Optimal Data Analysis
Vol. 8 (April 22, 2019), 94-96

2155-0182/10/$3.00

Multi-Layer Perceptron Neural Net Model Identifies Effect in Random Data

Ariel Linden, Dr.P.H. and Paul R. Yarnold, Ph.D.
Linden Consulting Group, LLC   Optimal Data Analysis, LLC

Prior research contrasted the ability of different classification algorithms [logistic regression (LR), random forests (RF), boosted regression (BR), support vector machines (SVM), classification tree analysis (CTA)] to correctly *fail to identify* a relationship between a binary class (dependent) variable and ten *randomly generated* attributes (covariates): only CTA found no relationship. In this paper, using the same ten-variable N=1,000 dataset, a Weka™ multi-layer perceptron (MLP) neural net model[1] using its default tuning parameters yielded (*a*rea *u*nder the *c*urve) AUC=0.724 in training analysis, and AUC=0.507 in ten-fold cross-validation. With the exception of CTA, all machine-learning algorithms assessed thus far have identified training effects in completely random data.

Predictive accuracy of CTA was compared with accuracy obtained using LR, RF, BR and SVM classification algorithms.[2,3] In that research, an artificial dataset was created with 500 “group 1” and 500 “group 2” observations as well as ten randomly created continuous covariates (attributes) which by design have no association with the binary dependent (class) variable. Among all algorithms tested *only CTA* correctly failed to discriminate among random groups.

This study assesses whether a consistent finding occurs for models identified by MLP analysis using the default tuning parameters: learning rate 0.3, momentum 0.2, 500 training epochs, and hidden neurons equal to (10 inputs and 2 classes/2).

As done previously[2,3], receiver operating characteristics (ROC) analysis was conducted treating actual class status as the reference variable, and treating predicted probabilities from the MLP model as the classification variable. A model that perfectly discriminates two groups has AUC=1.0, and a model that provides chance-level discrimination has AUC=0.50.

Results obtained by MLP in training analysis are summarized in Table 1: CTA was unable to identify a model so reproducibility analysis wasn’t possible.[2,3] The corresponding *e*ffect *s*trength for *s*ensitivity or ESS index (on which 0 is the classification accuracy expected by chance, and 100 is perfect accuracy) is ESS= 39.4, indicating an effect of moderate strength.[4]

Table 1: MLP Model Training Results

		Predicted Class		
		Group 1	Group 2	*Sensitivity*
Actual	Group 1	322	178	64.4
Class	Group 2	125	375	75.0
Predictive Value		72.0	67.8	

2155-0182/10/$3.00

Ten-fold cross-validation analysis results obtained for the MLP model are given in Table 2 (ESS=0.8, a very-near-chance result).

Table 2: MLP Model 10-Fold Results

		Predicted Class		
		Group 1	Group 2	*Sensitivity*
Actual	Group 1	272	228	54.4
Class	Group 2	268	232	46.4
Predictive Value		50.4	50.4	

The failure of the MLP model to replicate in cross-validation reasserts the need to conduct reproducibility analysis and supports general constraint of CTA models to only retain attributes having stable effects in training and LOO analysis within the model.[5-7]

This paper supports the results obtained in previous analysis[2,3] identifying an important limitation of machine learning algorithms often used for predicting binary outcomes (e.g., to obtain propensity scores[8-11]). That is, an MLP model is likely to find relationships between variables which really don't exist. These results should be independently replicated, and the limits of this phenomenon should be identified. Future research should assess the effect of the number of random attributes available to the algorithms, the number of significant digits used for measures (an index of precision of measurement), and the number of class category levels in the application, in affecting the training and validity AUC. Research should also investigate applications involving randomized *categorical* attributes having different numbers of levels.

Finally, the findings continue to support our recommendation to employ the ODA and CTA frameworks to draw causal inferences regarding treatment effects in observational data, and in data from randomized controlled trials.[8-24] A large, ever-increasing body of evidence supports the use of ODA and CTA to evaluate the efficacy of health-improvement interventions and policy initiatives.[25-27]

References

[1]Weka™ multi-layer perceptron neural net. Found at: http://www.cs.waikato.ac.nz/ml/weka.

[2]Linden A, Yarnold PR (2019). Some machine learning algorithms find relationships between variables when none exist -- CTA doesn't. *Optimal Data Analysis*, *8*, 64-67.

[3]Linden A, Yarnold PR (2019). Effect of sample size on discovery of relationships in random data by classification algorithms. *Optimal Data Analysis*, *8*, 76-80.

[4]Yarnold PR, Soltysik RC (2005). *Optimal Data Analysis: Guidebook with Software for Windows*. Washington, D.C.: APA Books.

[5]Yarnold PR (2016). Using UniODA to determine the ESS of a CTA model in LOO analysis. *Optimal Data Analysis*, *5*, 3-10.

[6]Yarnold PR (2016). Determining jackknife ESS for a CTA model with chaotic instability. *Optimal Data Analysis*, *5*, 11-14.

[7]Yarnold PR, Soltysik RC (2016). *Maximizing Predictive Accuracy*. Chicago, IL: ODA Books. DOI: 10.13140/RG.2.1.1368.3286

[8]Linden A, Yarnold PR (2016). Combining machine learning and propensity score weighting to estimate causal effects in multivalued treatments. *Journal of Evaluation in Clinical Practice*, *22*, 875-885.

[9]Linden A, Yarnold PR (2017). Using classification tree analysis to generate propensity score weights. *Journal of Evaluation in Clinical Practice*, *23*, 703-712.

[10]Yarnold PR, Linden A (2017). Computing propensity score weights for CTA models involving perfectly predicted endpoints. *Optimal Data Analysis*, *6*, 43-46.

2155-0182/10/$3.00

[11]Linden A, Yarnold PR (2018). Estimating causal effects for survival (time-to-event) outcomes by combining classification tree analysis and propensity score weighting. *Journal of Evaluation in Clinical Practice*, *24*, 380-387.

[12]Linden A, Yarnold PR (2016). Using data mining techniques to characterize participation in observational studies. *Journal of Evaluation in Clinical Practice*, *22*, 839-847.

[13]Linden A, Yarnold PR (2016). Using machine learning to assess covariate balance in matching studies. *Journal of Evaluation in Clinical Practice*, *22*, 848-854.

[14]Linden A, Yarnold PR (2016). Using machine learning to identify structural breaks in single-group interrupted time series designs. *Journal of Evaluation in Clinical Practice*, *22*, 855-859.

[15]Linden A, Yarnold PR, Nallomothu BK (2016). Using machine learning to model dose-response relationships. *Journal of Evaluation in Clinical Practice*, *22*, 860-867.

[16]Yarnold PR, Linden A (2016). Using machine learning to model dose-response relationships via ODA: Eliminating response variable baseline variation by ipsative standardization. *Optimal Data Analysis*, *5*, 41-52.

[17]Linden A, Yarnold PR (2016). Combining machine learning and matching techniques to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice*, *22*, 868-874.

[18]Linden A, Yarnold PR (2017). Minimizing imbalances on patient characteristics between treatment groups in randomized trials using classification tree analysis. *Journal of Evaluation in Clinical Practice*, *23*, 1309-1315.

[19]Linden A, Yarnold PR (2017). Modeling time-to-event (survival) data using classification tree analysis. *Journal of Evaluation in Clinical Practice*, *23*, 1299-1308.

[20]Linden A, Yarnold PR (2018). The Australian gun buy-back program and the rate of suicide by firearm. *Optimal Data Analysis*, *7*, 28-35.

[21]Linden A, Yarnold PR (2018). Identifying causal mechanisms in health care interventions using classification tree analysis. *Journal of Evaluation in Clinical Practice*, *24*, 353-361.

[22]Linden A, Yarnold PR (2018). Using machine learning to evaluate treatment effects in multiple-group interrupted time series analysis. *Journal of Evaluation in Clinical Practice*, *24*, 740-744.

[23]Linden A, Yarnold PR (2018). Using ODA in the evaluation of randomized controlled trials. *Optimal Data Analysis*, *7*, 46-49.

[24]Linden A, Yarnold PR (2018). Using ODA in the evaluation of randomized controlled trials: Application to survival outcomes. *Optimal Data Analysis*, *7*, 50-53.

[25]Linden A, Adams J, Roberts N (October, 2003). *Evaluation methods in disease management: determining program effectiveness*. Position Paper for the Disease Management Association of America (DMAA).

[26]Linden A, Roberts N (2005). A Users guide to the disease management literature: recommendations for reporting and assessing program outcomes. *American Journal of Managed Care,* 11, 81-90.

[27]Linden A, Adler-Milstein J (2008). Medicare disease management in policy context. *Health Care Finance Review*, 29, 1-11.

Author Notes

No conflict of interest was reported.