Preface

The amount of data stored in the world's databases doubles every 20 months, as estimated by Usama Fayyad, one of the founders of machine learning and coauthor of the book "Advances in knowledge discovery and data mining" (ed. by the American Association for Artificial Intelligence, Menlo Park, CA, USA, 1996), and clinicians, familiar with traditional statistical methods, are at a loss to analyze them.

Traditional methods have, indeed, difficulty to identify outliers in large datasets, and to find patterns in big data and data with multiple exposure/outcome variables. In addition, analysis rules for surveys and questionnaires, which are currently common methods of data collection, are, essentially, missing. Fortunately, the new discipline, machine learning, is able to cover all of these limitations.

In the past three years, we have completed three textbooks entitled "Machine Learning in Medicine Part One, Two, and Three" (ed. by Springer, Heidelberg, Germany, 2013). Although the textbooks were well received, it came to our attention that jaded physicians and students often lacked time to read the entire books, and requested a small book on the most important machine learning methods, without background information and theoretical discussions, and highlighting technical details.

For this reason, we have produced a small cookbook of around 100 pages containing similar information as that of the textbooks, but in a condensed form. The chapters do not have "summary, introduction, discussion, and reference" sections. Only the "example and results" sections have been maintained. Physicians and students wishing more information are referred to the textbooks.

So far medical professionals have been rather reluctant to use machine learning.

Ravinda Khattree, coauthor of the book "Computational methods in biomedical research" (ed. by Chapman & Hall, Baton Rouge, LA, USA, 2007), suggests that there may be historical reasons: technological (doctors are better than computers (?)), legal, and cultural (doctors are better trusted). Also, in the field of diagnosis making, few doctors may want a computer checking them, are interested in collaboration with a computer, collaborate with computer engineers.

In the current book, we will demonstrate that machine learning performs sometimes better than traditional statistics does. For example, if the data perfectly fit the cut-offs for node splitting, because, e.g., age > 55 years gives an exponential rise in infarctions, then decision trees, optimal binning, and optimal scaling will be

better analysis methods than traditional regression methods with age as continuous predictor. Machine learning may have little options for adjusting confounding and interaction, but you can add propensity scores and interaction variables to almost any machine learning method.

Twenty machine leaning methods relevant to medicine are described. Each chapter starts with purposes and scientific questions. Then, step-by-step analyses, using mostly simulated data examples, are given. In order for readers to perform their own analyses, the data examples are available at extras.springer.com. Finally, a paragraph with conclusion, and reference to the corresponding sites of the three textbooks written by the same authors, is given. We should emphasize that all of the methods described have been successfully applied in the authors' own research.

Lyon, November 2013

Ton J. Cleophas
Aeilko H. Zwinderman

http://www.springer.com/978-3-319-04180-3

Machine Learning in Medicine - Cookbook

Cleophas, T.J.; Zwinderman, A.H.

2014, XI, 137 p. 14 illus., Softcover
ISBN: 978-3-319-04180-3