach to robustify CCA is to estimate the population, sample covariance or correlation matrix. A number of approaches for robust canonical correlation analysis except the one just mentioned were compared and discussed by Branco[7]and Taskinen [8] obtained influence function and asymptotic distributional properties of CCCA based on robust estimates of the covariance matrix. Many researchers [7], [8] and [9] developed a few number of robust methods in CCA and suggested that from the view point of reobustness and computaton the performance of minimum covariance determinate (MCD) estimator is the best [7]. In our paper, we consider MCD estimator as an estimator of class of robust method.

III. Kernel Canonical Correlation Analysis (KCCA)

At the beginning of this century there have been advances in pattern recognition of data using kernel methods [10], [11]. It is a method that generalizes the classical CCA to nonlinear setting in a nonparametric way. The kernel-based canonical correlation analysis no longer requires the Gaussian distribution assumption on observation and, therefore, enhances the applicability of CCA greatly. Basics idea of KCCA is as follows: first the original pair of multivariate variables x and y are mapped into ϕ_x(x) and ϕ_y(y), elements of some high (possibly infinite) dimensional Hilbert spaces $\mathcal{H}_x$ and $\mathcal{H}_y$ respectively through a suitable kernel function without calculating $\phi_.(.)$'s in order to get benefit of the simple fact-what is non-linear in a vector space may be linear in a transformed high dimensional space. Resorting to kernel trick through forming kernel matrix linear transformation similar to classical case is implemented in the transformed high dimensional space. Using about similar calculation that, in fact, can find more complex transformation than CCA in original space (summarized in Fig. 1).

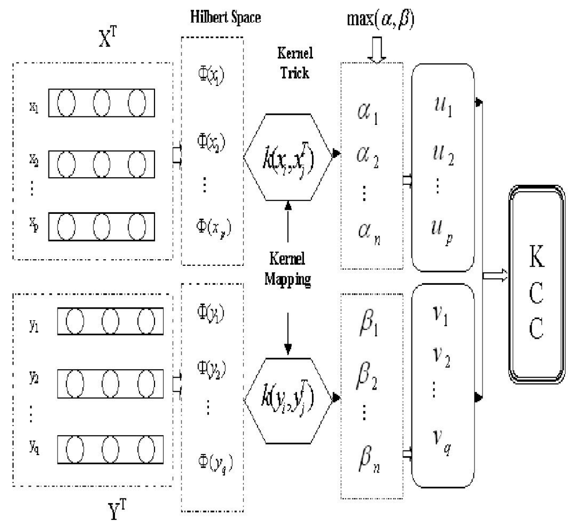

Fig. 1. System of Kernel CCA.

A. Derivation of the Kernel Canonical Correlation

Let us consider linear dependence of two multivariate data set (x_i, y_i) be the p dimensional and q dimensional x and y sets of variables having $i = 1, 2, \ldots$ n observations respectively. The classical canonical correlation is find the directions a and b so that the correlation between the projections of X onto a and that of Y onto b is maximized such that

$$\begin{aligned} \rho &= \underset{a \in \Re^p, b \in \Re^q}{max} \frac{Cov[a^T X, b^T Y}{\sqrt{Var[a^T X] Var[b^T Y]}} \\ &= \underset{a \in R^p, b \in R^q}{max} \frac{a^T \hat{V_{xy}} b^T}{\sqrt{a^T \hat{V_{xx}} a}\sqrt{b^T \hat{V_{yy}} b}} \end{aligned} \quad (1)$$

where $\hat{V_{xx}}, \hat{V_{yy}}$ and $\hat{V_{xy}}$ are the sample variance and covariance matrices respectively. Classical canonical correlation is highly depend on linearity assumption, because of this assumption it gives us naive measurement in case of nonlinear data set. Even robust method can fail to find worthy relationship in such type of data. We have to seek such a method that gives us accurate measurement in case of nonlinear data. Kernel canonical correlation is a such type of powerful tool that offers us an alternative solution for nonlinear data by first projecting the data into a higher dimensional feature space (Hilbert space). Let $\{(x_i, y_i)$; $i = 1, 2, \ldots n\}$ be the training sample with x_i and y_i taking values in $\mathcal{X}$ and $\mathcal{Y}$ respectively. Let us prepare kernels k_x on $\mathcal{X}$ and k_y on $\mathcal{Y}$ as $x_1 \ldots x_n \to \phi_x(x_1), \ldots \phi_x(x_n) \in \mathcal{H}_{k_x}$ and $y_1 \ldots y_n \to \phi_y(y_1), \ldots \phi_y(y_n) \in \mathcal{H}_{k_y}$. Now we apply classical CCA on Hilbert space $(\mathcal{H}_x, \mathcal{H}_y)$

$$f \in \mathcal{H}_x^{max}, g \in \mathcal{H}_y \frac{\sum_{i=1}^{n} \langle f, \tilde{\phi}_x(x_i)\rangle_{\mathcal{H}_x} \langle g, \tilde{\phi}_y(y_i)\rangle_{\mathcal{H}_y}}{\sqrt{\sum_{i=1}^{n} \langle f, \tilde{\phi}_x(x_i)\rangle^2_{\mathcal{H}_x}} \sqrt{\sum_{i=1}^{n} \langle g, \tilde{\phi}_y(y_i)\rangle^2_{\mathcal{H}_y}}} \quad (2)$$

where $\tilde{\phi}_x(x_i) = \phi_x(x_i) - \frac{1}{n}\sum_{i=1}^{n} \phi_x(x_i)$ and $\tilde{\phi}_y(y_i) = \phi_y(y_i) - \frac{1}{n}\sum_{i=1}^{n} \phi_y(y_i)$ Without any lose of generality, we can assume that $f = \sum_{i=1}^{n} \alpha_i \tilde{\phi}_x(x_i)$ and $g = \sum_{i=1}^{n} \beta_i \tilde{\phi}_y(y_i)$

$$\rho = \alpha \in \Re^{n}, \beta \in \Re^{n} \frac{\alpha^T \tilde{K}_x \tilde{K}_y \beta}{\sqrt{\alpha^T \tilde{K}_x^2 \alpha}\sqrt{\beta^T \tilde{K}_y^2 \beta}} \quad (3)$$

where $\tilde{K}_x^2$ and $\tilde{K}_y^2$ are the centered Gram matrices. It is equivalent to performing CCA on two vectors of dimension n with covariance matrix equal to $\begin{bmatrix} K_x^2 & K_x K_y \\ K_y K_x & K_y^2 \end{bmatrix}$. Thus we observe that we can perform a kernelized version of CCA by solving the following generalized eigenvalue problem.

$$\begin{bmatrix} 0 & K_x K_y \\ K_y K_x & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \rho \begin{bmatrix} K_x^2 & 0 \\ o & K_y^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad (4)$$

based on the Gram matrices.

B. Different Kernel Functions for KCCA

The kernel function is used in training and predicting. The parameter of this function can be set to any function of class kernel that computes the inner product in feature space between two vector arguments. In this work we consider, Gaussian RBF kernel, $k(x_i, x_j) = exp(-\frac{||x_i - x_j||^2}{\sigma})$, Laplacian kernel function, $k(x_i, x_j) = exp(-\frac{||x_i - x_j||)}{\sigma})$ and Polynomial function, $k(x_i, x_j) = (\langle x_i, x_j\rangle + c)^d$.

IV. RESULTS

In this section we consider the simulation, influence function, and estimate of influence function results respectively. We generate multivariate normal (MVN) data by considering two covariance matrices CVM_1 and CVM_2 that are given below. We also generate data from uniform distribution having $[-\pi, \pi]$. After that we transform generated data by sin and cos function in different way. We find bias, standard error, mean square error, qualitative robustness index(QRI)[12] of the canonical correlation estimators at different models as well as transformed data. The results are represented in different tables and boxplots.

$$CVM_1 = \begin{bmatrix} 1.0 & 0.8702 & -0.3657 & -0.3896 & -0.4931 & -0.2263 \\ 0.8703 & 1.0 & -0.3529 & -0.5522 & -0.6456 & -0.1915 \\ -0.3657 & -0.3529 & 1.0 & 0.1506 & 0.2250 & 0.0349 \\ -0.3896 & -0.5522 & 0.1506 & 1.0 & 0.6957 & 0.4957 \\ -0.4938 & -0.6456 & 0.22503 & 0.695 & 1.0 & 0.6692 \\ -0.2263 & -0.1915 & 0.03493 & 0.4957 & 0.669 & 1.0 \end{bmatrix}$$

$$\text{and } CVM_2 = \begin{bmatrix} 1.000 & 0.505 & 0.569 & 0.602 \\ 0.505 & 1.000 & 0.422 & 0.467 \\ 0.569 & 0.422 & 1.000 & 0.926 \\ 0.602 & 0.467 & 0.926 & 1.000 \end{bmatrix}$$

A. Simulation Results

Here we report a simulation study of different correlation coefficients. In our study we consider three different sample sizes $(50, 500,$ and $1000)$ with different types of Multivariate models. To perform simulation study we consider three experiments as follows.

Experiment-1 (Data generated from multivariate normal(MVN) and MVN with 6%contamination): In this experiment we generate data from MVN as ideal model and MVN with 6% fixed outlier as contaminated model. We consider three sample size $n = 50$, $n = 500$ and $n = 1000$ and simulated 1000 times. In case of $n = 1000$ we take fewer replications in order to avoid much calculation. After that we calculate bias, standard error, MSE as well as QRI which are given in the Table I and Table II respectively. From Table I, we see that, although in case of bias (Minimum values as bold) the performance of classical and robust are better but over all, we may conclude that kernel estimators are better than classical as well as robust estimators specially at contaminated models and large sample size.

Qualitative Robustness Index (QRI): To measure the effect of contamination on different estimators at different contamination models we use qualitative robustness index;

$$QRI = \frac{1}{\sum |\xi^{(s)}_{100\alpha} - \xi^{(s)c}_{100\alpha}|} \quad (5)$$

where $\xi^{(s)}_{100\alpha} = 100^{th}_{\alpha}$ percentile of the simulated sampling distribution of the estimators at standard model, and $\xi^{(s)c}_{100\alpha} = 100^{th}_{\alpha}$ percentile of the simulated sampling distribution of different estimators at contaminated model. We consider $\alpha = 0.005, 0.01, 0.025, 0.05, 0.1, 0.5, 0.9, 0.95, 0.975, 0.99, 0.995$ for our standard model and contaminated model. By considering qualitative robustness index (Maximum values as bold)in Table II, we see that both robust and kernel method perform better than classical method.

Experiment-2 (Data generated from multivariate normal and transformation with $CC = 0.79$ and $CC = 0.63$): In this experiment we generate data by using CVM_1 and CVM_2 and take transformation in the following way. In case of CVM_1, the data of x-set with three variables x_1, x_2 and x_3 transform by $sin(2x_1)$, $cos(x_2)$ and $sin(3x_3)$ as well as y-set as $cos(2y_1)$, $sin(y_2)$, $cos(3y_3)$. For CVM_2 we take transformation as $sin(x_1)$, $sin(3x_2)$ and $cos(2y_1)$, $cos(3y_2)$ for both x-set and y-set data respectively. By using canonical coefficient we sketch the boxpolts of five estimators, which are given in Fig. 2a, and Fig. 2b. having first population $CC = 0.79$ and $CC = 0.63$) respectively. From the Fig.2, we observe that in both cases classical and robust method give very naive estimates over all sample sizes. But the perfomace of kernel method is the best. For

TABLE I
Bias, standard error and mean square error vlues of simulated data

Estimators	Model	Bias			SE			MSE		
		$n = 50$	$n = 500$	$n = 1000$	$n = 50$	$n = 500$	$n = 1000$	$n = 50$	$n = 5000$	$n = 1000$
	MVN	**0.0183**	**0.0171**	**0.0011**	0.0486	0.0503	0.0123	0.0489	0.0506	0.0123
CC	MVNC	0.6658	0.6713	0.6628	0.0311	0.0098	0.0069	0.4744	0.4604	0.4462
	MVN	0.0554	0.0460	0.0016	0.0657	0.0751	0.0136	0.0688	0.0772	0.0136
RC	MVNC	**0.0473**	**0.0022**	**0.0012**	0.0717	0.0193	0.0132	0.0739	**0.0193**	0.0132
	MVN	0.1887	0.1887	0.10137	0.0134	0.0139	0.0144	0.0495	0.0495	0.0146
KG	MVNC	0.1909	0.0343	0.0160	0.0117	0.0246	**0.0120**	0.0481	0.0258	**0.0123**
	MVN	0.1044	0.2044	0.0591	**0**	**0**	0.0212	**0.4180**	**0.04180**	0.0247
KL	MVNC	0.2044	0.1764	0.0659	**0**	0.0215	0.0206	**0.0418**	0.0526	0.0249
	MVN	0.0924	0.0938	0.0047	0.0372	0.0385	**0.0121**	0.0457	0.0473	**0.0121**
KP	MVNC	0.1809	0.0286	0.0454	0.0153	**0.0019**	0.0136	0.0480	0.0198	0.0157

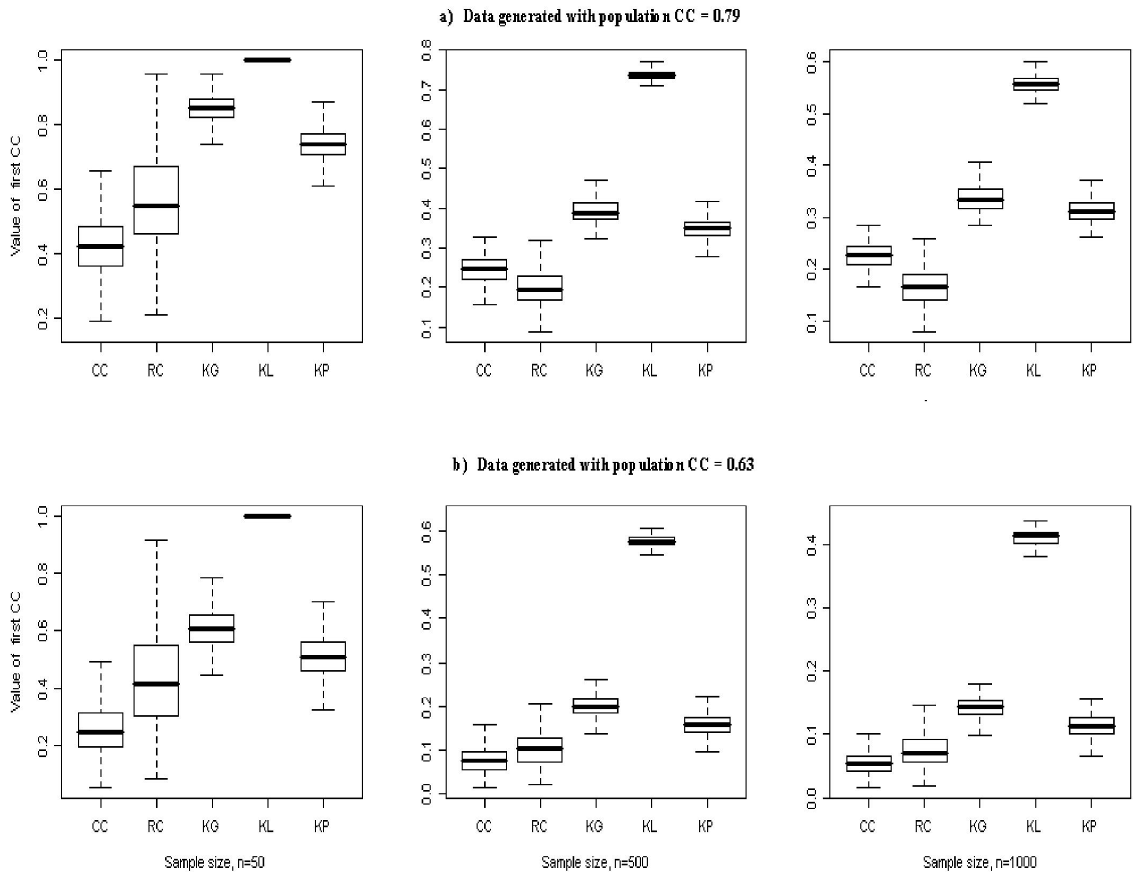

Fig. 2. Boxplots of canonical correlation coefficient of five estimators with population $CC = 0.79$ and $CC = 0.63$.

sample $n = 50$ the kernel estimator with Gaussian (KG) function gives smallest bias, on the other hand, for large sample Laplacian function (KL) gives smallest bias along with minimum variation in the both cases.

Experiment-3 (Generated from uniform distribution): In this section we generate data according to [4], [5] but size is 100. At first θ is generated from the uniformdistribution on $[-\pi, \pi]$ after that a pair of two dimensional variables x and y are generated by $x = [\theta, sin3\theta]$ and $y = e^{\theta/4}[cos2\theta, sin2\theta]$. Form Fig. 3, we see that classical method does not give linear relationship, robust method gives higher coefficient but relation for first canonical variates are nonlinear.

TABLE II
Qualitative Robustness Index

Estimators	$n = 50$	$n = 500$	$n = 1000$
CC	0.0015	0.0014	0.0015
RC	**3.4483**	0.0245	**0.0025**
KG	0.0833	0.0066	**0.0025**
KL	0.1	**0.0412**	0.0024
KP	0.0102	0.0083	0.0024

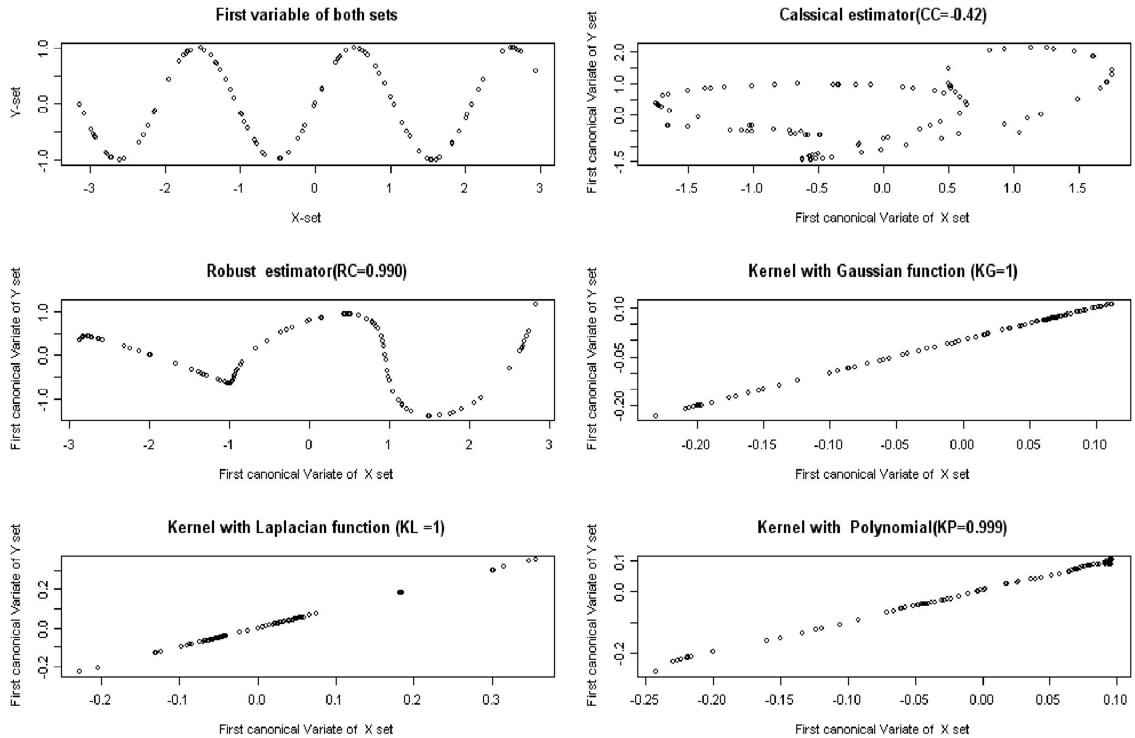

Fig. 3. Scatter plot of first canonical variate of data generated form uniform distribution with size $n = 100$.

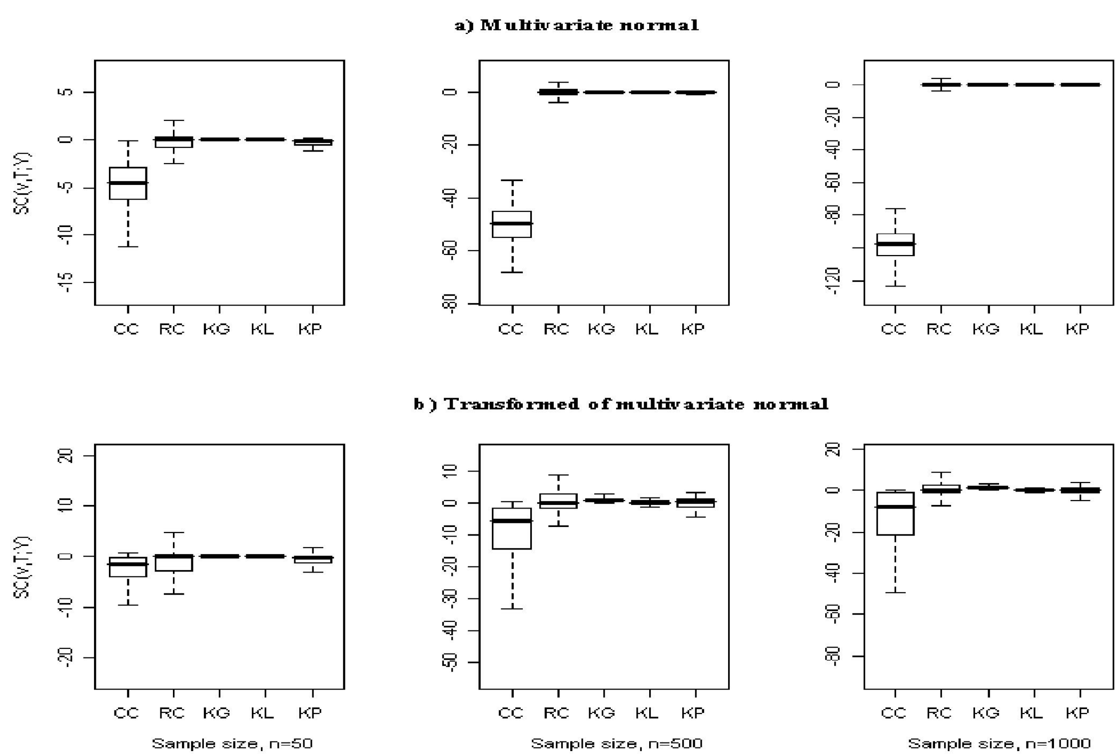

Fig. 4. Box plot of 200 numbers generated from $SCn(v; T_n : V)$where follows a) Multivariate normal and b) Transformation of Multivariate normal having $n = 50, 500$ and 1000.

On the other hand, kernel method shows the best performance providing linear relationship of canonical variates. Especially kernels with Gaussian (KG) and Laplacian function are found as the best ones in offering linear relationship. We have also observed that similar scatter plot for second canonical variate.

B. Sensitivity Analysis

In this section following [13], we perform a simulation study to confirm some aspects of our findings with the help of sensitivity curves. Firstly we generate 200 samples of size $n = 50$, $n = 500$ and $n = 1000$ from a MVN using CVM_1 and compute $SC_n(v, T_n; V)$ for each of these samples. We also consider an outlier values $v = 500, 1000, 1500$ and $0.5, 0.1, 0.15$ for x-set and y-set respectively. That is, the sample sizes are $n = 51$,$n = 501$ and $n = 1001$. Boxplots of these 200 numbers are shown in Fig. 4a, for MVN using CVM_1 and we repeat the work by taking a transformation as $sin(2x_1)$, $cos(x_2)$, $sin(3x_3)$ and $cos(2y_1)$, $sing(y_2)$, $cos(3y_3)$ for x-set and y-set respectively; boxplots of the values target are given in Fig. 4b, of classical estimator (CC), robust estimator (RC) kernel with Gaussian function (KG), kernel with Laplacian function (KL) and kernel with polynomial (KP).

V. Conclusion

In this paper we compare the performances of five (classical, robust and kernel with three functions) estimators of canonical correlation coefficient that are commonly used in the statistical literature as well as in pattern recognition. Their performances are investigated through simulation and influence function. We report a simulation where sampling distributions are taken as multivariate normal with two covariance matrices, CVM_1 and CVM_2 and its contaminated forms, and we also consider uniform distribution with three different sample sizes $(50, 500$ and $1000)$ and in different way of transformation. The estimators are compared with regard to bias, standard error, MSE and QRI. Over all, the class of kernel estimators, especially the Gaussian function for small sample and Laplacian function for large sample perform better than the class of classical as well as robust estimators. A study based on sensitivity curves shows that outliers have the least effect on kernel estimator. We recommend using as a measure of kernel canonical correlation coefficient especially when the sample is large and contamination is likely.

Acknowledgment

The authors wish to acknowledge the various helpful comments, assistance and support of the reviewers and Publication Committee.

References

[1] A. K. Jain, R. P. W. and J. Mao. Statistical Pattern Recognition: A Review, *IEEE Transactions on Pattern analysis and Machine Intelligence*, vol. 3, no. 1, pp. 4-37, January, 2000.

[2] H. Hotelling. Relations between two sets of variables, *Biometrika* , vol. 28, pp. 321-377, 1936.

[3] D. R. Hardoon, S. Szedmak, J. Shawe-Taylor. Canonical correlation analysis: an overview with application to learning methods, *Technical Report, Royal Holloway University of London*, CSD-TR-03-02, 2003.

[4] Akaho, S. (2001). A kernel method for canonical correlation analysis. In Proceeding of the 2000 Workshop on Information Based Induction Science (IBS2000), vol.17-18, pp.123-128, 2000.

[5] S. Akaho. A kernel method for canonical correlation analysis, *International meeting of psychometric Society*, 2007.

[6] S.Y. Huang, Y.R. Yeh,S. Eguchi. Robust Kernel Principal Component Analysis, Research Memorandum 1057, Institute of Statistical Mathematics, Japan 2008.

[7] A. J. Branco, C. Croux,P. Filzmoser, M.R. and Oliviera. Robust Canonical Correlations: A Comparative Study, *Computational Statistics*, vol. 20, pp. 203-229, 2005.

[8] S. Taskinen, C. Croux, A. Kankainen,E. Ollila, and H. Oja. Canonical Analysis based on Scatter Matrices, *Journal of Multivariate Analysis*, vol. 97, no.2, pp. 359-384, 2006.

[9] D. Skocaj, A. Leonardis, and S. Fidler. Robust Estimation of Canonical Correlation Coefficients, *Proceedings of OAGM04*, pp.15-22, 2004.

[10] F. R. Bach and M. I. Jordan. Kernel independent component analysis, *J.March. Learning Rss*, vol. 3, pp.1-48, 2002.

[11] K. Fukumizu, F. R. Bach, and A. Getton. Consistency of kernel canonical correlation, Research Memorandum 942, Institute of Statistical Mathematics, Japan 2005.

[12] M. Huda. *Bootstrapping and Robust Estimation in Factor Analysis: Application in Stock Price Data*, M.Sc. thesis, Department of statistics, University of Rajshahi, Bangladesh, 2005.

[13] M. A, Alam, M. Nasser, and A.H.M. R. Imon. Sensitivity and Influence Analysis of Estimators of Correlation Coefficients, *Journal of Applied Statistics and Probability*, vol. 3, pp. 119-136, May, 2008.