

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation: Data, models and molecular sciences	2
1.2	Optimization	6
1.2.1	Calculus	10
1.2.2	Iterative optimization	14
1.2.3	Iterative local optimization	16
1.2.4	Iterative global optimization	20
1.2.5	Constrained iterative optimization	31
1.3	Model functions	37
1.3.1	Linear model functions with one argument	38
1.3.2	Non-linear model functions with one argument	40
1.3.3	Linear model functions with multiple arguments	41
1.3.4	Non-linear model functions with multiple arguments	43
1.3.5	Multiple model functions	44
1.3.6	Summary	44
1.4	Data structures	45
1.4.1	Data for curve fitting	45
1.4.2	Data for machine learning	45
1.4.3	Inputs for clustering	47
1.4.4	Inspection, cleaning and splitting of data	48
1.5	Scaling of data	54
1.6	Data errors	55
1.7	Regression versus classification tasks	56
1.8	The structure of CIP calculations	58
1.9	A note on reproducibility	59
<b>2</b>	<b>Curve Fitting</b>	<b>61</b>
2.1	Basics	65
2.1.1	Fitting data	65
2.1.2	Useful quantities	66
2.1.3	Smoothing data	68

2.2	Evaluating the goodness of fit	70
2.3	How to guess a model function	76
2.4	Problems and pitfalls	88
2.4.1	Parameters' start values	89
2.4.2	How to search for parameters' start values	93
2.4.3	More difficult curve fitting problems	97
2.4.4	Inappropriate model functions	107
2.5	Parameters' errors	112
2.5.1	Correction of parameters' errors	112
2.5.2	Confidence levels of parameters' errors	113
2.5.3	Estimating the necessary number of data	114
2.5.4	Large parameters' errors and educated cheating	118
2.5.5	Experimental errors and data transformation	132
2.6	Empirical enhancement of theoretical model functions	135
2.7	Data smoothing with cubic splines	143
2.8	Cookbook recipes for curve fitting	154
<b>3</b>	<b>Clustering</b>	<b>157</b>
3.1	Basics	160
3.2	Intuitive clustering	163
3.3	Clustering with a fixed number of clusters	178
3.4	Getting representatives	185
3.5	Cluster occupancies and the iris flower example	194
3.6	White-spot analysis	206
3.7	Alternative clustering with ART-2a	209
3.8	Clustering and class predictions	220
3.9	Cookbook recipes for clustering	228
<b>4</b>	<b>Machine Learning</b>	<b>229</b>
4.1	Basics	237
4.2	Machine learning methods	242
4.2.1	Multiple linear and polynomial regression (MLR, MPR)	243
4.2.2	Three-layer feed-forward neural networks	246
4.2.3	Support vector machines (SVM)	251
4.3	Evaluating the goodness of regression	256
4.4	Evaluating the goodness of classification	260
4.5	Regression: Entering non-linearity	264
4.6	Classification: Non-linear decision surfaces	282
4.7	Ambiguous classification	286
4.8	Training and test set partitioning	298
4.8.1	Cluster representatives based selection	299
4.8.2	Iris flower classification revisited	304
4.8.3	Adhesive kinetics regression revisited	316
4.8.4	Design of experiment	320
4.8.5	Concluding remarks	335

4.9	Comparative machine learning	336
4.10	Relevance of input components and minimal models	349
4.11	Pattern recognition	355
4.12	Technical optimization problems	373
4.13	Cookbook recipes for machine learning	378
4.14	Appendix - Collecting the pieces	380
<b>5</b>	<b>Discussion</b>	<b>407</b>
5.1	Computers are about speed	407
5.2	Isn't it just ...?	417
5.2.1	... optimization?	418
5.2.2	... data smoothing?	418
5.3	Computational intelligence	429
5.4	Final remark	434
<b>A</b>	<b>CIP -Computational Intelligence Packages</b>	<b>437</b>
A.1	Basics	437
A.2	Experimental data	439
A.2.1	Temperature dependence of the viscosity of water	439
A.2.2	Potential energy surface of hydrogen fluoride	440
A.2.3	Kinetics data from time dependent IR spectra of the hydrolysis of acetanhydride	441
A.2.4	Iris flowers	448
A.2.5	Adhesive kinetics	449
A.2.6	Intertwined spirals	451
A.2.7	Faces	452
A.2.8	Wisconsin Diagnostic Breast Cancer (WDBC) data	455
A.2.9	Wisconsin Prognostic Breast Cancer (WPBC) data	456
A.2.10	QSPR data	457
A.3	Parallelized calculations	458
	<b>References</b>	<b>461</b>
	<b>Index</b>	<b>467</b>

<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