International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 16, 2018, pp 97-114

Mohammed Nasser: Papers and Publications

[Abstracts only]

1. Support Vector Machine and Random Forest Modeling for Intrusion Detection System (IDS)

-Md. Al Mehedi Hasan, Mohammed Nasser, Biprodip Pal and Shamim Ahmad

[Journal of Intelligent Learning Systems and Applications (Feb 2014): 45-52.]

Abstract. The success of any Intrusion Detection System (IDS) is a complicated problem due to its nonlinearity and the quantitative or qualitative network traffic data stream with many features. To get rid of this problem, several types of intrusion detection methods have been proposed and shown different levels of accuracy. This is why the choice of the effective and robust method for IDS is very important topic in information security. In this article, the authors have built two models for the classification purpose. One is based on Support Vector Machines (SVM) and the other is Random Forests (RF). Experimental results show that, either classifier is effective. SVM is slightly more accurate, but more expensive in terms of time. RF produces similar accuracy in a much faster manner if given modeling parameters. These classifiers can contribute to an IDS system as one source of analysis and increase its accuracy. In this article, KDD'99 Dataset is used and find out which one is the best intrusion detector for this dataset.

2. Comparison of the Finite Mixture of ARMA-GARCH, Back Propagation Neural Networks and Support-Vector Machines in Forecasting Financial Returns

- Altaf Hossain and Mohammed Nasser

[Journal of Applied Statistics, Volume 38, 2011 - Issue 3 , Pages 533-551
https://doi.org/10.1080/02664760903521435]

Abstract. The use of GARCH type models and computational-intelligence-based techniques for forecasting financial time series has been proved extremely successful in recent times. In this article, we apply the finite mixture of ARMA-GARCH model instead of AR or ARMA models to compare with the standard BP

and SVM in forecasting financial time series (daily stock market index returns and exchange rate returns). We do not apply the pure GARCH model as the finite mixture of the ARMA-GARCH model outperforms the pure GARCH model. These models are evaluated on five performance metrics or criteria. Our experiment shows that the SVM model outperforms both the finite mixture of ARMA-GARCH and BP models in deviation performance criteria. In direction performance criteria, the finite mixture of ARMA-GARCH model performs better. The memory property of these forecasting techniques is also examined using the behavior of forecasted values vis-à-vis the original values. Only the SVM model shows long memory property in forecasting financial returns.

3. A Diagnostic Measure for Influential Observations in Linear Regression

- A. A. M. Nurunnabi, A. H. M. Rahmatullah Imon and M. Nasser

[Communications in Statistics - Theory and Methods, Volume 40, 2011 - Issue 7 Pages 1169-1183, https://doi.org/10.1080/03610920903564727]

Abstract. In linear regression it is a common practice of measuring influence of an observation is to delete the case from the analysis and to investigate the change in the parameters or in the vector of forecasts resulting from this deletion. Pena (2005) introduced a new idea to measure the influence of an observation based on how this observation is being influenced by the rest of the data. In this article we propose a new influence measure extending the idea of Pena to group deletion for identifying multiple influential observations in linear regression. We investigate the usefulness of the proposed technique by two well-referred data sets, an artificial large data with high-dimension and heterogeneous sample points and by reporting a Monte Carlo simulation experiment.

4. Identification of Multiple Influential Observations in Logistic Regression

-A.A.M. Nurunnabi, A. H.M. Rahmatullah Imon and M. Nasser

[Journal of Applied Statistics, Volume 37, 2010 - Issue 10, Pages 1605-1624, https://doi.org/10.1080/02664760903104307]

Abstract. The identification of influential observations in logistic regression has drawn a great deal of attention in recent years. Most of the available techniques

like Cook's distance and difference of fits (DFFITS) are based on single-case deletion. But there is evidence that these techniques suffer from masking and swamping problems and consequently fail to detect multiple influential observations. In this paper, we have developed a new measure for the identification of multiple influential observations in logistic regression based on a generalized version of DFFITS. The advantage of the proposed method is then investigated through several well-referred data sets and a simulation study.

5. Regression Diagnostics for Multiple Model Step Data

- A.A.M. Nurunnabi, Sch. of Bus., Uttara Univ., Dhaka, Bangladesh and Mohammed Nasser

[International Conference on Digital Image Processing, 2009, ISBN: 978-0-7695- 3565-4, DOI: 10.1109/ICDIP.2009.71 , Bangkok, Thailand]

Abstract. In many vision and image problems there are multiple structures in a single data set and we need to identify the multiple models. To preserve most structures in presence of noise makes the estimation difficult. In such case for each structure, data which belong to other structures are also outliers in addition to the outliers for all the structures. Robust regression techniques are commonly used to serve the model building process for noisy data to the vision community, that fits the majority data and then to discover outliers, they tend to fail to cope with the situation. In this paper we show a newly proposed regression diagnostic measure is capable for identifying large fraction of outliers, and regression diagnostics may be a better choice to the robust regression. We demonstrate the whole thing through several artificial multiple model step data.

6. Outlier Diagnostics in Logistic Regression: A Supervised Learning Technique

- A.A.M. Nurunnabi and Mohammed Nasser

Abstract. The goal of supervised learning is to build a concise model of the distribution of class labels in terms of predictor features. Logistic regression is one of the most popular supervised learning technique that is used in classification. Fields like computer vision, image analysis and engineering sciences frequently encounter data with outliers (noise). Presence of outliers in the training sample may be the

cause of large training time, misclassification, and to design a faulty classifier. This article provides a new method for identifying outliers in logistic regression. The significance of the measure is shown by well-referred data sets.

7. Outlier Detection by Regression Diagnostics in Large Data

- A. A. M. Nurunnabi; Sch. of Bus., Uttara Univ., Dhaka, Bangladesh and Mohammed Nasser

[International Conference on Future Computer and Communication, 2009, ISBN: 978-0-7695-3591-3, DOI: 10.1109/ICFCC.2009.60, Kuala Lumpar, Malaysia]

Abstract. Regression analysis is a well known supervised learning technique. To estimate and justify an effective model from regression analysis it is necessary to check and preprocess the data set. Without outliers (noise) it is impossible to get a real data. Areas in bio-informatics, astronomy, image analysis, computer vision etc, large or fat data appear with unusual observations (outliers) very naturally. In these industries robust regression are commonly used in model building process. But robust regression methods are not good enough in large and/or high dimensional data. Checking raw data for outliers in regression is regression diagnostics. Robust regression and regression diagnostics are two complementary ideas and any one is not enough for studying a contaminated data. Most of the popular diagnostic methods are not sufficient for large data because of masking and swamping. In this article, both of the above ideas are shortly discussed and we show a new measure can effectively identify outliers (influential observations) in linear regression for large data.

8. A New Influence Statistic in Linear Regression

- A.A.M. Nurunnabi, A.H.M.R. Imon and M. Nasser

[Conference: Proceedings of International Conference on Statistical Sciences (ICSS'2008), pp. 165-173., Dhaka, Bangladesh]

Abstract. Since Cook's (1977) seminal article a large number of identification methods for influential observations have been introduced and investigated extensively over the period. It is well evident that most of the popular diagnostic measures are based on single case deletion and mislead the analysis in presence of

multiple influential observations because of masking and/or swamping. We propose a new diagnostic measure for influential observations in linear regression.

9. A New Graphical Multivariate Outlier Detection Technique Using Singular Value Decomposition

- Nishith Kumar and Mohammed Nasser

[International Journal of Engineering Research & Technology (IJERT), Vol. 1 Issue 6, August – 2012]

Abstract. It is well recognized that even a high quality data set tends to contain a remarkable percent of unusual observations. There are various techniques to detect multivariate outliers. But they are highly depends on mathematics. We propose a new graphical technique to detect multivariate outliers which is easy to understand without hard mathematics, it can be applied to data for both supervised and unsupervised learning, it can be directly applied to separate extreme outliers from general outliers.

10. An Improved Version of Kurtosis Measure and their Application in ICA

- Md Shamim Reza, Mohammed Nasser and Md. Shahjaman

[International Journal of Wireless Communication and Information Systems (IJWCIS) Vol 1 No 1 April, 2011]

Abstract. Kurtosis plays important role in defining shape characteristics of a probability distribution, and also in extracting as well as sorting independent components. From recent research on various versions of classical kurtosis we see that all the measures substantially underestimate kurtosis parameter and exhibit high variability when underlying population distribution is highly skewed or heavy tailed. This is unwanted for independent component analysis (ICA). In this paper, we propose a bootstrap bias corrected kurtosis estimator.

11. Robust Regression Estimation in Generalized Linear Models

- Aishah Hamzah and Mohammed Nasser

[International Encyclopedia of Statistical Science 2011: 1242-1244]

Abstract. In experimental settings where data are collected with the goal of making inferences about some aspects of an underlying population it is always important to design the study in such a way as to obtain as much useful information as possible while minimizing the overall cost of the experiment. This is particularly true when the initial step in collecting these data is to select the particular units from the finite or infinite population on which measurements.

12. Outlier Detection in Linear Regression

- A.A.M Nurunnabi, A.H.M.R. Imon, A.B.M.S. Ali and Mohammed Nasser

Abstract. Regression analysis is one of the most important branches of multivariate statistical techniques. It is widely used in almost every field of research and application in multifactor data, which helps to investigate and to fit an unknown model for quantifying relations among observed variables. Nowadays, it has drawn a large attention to perform the tasks with neural networks, support vector machines, evolutionary algorithms, et cetera. ….

13. Regression Diagnostics in Large and High Dimensional Data

- A.A.M. Nurunnabi and Mohammed Nasser

[11th International Conference on Computer and Information Technology, 2008 DOI: 10.1109/ICCITECHN.2008.4802969, Khulna, Bangladesh]

Abstract. “Learning methods” play a key role in the fields of statistics, data mining, and artificial intelligence, intersecting with areas of engineering and other disciplines. These methods for analyzing and modeling data come in two flavors: supervised and unsupervised learning. Regression analysis and classification are two well known supervised learning techniques. To get an effective model from regression analysis it is necessary to check and preprocess the data set in astronomy, bio-informatics, image analysis, computer vision etc.

14. Estimators of Influence Function

- Mohammed Nasser and Mesbahul Alam

[Communication in Statistics- Theory and Methods 35(1):21-32 · February 2006]

Abstract. This article discusses estimators of influence function with some new counter-examples and tries to uphold their usefulness mathematically as well as through simulation. It is suggested that some estimators of influence function of uniformly Fréchet differentiable functional has more desirable properties.

15. Identification of Differentially Expressed Gene Using Robust Singular Value Decomposition

- Nishith Kumar, Tofazzal Hossain, Eric J. Beh, Masahiro Sugimoto and Mohammed Nasser

[Current Bioinformatics, Volume 11, Issue 3, 2016, Page: 366 – 374, DOI: 10.2174/1574893611999160610124913]

Abstract. Identification of differentially expressed genes (DEG) in transcriptomic analyses is one of the important tasks to find out significantly activated/deactivated pathways. Outliers and/or the missing values are commonly observed in microarray data; however, most available statistical methods did not deal with these issues and, therefore, their analytical results were frequently skewed and deteriorated. Here, we developed a novel technique robust against outliers and missing values: a dimension reduction procedure based on robust singular …..

16. Identification and Classification of Multiple Outliers, High Leverage Points and Influential Observations in Linear Regression

- A.A.M. Nurunnabi, M. Nasser and A.H.M.R. Imon

[Journal of Applied Statistics, Volume 43, 2016 - Issue 3, https://doi.org/10.1080/02664763.2015.1070806]

Abstract. Detection of multiple unusual observations such as outliers, high leverage points and influential observations (IOs) in regression is still a challenging task for statisticians due to the well-known masking and swamping effects. In this paper we introduce a robust influence distance that can identify multiple IOs, and propose a six-fold plotting technique based on the well-known group deletion approach to classify regular observations, outliers, high leverage points and IOs simultaneously in linear regression.

17. Modeling and Predicting of Children Ever Born in Bangladesh

- Faisal Ahmmed and Mohammed Nasser

[Conference: International Conference on Statistical Data Mining for Bioinformatics Health Agriculture and Environment, 2012, University of Rajshahi, Bangladesh]

Abstract. This paper explores to model as predict the average children of ever married women in Bangladesh using the different important factors associated with childbearing pattern. Bangladesh Demography Health Survey data (2007) have been used for this study. Firstly we identify the important factors associated with the children ever born. Then generalized linear model especially Poisson Model is applied to know the effect of these factors and calculate the predicting values. Support Vector Machines (SVM) is used.

18. Outlier Detection in Logistic Regression

- A.A.M. Nurunnabi, A.B.M.S. Ali, A.H.M. Rahmatullah Imon and Mohammed Nasser

[In book: Multidisciplinry Computational Intelligence Techniques: Applications in Business, Engineering and Medicine, Chapter: 16, January 2012, Publisher: IGI Global, USA, DOI: 10.4018/978-1-4666-1830-5.ch016]

Abstract. The use of logistic regression, its modelling and decision making from the estimated model and subsequent analysis has been drawn a great deal of attention since its inception. The current use of logistic regression methods includes epidemiology, biomedical research, criminology, ecology, engineering, pattern recognition, machine learning, wildlife biology, linguistics, business and finance, et cetera. Logistic regression diagnostics have attracted both theoreticians and practitioners in recent years.

19. Outlier Diagnostics in Logistic Regression: A Supervised Learning

-A.A.M. Nurunnabi and Mohammed Nasser

[International Conference on Machine Learning and Computing, IPCSIT vol.3, 2011, IACSIT Press, Singapore]

Abstract. The goal of supervised learning is to build a concise model of the distribution of class labels in terms of predictor features. Logistic regression is one of the most popular supervised learning technique that is used in classification.

Fields like computer vision, image analysis and engineering sciences frequently encounter data with outliers (noise). Presence of outliers in the training sample may be the cause of large training time, misclassification, and to design a faulty classifier. This article provides a new methodology.

20. Applications of Robust Regression in Business, Economics and Social Sciences

- Mohammed Nasser and A.H.M. Rahmatullah Imon

[North South Business Review, Vol. 2, No. 1, July, 2007, ISSN 1991-4938]

Abstract. Robust regression techniques are not widely used in business, economics or in social sciences. It is a reliable alternative, where ordinary least squares (LS) do not work due to unusual observations and the violations of normality assumptions of error distributions. We demonstrate the importance of robust regression techniques by studying and comparing with OLS. Three examples are taken from the literature in areas of business, economics and social sciences.

21. Knowledge Inequality between Male and Female on HIV/AIDS in Bangladesh

- A.A.M. Nurunnabi, A.H.M. Rahmatullah Imon and Mohammed Nasser

Abstract. Women accounted for nearly 41% of all people living with HIV worldwide in 1997, but this figure increased up to more than 50% by 2004. It is generally believed that the challenges to fight against HIV/AIDS are closely related with many economic and social factors of a country. But it is now evident that lack of knowledge and awareness about the causes and preventions regarding……

22. Recurrent Support and Relevance Vector Machines Based Model with Application to Forecasting Volatility of Financial Returns

-Altaf Hossain and Mohammed Nasser

[Journal of Intelligent Learning Systems and Applications, 2011, 3, 230-241 doi:10.4236/jilsa.2011.34026 Published Online November 2011 (http://www.SciRP.org/journal/jilsa)]

Abstract. In the recent years, the use of GARCH type (especially, ARMA-GARCH) models and computational-intelligence-based techniques—Support

Vector Machine (SVM) and Relevance Vector Machine (RVM) have been successfully used for financial forecasting. This paper deals with the application of ARMA-GARCH, recurrent SVM (RSVM) and recurrent RVM (RRVM) in volatility forecasting. Based on RSVM and RRVM, two GARCH methods are used and are compared with parametric GARCHs (Pure and ARMA-GARCH) in terms of their ability to forecast multi-periodically. These models are evaluated on four performance metrics: MSE, MAE, DS, and linear regression R squared. The real data in this study uses two Asian stock market composite indices of BSE SENSEX and NIKKEI225. This paper also examines the effects of outliers on modeling and forecasting volatility. Our experiment shows that both the RSVM and RRVM perform almost equally, but better than the GARCH type models in forecasting. The ARMA-GARCH model is superior to the pure GARCH and only the RRVM with RSVM hold the robustness properties in forecasting.

23. A Comparative Study of Kernel and Robust Canonical Correlation Analysis

-Ashad M. Alam, Mohammed Nasser and Kenji Fukumizu

[Journal of Multimedia, Vol. 5, No. 1, February 2010]

Abstract. A number of measures of canonical correlation coefficient are now used in multimedia related fields like object recognition, image segmentation facial expression recognition and 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 i) simulation from the view point of sensitivity, breakdown analysis as well as ii) using real data sets. In this paper an attempt has been made to compare performances of kernel canonical correlation coefficients (Gaussian function, Laplacian function and Polynomial function) with that of robust and classical canonical correlation coefficient measures using simulation with five sample sizes (50, 500, 1000, 1500 and 2000), influence function, breakdown point along with several real data and a multi-modal data sets, focusing on the specific case of segmented images with associated text. We investigate the bias, mean square error (MISE), qualitative robustness index (RI), sensitivity curve of each estimator under a variety of

situations and also employ box plots and scatter plots of canonical variates to judge their performances. We have observed 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 and break down is high in case of nonlinear data.

24. New second-order optimality conditions in multi objective optimization problems: Differentiable case

-M. M. Rizvi and M. Nasser

[J. Indian Inst. Sci., May–June 2006, 86, 279–286]

Abstract. To get positive Lagrange multipliers associated with each of the objective function, Maeda [Constraint qualification in multi objective optimization problems: Differentiable case, *J. Optimization Theory Appl*., 80, 483–500 (1994)], gave some special sets and derived some generalized regularity conditions for first-order Karush–Kuhn–Tucker (KKT)-type necessary conditions of multi objective optimization problems. Basing on Maeda's set, Bigi and Castellani [Second order optimality conditions for differentiable multi objective problems, *RAIRO, Op. Res.*, 34, 411–426 (2000)], tried to get the same result for second-order optimality conditions but their treatment was not convincing. In this paper, we have generalized these regularity conditions for second-order optimality conditions under different sets and obtained positive Lagrange multipliers for the objective function.

25. A New Singular Value Decomposition Based Robust Graphical Clustering Technique and Its Application in Climatic Data

-Nishith Kumar, Mohammed Nasser and Subaran Chandra Sarker

[www.ccsenet.org/jgg Journal of Geography and Geology Vol. 3, No. 1; September 2011]

Abstract. An attempt is made to study mathematical properties of singular value decomposition (SVD) and its data exploring capacity and to apply them to make exploratory type clustering for 10 climatic variables and thirty weather stations in Bangladesh using a newly developed graphical technique. Findings in SVD and

Robust singular value decomposition (RSVD) based graphs are compared with that of classical K-means cluster analysis, its robust version, partition by medoids (PAM) and classical factor analysis, and the comparison clearly demonstrates the advantage of SVD over its competitors. Lastly the method is tested on well known Hawkins-Bradu-Kass (1984) data.

26. Farmers' Perception Regarding Climate Change and Crop Production, Especially for Wheat in Dinajpur District

-J. A. Syeda and M. Nasser

[J. Environ. Sci. & Natural Resources, 5(2): 129 - 136, 2012 ISSN 1999-7361]

Abstract. An attempt was made to depict the valuable experience of farmers about climate change, environment and agricultural production, particularly wheat by conducting an opinion survey among 50 years and above aged farmers and agricultural workers in selected mauzas of Dinajpur district. Three hundred thirteen (313) respondents were interviewed in the survey. All the respondents opined regarding climate change in Dinajpur district over time. All of them opined that crop land, crop cultivation and crop yield were affected due to climatic change and changing of climate might pose a big and devastating threat to the production of wheat. Besides, the three case studies were accomplished to explore new ideas about climate change and the behavior of nature and human culture. They had also similar types of experience about climate change.

27. Qualitative Robustness in Estimation

-Mohammed Nasser, Nor Aishah Hamzah and Md. Ashad Alam

[Statistics in the Twenty-First Century: Special Volume, In Honour of Distinguished Professor Dr. Mir Masoom Ali, On the Occasion of his 75th Birthday Anniversary, PJSOR, Vol. 8, No. 3, pages 619-634, July 2012]

Abstract. Qualitative robustness, influence function, and breakdown point are three main concepts to judge an estimator from the viewpoint of robust estimation. It is important as well as interesting to study relation among them. This article attempts to present the concept of qualitative robustness as forwarded by first proponents and its later development. It illustrates intricacies of qualitative

robustness and its relation with consistency, and also tries to remove commonly believed misunderstandings about relation between influence function and qualitative robustness citing some examples from literature and providing a new counter-example. At the end it places a useful finite and a simulated version of qualitative robustness index (QRI). In order to assess the performance of the proposed measures, we have compared fifteen estimators of correlation coefficient using simulated as well as real data sets.

28. Intrusion Detection Using Combination of Various Kernels Based Support Vector Machine

-Md. Al Mehedi Hasan, Mohammed Nasser, Biprodip Pal and Shamim Ahmad

[International Journal of Scientific & Engineering Research, Volume 4, Issue 9, September-2013 ISSN 2229-5518]

Abstract. The success of any Intrusion Detection System (IDS) is a complicated problem due to its nonlinearity and the quantitative or qualitative network traffic data stream with many features. To get rid of this problem, several types of intrusion detection methods have been proposed and shown different levels of accuracy. This is why, the choice of the effective and robust method for IDS is very important topic in information security. In this paper, a combining classification approach to network intrusion detection based on the fusion of multiple classifiers is proposed. This approach makes a combination of various kernel based Support Vector Machine (SVM) classifier for intrusion detection system using majority voting fusion strategy. The experimental results indicate that combined approach effectively generates a more accurate model compared to single kernel based SVM classifier for the problem of intrusion detection.

29. On the KDD'99 Dataset: Support Vector Machine Based Intrusion Detection System (IDS) with Different Kernels

-Md. Al Mehedi Hasan, Mohammed Nasser and Biprodip Pal

[International Journal of Electronics Communication and Computer Engineering Volume 4, Issue 4, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209]

Abstract. The success of any Intrusion Detection System (IDS) is a complicated problem due to its nonlinearity and the quantitative or qualitative network traffic data stream with many features. To get rid of this problem, several types of intrusion detection methods have been proposed and shown different levels of accuracy. This is why, the choice of the effective and robust method for IDS is

very important topic in information security. Support vector machine (SVM) has been employed to provide potential solutions for the IDS problem. However, the practicability of SVM is affected due to the difficulty of selecting appropriate kernel and its parameters. Thus, this paper is aimed to use different kernel on the KDD'99 Dataset and find out which is best for SVM based intrusion detection system. In this work, we have developed a new data set, KDD99Train+ and KDD99Test+, which does not include any redundant records in the train set as well as in the test set which was an inherent problem of
KDD'99 dataset, so the classifiers will not be biased towards more frequent records. The experimental results indicate that RBF kernel can achieve higher detection rate than others kernel like Linear and polynomial kernel in the same time. RBF kernel also shows lower false negative rate than polynomial kernel.

30. Multiple Outliers Detection: Application to Research & Development Spending and Productivity Growth

-A. A. M. Nurunnabi and Mohammed Nasser

[BRAC University Journal, vol. V, no. 2, 2008, pp. 31-39]

Abstract. Multiple outliers are frequently encountered in applied studies in business and economics. Most of the practitioners depend on ordinary least squares (OLS) method for parameter estimation in regression analysis without identifying outliers properly. It is evident that OLS totally fails even in presence of single outlying observation. Single observation outlier detection methods are failed to identify multiple outliers due to masking and swamping effects. This paper analytically and numerically compares the sensitivity of the most popular diagnostic statistics. Data set from Griliches and Lichtenberg (1984) is used to show that we need to take extra care for model building process in presence of multiple outliers.

31. Application of Robust Regression and Bootstrapping in Purchasing Power Parity Analysis

-A. A. M. Nurunnabi and Mohammed Nasser

[Daffodil International University Journal of Business and Economics, Vol. 2, No. 1, January 2007]

Abstract. This article is an attempt to show how robust regression, a computer based statistical technique introduced by P. J. Huber in 1973 and later developed by Rousseeuw (1984), Rousseeuw and Yohai (1984), and many others, can helps

us in cases where OLS totally fails due to outliers, leverage points and non-normality of error distribution. To infer from the estimators obtained from robust regression we generally need, especially for small samples, bootstrapping (resampling) technique that is also a computer intensive statistical technique introduced by Efron (1979), and later developed in many directions. This talk illustrates the whole thing by an example using data extracted from the Big Mac. Index with a purchasing power parity analysis.

32. Feature Selection for Intrusion Detection Using Random Forest

-Md. Al Mehedi Hasan, Mohammed Nasser, Shamim Ahmad, Khademul Islam Molla

[Journal of Information Security, 2016, 7, 129-140, Published Online April 2016 in SciRes. http://www.scirp.org/journal/jis]

Abstract. An intrusion detection system collects and analyzes information from different areas within a computer or a network to identify possible security threats that include threats from both outside as well as inside of the organization. It deals with large amount of data, which contains various irrelevant and redundant features and results in increased processing time and low detection rate. Therefore, feature selection should be treated as an indispensable pre-processing step to improve the overall system performance significantly while mining on huge datasets. In this context, in this paper, we focus on a two-step approach of feature selection based on Random Forest. The first step selects the features with higher variable importance score and guides the initialization of search process for the second step whose outputs the final feature subset for classification and interpretation. The effectiveness of this algorithm is demonstrated on KDD'99 intrusion detection datasets, which are based on DARPA 98 dataset, provides labeled data for researchers working in the field of intrusion detection. The important deficiency in the KDD'99 data set is the huge number of redundant records as observed earlier. Therefore, we have derived a data set RRE-KDD by eliminating redundant record from KDD'99 train and test dataset, so the classifiers and feature selection method will not be biased towards more frequent records. This RRE-KDD consists of both KDD99Train+ and KDD99Test+ dataset for training and testing purposes, respectively. The experimental results show that the Random Forest based proposed approach can select most important and relevant features useful for classification, which, in turn, reduces not only the number of input features and time but also increases the classification accuracy.

33. Image Compression and Denoising Using Wavelet Transformation, Fourier Transformation and Singular Value Decomposition (Classical and Robust)

-Md. Tofazzal Hossain, Nishith Kumar and Mohammed Nasser

[Global Journal of Quantitative Science, Vol. 3. No.1. Mar 2016 Issue. Pp 8-16.]

Abstract. In the present age of technology, we have to store, transmit and analyze huge images that necessitate image compression and denoising. There are many techniques for image compression and denoising such as Wavelet transformation, Fourier transformation, classical and robust singular value decomposition, Neural networks, Principal Component Analysis etc. Since in literature no comparison among wavelet transformation, Fourier transformation, and singular value decomposition have been found, in this paper, an attempt has been made to compare the three most popular as well as influential techniques in image compression and denoising. In image compression and random denoising, the performance of wavelet transformation is better but in text denoising, robust singular value decomposition (RSVD) is better among the four techniques. Therefore, we recommend to use wavelet transformation for image compression and random denoising and robust singular value decomposition (RSVD) for text denoising.

34. Multivariate Data Visualization: Correspondence Analysis, Classical and Robust Singular Value Decomposition and Depth based Approach

-Nishith Kumar and Mohammed Nasser

[International Journal of Advanced Research (2016), Volume 4, Issue 1, 416- 425, ISSN 2320-5407]

Abstract. Now a day's Data Mining is one of the challenging area in statistics as well as in computer science. Data visualization is one of the most important parts in Data Mining. We can only visualize two or three dimensional data but for Data Mining much more than three dimensions is usual rather than exception. So data reduction is very important for visualization of multivariate data. For this reason, in this article I would like to introduce four well known, effective and

sophisticated scientific data reduction methods, correspondence analysis, singular value decomposition (SVD), robust singular value decomposition (RSVD) and depth for data visualization, pattern recognition and outlier detection. Since these techniques are used for data reduction technique so by using these techniques we can visualize data taking only two or three dimensions that maximize the total variation of data. In many cases two or three singular values or eigen values cannot explain most (greater than 80%) of the variation of data. In that case we can use L1 depth, half space depth and kernel based depth for visualizing data. . In this paper we have used four well known real dataset (fisher's iris data, Wisconsin breast cancer data, Glass identification data and Seeds data) for visualization and also rigorously explained the results on the motion of the aforementioned techniques.

35. Time Series Properties of Some Climatic Variables in Dinajpur District

-M. J. Rahman, J. A. Syeda and M. Nasser

[J. Environ. Sci. & Natural Resources, 6(2): 127-137, 2013 ISSN 1999-7361]

Abstract. Several direct empirical time series investigations of global climate change and its impact have been studied by several world famous researchers. Some researches regarding local climatic change and its impact have been published but the time series properties of the variables related to national as well as local climate are yet to be able to have proper attention. The presence or absence of unit roots in these time series or inappropriate statistical tools may challenge the validity of the interpretations of their results and implies that cointegration analysis can be used to investigate the relation among variables. This article attempts to deduce time series properties of temperature, rainfall and humidity of Dinajpur district.

36. Trend and Variability Analysis and Forecasting of Sunshine-Hour in Bangladesh

-J. A. Syeda and M. Nasser

[J. Environ. Sci. & Natural Resources, 5(2): 109-118, 2012 ISSN 1999-7361]

Abstract. An attempt has been taken to investigate the trend and variability pattern for annual and seasonal (Three crop seasons) average sunshine-hours (ASSH) for six divisional stations of Bangladesh: Dhaka, Khulna, Rajshahi, Barisal, Sylhet and Chittagong. The monthly ASSH (2008-2011) are forecasted applying the univariate Box-Jenkin's ARIMA (autoregressive integrated moving average) modelling technique on the basis of the minimum root mean square

forecasting error. Several diagnostic techniques are used for detection of outlier and residual's stationarity and normality. The significant negative rates are observed for linear trend (LT) of Annual and Seasonal Average ASSH in all the six stations with only one minor positive rate for Prekharif season in Khulna. The findings support that the climate of Bangladesh is changing in terms of sunshine-hour which may have tremendous effects in the agricultural production.

37. Simulation Based Comparison among Fifteen Estimators of Correlation Coefficient

-Md. Ashad Alam and Mohammed Nasser

[International Journal of Statistical Sciences, ISSN 1683-5603, Vol. 8, 2008, pp 61-74]

Abstract. In this paper, fifteen estimators of correlation coefficient available in the literature have been compared through simulation. We have considered simulating sampling distribution at bivariate standard normal and its five contaminated forms with three different correlation coefficients and three different sample sizes. We also have considered a real population of size 1491.The estimators have been compared with regard to bias, standard error, mean square error and length of 90% percentile interval. The class of robust estimators, especially the normal score estimator has been found superior to the class of non-robust estimators over all. To measure correlation coefficient we have recommend using for normal score estimator especially when sample is large and contamination is probable.

38. Heuristics of Influence Function

-M. Nasser

[International Journal of Statistical Sciences, ISSN 1683–5603, Vol. 6 (Special Issue), 2007, pp 305-314]

Abstract. This article discusses heuristics of influence functions, an important concept in robust statistics, with some new counter-examples. It re-emphasises the caution raised by Davies (1993, 1994 and 1998) about using heuristics of influence function thoughtlessly. It is suggested that the influence functions of more smooth functionals have more desirable properties. Finally some examples of more smooth functionals including a new class of uniformly Fr´echet differentiable L- location functional is presented.