Machine and deep learning in MS research are just powerful statistics – No

Brian C Healy

Multiple Sclerosis Journal
2021, Vol. 27(5) 663–664
DOI: 10.1177/1352458520978648

Correspondence to:
BC Healy
Partners Multiple Sclerosis Center, Brigham and Women's Hospital, 09002O, 60 Fenwood Road, Boston, MA 02115, USA.
bchealy@mgh.harvard.edu

Brian C Healy
Partners Multiple Sclerosis Center, Brigham and Women's Hospital, Boston, MA, USA/Department of Neurology, Harvard Medical School, Boston, MA, USA/Biostatistics Center, Massachusetts General Hospital, Boston, MA, USA

Over the past several years, there has been an explosion in the use of machine and deep learning in medical research. A search of PubMed in October of 2020 for the term 'deep learning' returned less than 20 papers per year prior to 2012, but more than 5000 papers already in 2020. Although traditional statistics, machine learning and deep learning often try to solve similar problems; machine and deep learning have become more common in the literature because they are more than just powerful statistics.

In order to understand the similarities and differences between statistics, machine learning, and deep learning, it helps to define each term. Statistics is defined as 'a branch of mathematics dealing with collection, analysis, interpretation and presentation of masses of numerical data'.[1] This very broad definition shows that statistics encompasses a wide range of topics related to data analysis and data science, which is why many fields overlap with statistics. Machine learning and deep learning generally focus on a subset of tasks related to statistics. Machine learning has been defined as 'a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data' (p. 1).[2] Machine learning encompasses many approaches, and one of the first approaches discussed in many machine-learning textbooks is regression (linear regression for continuous outcomes and logistic regression for dichotomous outcomes), which is also commonly described in introductory statistics courses. Deep learning is a specific field of machine learning that uses neural networks to build prediction models including many hidden layers, which allow very complex relationships between the predictors and the outcome to be discovered.[3]

Most research questions in multiple sclerosis can be divided into one of three classes: description, causal inference or prediction.[4] Studies focused on description will provide summary statistics (and potentially associated 95% confidence intervals) related to specific characteristics of a group or attempt to cluster subjects into groups. For causal inference, statistical methods for group comparisons like the *t*-test or log-rank test can be used to estimate the causal effect from randomized clinical trials, while more complex methods like propensity score matching or inverse probability weighting are needed to control for confounding in observational studies.[5] For questions related to description and causal inference, machine learning and deep learning may have a role, but these approaches are generally not needed.

For prediction, the set of possible methods is broad, ranging from traditional statistical approaches like linear/logistic regression to machine learning and deep learning. The reason for the large number of potential models for prediction is that no approach is best in all settings. In the textbook *An Introduction to Statistical Learning*, the authors place analytic approaches on a graph showing that increasing flexibility of an approach often occurs at the cost of decreasing interpretability (Figure 2.7).[6] Most of the approaches researchers would consider as statistics (regression/least squares, best subsets regression, lasso) have high interpretability with corresponding limited flexibility. The high interpretability is due to the analyst specifying a model, and the model's coefficients having a specific scientific meaning. The limited flexibility comes from the fact that these models rarely include complex relationships between a predictor and the outcome. Conversely, approaches that are commonly considered machine learning like random forests, support vector machines and deep learning have high flexibility and limited interpretability. These approaches place limited structure on the data and allow the data to determine the relationships among the variables. This flexibility means these models can have improved predictive accuracy compared to traditional statistical approaches by identifying complex relationships,[7] but these relationships are often difficult to interpret and describe to users.

If machine learning aims to address the same prediction problems as traditional statistical approaches by increasing the flexibility at the cost of interpretability, why can't we conclude that machine learning is just powerful statistics? Although the techniques are similar, machine learning and traditional statistics have three important distinctions: (1) the main goal of the analysis, (2) the amount of user involvement and (3) the problems that are best suited to each approach. First, machine learning is often focused on maximizing predictive accuracy of a model, and statistics is often focused on making an inference about a population statistic based on a sample.[8] Furthermore, statistics is often interested in more than just the predictive accuracy of a model in cases where the goal is description or causal inference. Second, machine-learning models have less user input compared to traditional statistical models.[9] This leads to the increased flexibility described in the previous paragraph, but it also has the corresponding cost of reduced interpretability. Third, machine learning and deep learning have been extremely successful in developing prediction models in specific cases like image recognition and natural language processing.[10] In these problems, there are well-defined phenotypes, large numbers of predictors per observations and many observations to train the algorithm. The large number of predictors and observations means that very complicated models can be fit to predict the outcome. Traditional statistical models could be used in these settings, but it would be very difficult to specify a sufficiently complicated model. In other problems with a small number of predictors and a small data set, the goal is often to understand the importance of a single predictor. In this case, machine-learning and deep learning approaches might not be the best choice.

Machine learning and statistics have similarities, and many important scientific problems can be addressed using either approach. Although there is overlap between statistics and machine learning, machine learning and deep learning are more than just powerful statistics because they are designed to address specific questions in medical research and can efficiently find solutions to problems that would be challenging, if not impossible, using traditional statistics. The key for a data analyst is to identify the type of problem that is being confronted (description, causal inference and prediction), the main goal of the analysis and the best analytic approach, given the type of problem and goal. Both statistics and machine learning will be necessary to solve many issues in MS.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

References

1. Statistics noun, plural in form but singular or plural in construction, https://www.merriam-webster.com/dictionary/statistics (accessed 11 November 2020).
2. Murphy KP. *Machine learning: A probabilistic perspective (Adaptive computation and machine learning series)*. Cambridge, MA: MIT Press, 2012.
3. Goodfellow I, Bengio Y and Courville A. *Deep learning: Adaptive computation and machine learning*. Cambridge, MA: MIT Press, 2016.
4. Hernan M, Hsu J and Healy B. A second chance to get causal inference right: A classification of data science tasks. *Chance* 2019; 32(1): 42–49.
5. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011; 46(3): 399–424.
6. James G, Witten D, Hastie T, et al. *Springer texts in statistics: An introduction to statistical learning: With applications in R*. New York: Springer, 2013.
7. Zhao Y, Healy BC, Rotstein D, et al. Exploration of machine learning techniques in predicting multiple sclerosis disease course. *PLoS ONE* 2017; 12(4): e0174866.
8. Bzdok D, Altman N and Krzywinski M. Statistics versus machine learning. *Nat Methods* 2018; 15(4): 233–234.
9. Beam AL and Kohane IS. Big data and machine learning in health care. *JAMA* 2018; 319(13): 1317–1318.
10. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med* 2019; 25(1): 24–29.

Visit SAGE journals online journals.sagepub.com/home/msj

SAGE journals

Multiple Sclerosis Journal

2021, Vol. 27(5) 664–666

DOI: 10.1177/1352458521996699

Machine and deep learning in multiple sclerosis research are just powerful statistics – Commentary

Massimiliano Copetti and Dawei Liu