Optimal Data Analysis
Vol. 7 (March 2, 2018), 14-15

2155-0182/10/$3.00

Learning About the ODA Paradigm

Paul R. Yarnold, Ph.D.
Optimal Data Analysis, LLC

This note suggests tips on learning about ODA, from a cursory understanding through paradigm mastery.

Optimal discriminant (or data) analysis (ODA) is a machine-learning algorithm introduced more than a quarter-century ago.[1]

Gist

Here is link to a free, brief article with the fastest, broadest description of the ODA paradigm.[2]

Gestalt

Two books and a handful of articles cover the ontogenesis of the paradigm through its current metamorphosis as the conceptual equivalent of quantum mechanics for classical data.

The 2005 book, written when ODA was in its infancy, efficiently provides one with sufficient information to rapidly understand, conduct, and interpret ODA analysis results, for data reflecting many widely-used experimental designs.[3]

The 2016 book—written after novometric (Latin: new measure) theory was completed—summarizes everything known about the ODA paradigm at the time of its publication.[4]

Finally, a suite of the most recent articles discuss using ODA in causal inference research.[5-15]

Mastery

Sequentially working each chapter of both of the books—and reading articles on ODA cited within them, and experimenting with software settings (the 2005 book includes the UniODA™ software system), is highly recommended. Run various command configurations (e.g., vary the use of prior odds weighting, directional hypotheses, number of class categories, etcetera) for every example problem, to become familiar with the effect of parameter settings on analysis results.

It is a capital idea to attempt to replicate the analysis illustrated in each example problem using other data sets, and publish the results. And, it is an excellent idea to experiment using data with which one is personally familiar, including prior studies for which data are available. Familiarity with the data will simplify conducting ODA and interpreting resulting models, and thereby provide an excellent basis for objective comparison of legacy versus optimal statistical methods.

In general, when learning or introducing a new statistical paradigm it is a good idea to begin with a simple design and subsequently add complexity.

References

[1]Yarnold PR, Soltysik RC (1991). Theoretical distributions of optima for univariate discrimination of random data. *Decision Sciences*, *22*, 739-752.

2155-0182/10/$3.00

[2]Yarnold PR (2017). What is optimal data analysis? *Optimal Data Analysis*, *6*, 26-42. URL: http://optimalprediction.com/files/pdf/V6A7.pdf

[3]Yarnold PR, Soltysik RC (2005). *Optimal data analysis: A guidebook with software for Windows*. Washington, DC, APA Books.

[4]Yarnold PR, Soltysik RC (2016). *Maximizing predictive accuracy*. Chicago, IL: ODA Books. DOI: 10.13140/RG.2.1.1368.3286

[5]Linden A, Yarnold PR (2016). Using data mining techniques to characterize participation in observational studies. *Journal of Evaluation in Clinical Practice*, *6*, 839-847.

[6]Linden A, Yarnold PR (2016). Using machine learning to assess covariate balance in matching studies. *Journal of Evaluation in Clinical Practice*, *6*, 848-854.

[7]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*, *6*, 855-859.

[8]Linden A, Yarnold PR, Nallamothu BK (2016). Using machine learning to model dose-response relationships. *Journal of Evaluation in Clinical Practice*, *6*, 860-867.

[9]Linden A, Yarnold PR (2016). Combining machine learning and matching techniques to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice*, *6*, 868-874.

[10]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*, *6*, 875-885.

[11]Linden A, Yarnold PR (2017). Using classification tree analysis to generate propensity score weights. *Journal of Evaluation in Clinical Practice*. DOI: 10.1111/jep.12744

[12]Linden A, Yarnold PR (2017). Modeling time-to-event (survival) data using classification tree analysis. *Journal of Evaluation in Clinical Practice*. DOI: 10.1111/jep.12779

[13]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*. DOI: 10.1111/jep.12792

[14]Linden A, Yarnold PR (2017). Identifying causal mechanisms in health care interventions using classification tree analysis. *Journal of Evaluation in Clinical Practice*. DOI: 10.1111/jep.12848

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

Author's Notes

No IRB review is needed for this programming note, and no conflict of interest was reported.

Mail: Optimal Data Analysis, LLC
6348 N. Milwaukee Ave., Suite 163
Chicago, IL 60646