

Preface

The amount of medical data is estimated to double every 20 months, and clinicians are at a loss to analyze them. Fortunately, user friendly statistical software has been helpful for the past 30 years. However, traditional statistical methods have 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 medical data collection, are, essentially, missing. Fortunately, a new discipline, machine learning, is able to cover all of these limitations. It involves computationally intensive methods like factor analysis, cluster analysis, and discriminant analysis. It is currently mainly the domain of computer scientists, and is already commonly used in social sciences, marketing research, operational research, and applied sciences. It is little used in medical research, probably due to the traditional belief of clinicians in clinical trials where multiple variables even out by the randomization process, and are not taken into account. In contrast, modern medical computer files often involve hundreds of variables like genes and other laboratory values, and computationally intensive methods are required.

In the past 2 years we have completed a series of three textbooks entitled *Machine Learning in Medicine Part One, Two, and Three* (ed. by Springer Heidelberg Germany, 2012–2013). Also, we produced two-100 page cookbooks, entitled *Machine Learning in Medicine—Cookbook One and Two*. These cookbooks were

- (1) without background information and theoretical discussions,
- (2) highlighting technical details,
- (3) with data examples available at extras.springer.com for readers to perform their own analyses,
- (4) with references to the above textbooks for those wishing background information.

The current volume, entitled *Machine Learning in Medicine—Cookbook Three* was written in a way much similar to that of the first two, and it reviews concised versions of machine learning methods so far, like spectral plots, Bayesian networks,

support vector machines (Chaps. 9, 12, 13). Also, a first description is given of several new methods already employed by technical and market scientists, and of their suitabilities for clinical research, like ordinal scalings for inconsistent intervals, loglinear models for varying incident risks, iteration methods for cross-validations (Chaps. 4–6, 16).

Additional new subjects are the following. Chapter 1 describes a novel method for data mining using visualization processes instead of calculus methods. Chapter 2 describes the use of trained clusters, a scientifically more appropriate alternative for traditional cluster analysis. Chapter 11 describes evolutionary operations (evops), and the evop calculators, already widely used in chemical and technical process improvements.

Similar to the first cookbook, the current work will assess in a nonmathematical way the stepwise analyses of 20 machine learning methods, that are, likewise, based on three major machine learning methodologies:

Cluster Methodologies (Chaps. 1, 2),
Linear Methodologies (Chaps. 3–8),
Rules Methodologies (Chaps. 9–20).

In extras.springer.com the data files of the examples (14 SPSS files) are given (both real and hypothesized data). Furthermore, 4 csv type excel files are available for data analysis in the Konstanz Information Miner, a widely approved free machine learning software package on the Internet since 2006.

The current 100-page book entitled *Machine Learning in Medicine—Cookbook Three*, and its complementary “Cookbooks One and Two” are written as training companions for 60 important machine learning methods relevant to medicine. We should emphasize that all of the methods described have been successfully applied in the authors’ own research.

Lyon, August 2014

Ton J. Cleophas
Aeilko H. Zwinderman

<http://www.springer.com/978-3-319-12162-8>

Machine Learning in Medicine - Cookbook Three

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

2014, XIII, 131 p. 37 illus., Softcover

ISBN: 978-3-319-12162-8