ECRONICON OPEN ACCESS

EC CLINICAL AND MEDICAL CASE REPORTS

Research Article

Low Classification Accuracy by Logistic Regression, Support Vector Classifier, and Multi-Layer Perceptron, But Not Decision Tree, on Random Attributes from Hadamard Matrix

Maurice HT Ling*

HOHY PTE LTD, Singapore

***Corresponding Author:** Maurice HT Ling, HOHY PTE LTD, Singapore.

Received: October 16, 2020; **Published:** November 28, 2020

Abstract

The use of machine learning classifiers is increasing with evidence of overtaking human judgement. This can be risky if workings and implications of machine learning classifiers remain a black box. Here, a case where a balanced and algorithmically generated data set, Hadamard matrix, classifies poorer than random using logistic regression (accuracy < 17.4%), support vector classifier (accuracy < 23.4%) and in most cases of multi-layer perceptron (accuracy < 27.9%) but not in decision tree (accuracy > 77.3%); despite perfect (100%) internal classification accuracy for both support vector classifier and multi-layer perceptron; is reported. This suggests a systematic and yet currently unexplained source of error.

Keywords: *Logistic Regression; Hadamard Matrix; Vector Classifier; Multi-Layer Perceptron; Support Vector Classifier*

Introduction

The application of machine learning is gaining prevalence in a wide variety of fields, from material science [1] to finance [2], healthcare [3] and even to law enforcement [4]. With pervasive use of machine learning, especially in decision support systems [5], there is a call to ensure that results from machine learning algorithms are interpretable [6]. This call is both timely and urgent as several studies [7,8] suggest a tendency to accept an advice from technology than from another person. Everyone experienced personal embarrassing episodes from auto-correction and/or auto-completion. While these can be embarrassing, they are usually harmless. What if such reliance can result in life-threatening situations? Matthew Grissinger [9] reports two incidences where wrong medication was administered to patients against correct prescription orders as a result of erroneous auto-completion of medication names on the display of automated dispensing cabinets. A meta-analysis also suggests that clinicians overrode their own correct decisions in favour of erroneous advice from technology between 6% and 11% of the time [10].

These boil down to the core question of whether the path from attributes to results can be described [11]. Here, a case where a balanced and algorithmically generated data set, Hadamard matrix, classifies poorly using of logistic regression (accuracy < 17.4%), support vector classifier (accuracy < 23.4%), and most instances of multi-layer perceptron (accuracy < 27.9%) but not decision tree (accuracy > 77.3%); despite perfect (100%) accuracy between input attributes and labels for support vector classifier and multi-layer perceptron is reported.

Methods

A 1024-by-1024 Hadamard matrix [12] constructed using Sylvester's construction [13] in SciPy [14] version 1.4.1 and converted into a Pandas data frame with Pandas [15] version 1.0.4. Random number of columns; representing 10% to 100% of the matrix, in increment of 10%; were randomly selected. Each selection will have varying numbers of columns with 1024 rows each, denoting 1024 data points. Within this set of selected columns, a randomly selected column was used as label while the rest of the columns were used as attributes. The resulting 1024 pairs of label and attribute list were then used to train four classifiers from Scikit-Learn [16,17] version 0.23.1; namely, (a) Support Vector Classifier [18], (b) Multilayer Perceptron [19], (c) Decision Tree [20], and (d) Logistic Regression. Defaults parameters were used. For each classifier, the mean accuracy score from the input attributes and labels, and 10-fold cross validation [21] accuracy were reported. A total of 10 replicates were performed, resulting in 100 runs (10 replicates of 10 number of columns) for each classifier.

Results and Discussion

The Hadamard matrix used in this study was constructed using Sylvester's construction [13] in SciPy [14]. Hence, the first column in the matrix consists of only 1s while the rest of the 1023 columns (2^{nd} column to 1024^{th} column) consist of equal numbers of 1s and 0s. There is no correlation ($r = 0$) between any two rows or between any two columns when the entire matrix is taken but a weak ($r = -0.00098$) and insignificant ($df = 1021$, $t = 0.0288$, p-value = 0.9770) correlation between the rows when the first column is removed. As a random selection of columns were made for each replicate, the values are balanced for each attribute (as columns) unless the first column is selected, with no correlations between attributes or individual samples (as rows).

Of the four classifiers evaluated using 10-fold cross validation (Figure 1), only decision tree consistently performs better than random (accuracy > 77.3%) regardless of the number of randomly selected attributes used. Both logistic regression (accuracy < 17.4%) and support vector classifier (accuracy < 23.4%) consistently performs poorer than random. Multi-layer perceptron performs poorer than random (accuracy < 27.9%) when the number of attributes selected exceed 30% of the total Hadamard matrix. The 10-fold cross validation results from support vector classifier, and multi-layer perceptron are surprising considering perfect (100%) accuracy between input attributes and labels in every case.

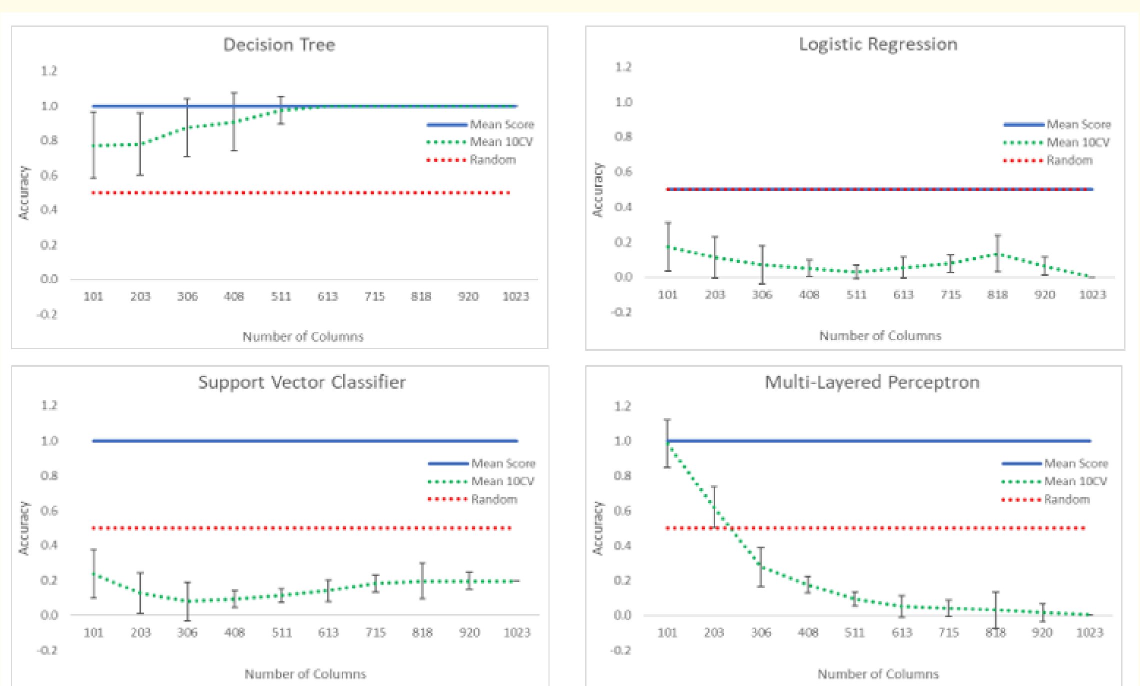

Figure 1: *Binary classification accuracy. number of columns (x-axis) of 101, 203, 306, 408, 511, 613, 715, 818, 920, and 1023 represents the number of randomly selected columns from 1024 x 1024 Hadamard matrix. Mean Score (solid blue line) is the grand mean (mean of means) of 10 replicates of accuracy on given data and label. Mean 10CV (dotted green line) is the grand mean of 10 replicates of 10-fold cross-validation accuracy. Random (dotted red line) represents 50% theoretical random classification accuracy. Error bars denotes standard error.*

The mean accuracy between the input attributes and labels is commonly used by the classification algorithm to determine when iteration halts, which can be considered as internal classification accuracy. In contrast, 10-fold cross-validation [22] requires that the test set is disjoint from the training set, which can be considered as external classification accuracy. Since Hadamard matrices are constructed procedurally [23,24] and using Sylvester's construction [13] in this case, it is mathematically defined; hence, poorer than random external classification accuracies from both logistic regression and support vector classifier suggest an unexplained yet systemic cause. This is further supported by the large difference of more than 70% between internal and external classification accuracy. A better than random external classification accuracy in multi-layer perceptron when the number of attributes is lesser than 30% of total suggests the reduced role of the underlying systemic cause but the reverse trend is seen in decision tree. Moreover, the internal classification accuracy from logistic regression is not better than random, suggesting that a strong systemic effect from the data set – Hadamard matrix.

Without a clear understanding of this currently unexplained systemic effect, it may be possible for unexpected failures in currently machine learning classification algorithms. This may have serious implications to critical systems; such as, dispensing systems [9] and clinical decision support systems [10]. Currently, the only way to avoid expected errors is by extensive validation using previously unseen data sets; thus, underpinning the importance of cross-validation [25] as the first validation step. This is followed by an in-depth analysis and quantification of the classification errors [26] through the use of confusion matrix [27].

Conclusion

Here, a case where a balanced and algorithmically generated data set, Hadamard matrix, classifies poorer than random using logistic regression, support vector classifier, and multi-layer perceptron is reported. This indicates the presence of a systematic and yet currently unexplained source of classification error that requires further studies.

Conflict of Interest

The authors declare no conflict of interest.

Bibliography

1. Schmidt J., *et al*. "Recent Advances and Applications of Machine Learning in Solid-State Materials Science". *NPJ Computational Materials* 5.1 (2019): 83.

2. Gerlein EA., *et al*. "Evaluating Machine Learning Classification for Financial Trading: An Empirical Approach". *Expert Systems with Applications* 54 (2016): 193-207.

3. Aljaaf AJ., *et al*. "Applied Machine Learning Classifiers for Medical Applications: Clarifying the Behavioural Patterns Using A Variety of Datasets". In: 2015 International Conference on Systems, Signals and Image Processing (IWSSIP). London, United Kingdom: IEEE (2015): 228-232.

4. Vestby A and Vestby J. "Machine Learning and the Police: Asking the Right Questions". *Policing An International Journal of Police* (2019): paz035.

5. Merkert J., *et al*. "A Survey of the Application of Machine Learning in Decision Support Systems. In: 23rd European Conference on Information Systems (ECIS 2015). University of Münster, Münster, Germany (2015): 133.

6. Stiglic G., *et al*. "Interpretability of Machine Learning-Based Prediction Models in Healthcare". *WIREs Data Mining and Knowledge Discovery* 10.5 (2010): e1379.

7. Logg JM. "Theory of Machine: When Do People Rely On Algorithms?" Harvard Business School (2017).

8. Logg JM., *et al*. "Algorithm Appreciation: People Prefer Algorithmic to Human Judgment". *Organizational Behavior and Human Decision Processes* 151 (2019): 90-103.
9. Grissinger M. "Understanding Human Over-Reliance On Technology". *PT Peer-Rev Journal Formula Manag* 44.6 (2019): 320-375.
10. Goddard K., *et al*. "Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators". *Journal of the American Medical Informatics Association* 19.1 (2012): 121-127.
11. Murdoch WJ., *et al*. "Definitions, Methods, and Applications in Interpretable Machine Learning". *Proceedings of the National Academy of Sciences* 116.44 (2019): 22071-22080.
12. Hadamard J. "Resolution D'une Question Relative Aux Determinants". *Bulletin des Sciences Mathématiques* 17 (1893): 240-246.
13. Sylvester JJ. "Thoughts on Inverse Orthogonal Matrices, Simultaneous Sign Successions, and Tessellated Pavements in Two or More Colours, with Applications to Newton's Rule, Ornamental Tile-Work, and the Theory of Numbers". *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 34.232 (1867): 461-475.
14. Virtanen P., *et al*. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". *Nature Methods* 17.3 (2020): 261-272.
15. McKinney W. "Data Structures for Statistical Computing in Python. In: Proceedings of the 9th Python in Science Conference". Austin, Texas, USA (2010): 51-56.
16. Buitinck L., *et al*. "API Design for Machine Learning Software: Experiences from the Scikit-learn Project". In: ECML PKDD Workshop: Languages for Data Mining and Machine Learning (2013): 108-122.
17. Pedregosa F., *et al*. "Scikit-learn: Machine learning in Python". *Journal of Machine Learning Research* 12 (2011): 2825-2830.
18. Cortes C and Vapnik V. "Support-Vector Networks". *Machine Learning* 20.3 (1995): 273-297.
19. Popescu M-C., *et al*. "Multilayer Perceptron and Neural Networks". *WSEAS Transactions on Circuits and Systems* 8.7 (2009): 579-588.
20. Breiman L., *et al*. "Classification and Regression Trees". CRC Press (1984).
21. Koul A., *et al*. "Cross-Validation Approaches for Replicability in Psychology". *Frontiers in Psychology* 9 (2018): 1117.
22. Arlot S and Celisse A. "A Survey of Cross-Validation Procedures for Model Selection". *Statistics Surveys* 4 (20010): 40-79.
23. Manjhi P and Kumar A. "On the Construction of Hadamard Matrices". *International Journal of Pure and Applied Mathematics* 120.1 (2018): 51-58.
24. Miyamoto M. "A Construction of Hadamard Matrices". *Journal of Combinatorial Theory, Series A* 57.1 (1991): 86-108.
25. De Rooij M and Weeda W. "Cross-Validation: A Method Every Psychologist Should Know". *Advances in Methods and Practices in Psychological Science* 3.2 (2020): 248-263.
26. Ling MH., *et al*. "Parts-of-Speech Tagger Errors Do Not Necessarily Degrade Accuracy in Extracting Information from Biomedical Text". *Python Pap* 3.1 (2008): 64-79.
27. Stehman SV. "Selecting and Interpreting Measures of Thematic Classification Accuracy". *Remote Sensing of Environment* 62.1 (1997): 77-89.

Volume 3 Issue 12 December 2020