

Contents

1	Introduction	1
1.1	Motivation: Data, Models and Molecular Sciences	2
1.2	Optimization	6
1.2.1	Calculus	9
1.2.2	Iterative Optimization	13
1.2.3	Iterative Local Optimization	15
1.2.4	Iterative Global Optimization	19
1.2.5	Constrained Iterative Optimization	30
1.3	Model Functions	36
1.3.1	Linear Model Functions with One Argument	37
1.3.2	Non-linear Model Functions with One Argument	39
1.3.3	Linear Model Functions with Multiple Arguments	40
1.3.4	Non-linear Model Functions with Multiple Arguments	42
1.3.5	Multiple Model Functions	43
1.3.6	Summary	43
1.4	Data Structures	44
1.4.1	Data for Curve Fitting	44
1.4.2	Data for Machine Learning	44
1.4.3	Inputs for Clustering	46
1.4.4	Inspection of Data Sets and Inputs	46
1.5	Scaling of Data	47
1.6	Data Errors	47
1.7	Regression versus Classification Tasks	49
1.8	The Structure of CIP Calculations	51

2	Curve Fitting	53
2.1	Basics	57
2.1.1	Fitting Data	57
2.1.2	Useful Quantities	58
2.1.3	Smoothing Data	60
2.2	Evaluating the Goodness of Fit	62
2.3	How to Guess a Model Function	68
2.4	Problems and Pitfalls	80
2.4.1	Parameters' Start Values	81
2.4.2	How to Search for Parameters' Start Values	85
2.4.3	More Difficult Curve Fitting Problems	89
2.4.4	Inappropriate Model Functions	99
2.5	Parameters' Errors	104
2.5.1	Correction of Parameters' Errors	104
2.5.2	Confidence Levels of Parameters' Errors	105
2.5.3	Estimating the Necessary Number of Data	106
2.5.4	Large Parameters' Errors and Educated Cheating	110
2.5.5	Experimental Errors and Data Transformation	124
2.6	Empirical Enhancement of Theoretical Model Functions	127
2.7	Data Smoothing with Cubic Splines	135
2.8	Cookbook Recipes for Curve Fitting	146
3	Clustering	149
3.1	Basics	152
3.2	Intuitive Clustering	155
3.3	Clustering with a Fixed Number of Clusters	170
3.4	Getting Representatives	177
3.5	Cluster Occupancies and the Iris Flower Example	186
3.6	White-Spot Analysis	198
3.7	Alternative Clustering with ART-2a	201
3.8	Clustering and Class Predictions	212
3.9	Cookbook Recipes for Clustering	220
4	Machine Learning	221
4.1	Basics	228
4.2	Machine Learning Methods	234
4.2.1	Multiple Linear Regression (MLR)	234
4.2.2	Three-Layer Perceptron-Type Neural Networks	236
4.2.3	Support Vector Machines (SVM)	241
4.3	Evaluating the Goodness of Regression	245
4.4	Evaluating the Goodness of Classification	250
4.5	Regression: Entering Non-linearity	253
4.6	Classification: Non-linear Decision Surfaces	263
4.7	Ambiguous Classification	267

4.8	Training and Test Set Partitioning	278
4.8.1	Cluster Representatives Based Selection	280
4.8.2	Iris Flower Classification Revisited	285
4.8.3	Adhesive Kinetics Regression Revisited	296
4.8.4	Design of Experiment	304
4.8.5	Concluding Remarks	320
4.9	Comparative Machine Learning	320
4.10	Relevance of Input Components	332
4.11	Pattern Recognition	339
4.12	Technical Optimization Problems.....	356
4.13	Cookbook Recipes for Machine Learning	360
4.14	Appendix - Collecting the Pieces	362
5	Discussion	381
5.1	Computers Are about Speed	381
5.2	Isn't It Just ...?	391
5.2.1	... Optimization?	392
5.2.2	... Data Smoothing?	392
5.3	Computational Intelligence	403
5.4	Final Remark	408
A	CIP - Computational Intelligence Packages	409
A.1	Basics	409
A.2	Experimental Data	411
A.2.1	Temperature Dependence of the Viscosity of Water	411
A.2.2	Potential Energy Surface of Hydrogen Fluoride ...	412
A.2.3	Kinetics Data from Time Dependent IR Spectra of the Hydrolysis of Acetanhydride.....	413
A.2.4	Iris Flowers	420
A.2.5	Adhesive Kinetics	420
A.2.6	Intertwined Spirals	422
A.2.7	Faces	423
A.2.8	Wisconsin Diagnostic Breast Cancer (WDBC) Data	426
	Index	433