

Preface

Preface to the first edition

The analysis of experimental data is at heart of science from its beginnings. But it was the advent of digital computers in the second half of the 20th century that revolutionized scientific data analysis twofold: Tedious pencil and paper work could be successively transferred to the emerging software applications so sweat and tears turned into automated routines. In accordance with automation the manageable data volumes could be dramatically increased due to the exponential growth of computational memory and speed. Moreover highly non-linear and complex data analysis problems came within reach that were completely unfeasible before. Non-linear curve fitting, clustering and machine learning belong to these modern techniques that entered the agenda and considerably widened the range of scientific data analysis applications. Last but not least they are a further step towards computational intelligence.

The goal of this book is to provide an interactive and illustrative guide to these topics. It concentrates on the road from two-dimensional curve fitting to multidimensional clustering and machine learning with neural networks or support vector machines. Along the way topics like mathematical optimization or evolutionary algorithms are touched. All concepts and ideas are outlined in a clear cut manner with graphically depicted plausibility arguments and a little elementary mathematics. Difficult mathematical and algorithmic details are consequently banned for the sake of simplicity but are accessible by the referred literature. The major topics are extensively outlined with exploratory examples and applications. The primary goal is to be as illustrative as possible without hiding problems and pitfalls but to address them. The character of an illustrative cookbook is complemented with specific sections that address more fundamental questions like the relation between machine learning and human intelligence. These sections may be skipped without affecting the main road but they will open up possibly interesting insights beyond the mere data massage.

All topics are completely demonstrated with the aid of the computing platform Mathematica and the Computational Intelligence Packages (CIP), a high-level function library developed with Mathematica's programming language on top of Mathematica's algorithms. CIP is open-source so the detailed code of every method is freely accessible. All examples and applications shown throughout the book may be used and customized by the reader without any restrictions. This leads to an interactive environment which allows individual manipulations like the rotation of 3D graphics or the evaluation of different settings up to tailored enhancements for specific functionality.

The book tries to be as introductory as possible calling only for a basic mathematical background of the reader - a level that is typically taught in the first year of scientific education. The target readerships are students of (computer) science and engineering as well as scientific practitioners in industry and academia who deserve an illustrative introduction to these topics. Readers with programming skills may easily port and customize the provided code. The majority of the examples and applications originate from teaching efforts or solution providing. The outline of the book is as follows:

- The introductory **chapter 1** provides necessary basics that underlie the discussions of the following chapters like an initial motivation for the interplay of data and models with respect to the molecular sciences, mathematical optimization methods or data structures. The chapter may be skipped at first sight but should be consulted if things become unclear in a subsequent chapter.
- The main chapters that describe the road from curve fitting to machine learning are chapters 2 to 4. The curve fitting **chapter 2** outlines the various aspects of adjusting linear and non-linear model functions to experimental data. A section about mere data smoothing with cubic splines complements the fitting discussions.
- The clustering **chapter 3** sketches the problems of assigning data to different groups in an unsupervised manner with clustering methods. Unsupervised clustering may be viewed as a logical first step towards supervised machine learning - and may be able to construct predictive systems on its own. Machine learning methods may also need clustered data to produce successful results.
- The machine learning **chapter 4** comprises supervised learning techniques, in particular multiple linear regression, three-layer feed-forward neural networks and support vector machines. Adequate data preprocessing and their use for regression and classification tasks as well as the recurring pitfalls and problems are introduced and thoroughly discussed.
- The discussions **chapter 5** supplements the topics of the main road. It collects some open issues neglected in the previous chapters and opens up the scope with more general sections about the possible discovery of new knowledge or the emergence of computational intelligence.

The scientific fields touched in the present book are extensive and in addition constantly and progressively refined. Therefore it is inevitable to neglect an awful lot of important topics and aspects. The concrete selection always mirrors an author's

preferences as well as his personal knowledge and overview. Since the missing parts unfortunately exceed the selected ones and people always have strong feelings about what is of importance the final statement has to be a request for indulgence.

Recklinghausen, April 2011

Achim Zielesny

Preface to the second edition

The first edition was friendly reviewed as a useful introductory cookbook for the novice reader. The second edition tries to keep this character and resists the temptation to heavily expand topics or lift the discussion to more subtle academic levels. Besides numerous minor additions and corrections throughout the whole book (together with the unavoidable introduction of some new errors) the only substantial extension of the second edition is the addition of Multiple Polynomial Regression (MPR) in order to support the discussions concerning the method crossover from linear and near-linear up to highly non-linear machine learning approaches. As a consequence several examples and applications have been reworked to improve readability and line of reasoning. Also the construction of minimal predictive models is outlined in an updated and more comprehensible manner.

The second edition is based on the extended version 2.0 of the Computational Intelligence Packages (CIP) which now allows parallelized calculations that lead to an often considerably improved performance with multiple (or multicore) processors. Specific parallelization notes are given throughout the book, the description of CIP is accordingly extended and reworked examples and applications make now use of the new functionality.

With this second edition the book hopefully strengthens its original intent to provide a clear and straight introduction to the fascinating road from curve fitting to machine learning.

Recklinghausen, February 2016

Achim Zielesny

<http://www.springer.com/978-3-319-32544-6>

From Curve Fitting to Machine Learning
An Illustrative Guide to Scientific Data Analysis and
Computational Intelligence

Zielesny, A.

2016, XV, 498 p. 343 illus., 200 illus. in color.,

Hardcover

ISBN: 978-3-319-32544-6