

# Contents

<i>Preface</i>	page xi
<i>List of abbreviations</i>	xiii
<b>1 Determination of the accurate location of an aircraft</b>	<b>1</b>
1.1 Introduction	1
1.2 Modelling the problem as a least squares problem	4
1.2.1 How to solve a mixed problem: equation system and minimization	6
1.3 Solving the non-linear least squares problem	6
1.4 Error analysis/confidence analysis	8
1.4.1 An approximation for a complicated $S$	10
1.4.2 Interactive exercise	12
Further reading	13
<b>2 When to replace equipment</b>	<b>14</b>
2.1 The replacement paradox	14
2.2 Car replacement	15
2.2.1 Simulation	18
2.2.2 Linear approximations	18
2.2.3 Mathematical background	20
2.3 Computer replacement and Moore's law	22
2.3.1 Piecewise linear approximation	23
<b>3 Secondary structure prediction using least squares and singular value decomposition</b>	<b>25</b>
3.1 Extremely brief introduction to protein biochemistry	25
3.1.1 The structure of myoglobin ( <i>Physeter catodon</i> ) – P02185	27
3.1.2 Additional information about the biochemical structure of $\alpha$ -helices	28
3.2 General introduction to modelling and prediction	30
3.2.1 SSP as modelling problem	32
3.3 Least squares – singular value decomposition	33
3.3.1 The need for weighting	35
3.3.2 Singular value decomposition and eigenvalue decomposition	36

3.3.3	Criteria for discarding singular values	40
3.3.4	A concrete example	43
<b>4</b>	<b>Secondary structure prediction using least squares and best basis</b>	<b>49</b>
4.1	Definition of “best basis”	49
4.1.1	Stepwise regression	49
4.1.2	The difficulty of the problem	50
4.2	Mathematical formulation using the best basis approach	51
4.3	Algorithms for finding local minima	52
4.3.1	Early abort	52
4.3.2	(Discrete) steepest descent	53
4.3.3	Improved EA or SD algorithm (avoiding repeated paths)	54
4.3.4	Exploring the solution space	54
4.4	The neighbor function	55
4.5	The confidence interval	56
4.5.1	Analysis of the variance of the coefficients	57
4.6	Confidence analysis with random columns	58
4.7	Final observations on the least squares methods	59
4.7.1	Refining the predictor function	59
4.7.2	Validation	60
4.7.3	Other models	61
4.8	Linear classification/discrimination	62
4.8.1	Trying to do better	63
4.8.2	Constructing initial solutions using the sigmoid function	64
<b>5</b>	<b>Secondary structure prediction with learning methods (nearest neighbors)</b>	<b>66</b>
5.1	Formulation of the problem	67
5.1.1	The general idea	67
5.1.2	Application to helix prediction	68
5.2	Searching and data structures for NN problems	69
5.2.1	Sequential search	69
5.2.2	Range searching	69
5.2.3	Quad trees	70
5.2.4	$k$ -d Trees ( $k$ -dimensional trees)	70
5.2.5	Nearest neighbor trees (NN trees)	71
5.3	Building NN trees	74
5.4	Searching the NN tree	75
5.4.1	Searching the $k$ nearest neighbors	76
5.4.2	Concluding a value from its neighbors	77
5.5	Practical considerations	79
5.5.1	Bucketing	79

5.5.2	Normalizing dimensions	79
5.5.3	Using random directions	79
5.6	NN trees used for clustering	80
<b>6</b>	<b>Secondary structure prediction with linear programming (LP)</b>	<b>82</b>
6.1	Introduction	82
6.1.1	An introductory example	83
6.2	The simplex algorithm	85
6.2.1	Slack variables	86
6.3	SSP formulated as linear programming	89
6.4	Inconsistent data	90
6.4.1	Using slack variables for each inconsistency	91
6.4.2	Alternative method – removing inconsistencies	91
6.5	Prediction	92
6.6	Comparison of linear programming with least squares	92
6.7	Example	93
	Further reading	94
<b>7</b>	<b>Stock market prediction</b>	<b>96</b>
7.1	Introduction	96
7.1.1	Definitions	97
7.1.2	Examples of information available online	98
7.1.3	Rationale for stock market prediction	99
7.1.4	Organization of the modelling	100
7.2	The optimal transaction sequence (OTS)	101
7.2.1	Defining the problem	101
7.2.2	Calculating the OTS for the past	101
7.2.3	Interactive exercise “Optimal transaction sequence”	102
7.3	Approximating the OTS	103
7.3.1	Decision versus numerical approach	104
7.3.2	Computing good (optimal) parameters	109
7.3.3	From the OTS to reality	111
7.4	Simulation	111
7.4.1	Two simple models	112
7.4.2	Optimization of simulation parameters	114
7.5	Confidence analysis	124
7.5.1	The notion of “transfer of error”	125
7.5.2	Monte Carlo estimation	125
7.6	Dynamic programming	127
7.6.1	The idea of dynamic programming in more detail	128
7.7	Examples of dynamic programming	129

7.7.1	Optimal binary search trees	129
7.7.2	Optimal order of matrix multiplication	132
7.7.3	String matching	135
	Further reading	138
<b>8</b>	<b>Phylogenetic tree construction</b>	<b>139</b>
8.1	Introduction	139
8.1.1	Applications	139
8.1.2	What it is about	142
8.1.3	Classification types	143
8.1.4	Phylogenies from protein sequences	144
8.2	Tree building	145
8.2.1	Rooted versus unrooted trees	146
8.2.2	The constructor–heuristic–evaluator (CHE) optimization strategy	148
8.3	Trees based on distance information	149
8.3.1	Measures of distance	149
8.3.2	PAM distance in detail	153
8.3.3	Calculating distances from sequence alignment	157
8.3.4	Dayhoff matrices	158
8.3.5	Sequence alignment using Dayhoff matrices and maximum likelihood	160
8.3.6	Estimating distances between sequences by maximum likelihood	162
8.3.7	Sequence alignment – global alignment	164
8.3.8	Sequence alignment – local alignment	164
8.3.9	Sequence alignment – cost-free end alignment	164
8.3.10	Distance and variance matrices	165
8.4	Creating an initial topology	166
8.4.1	The UPGMA algorithm	166
8.4.2	The WPGMA algorithm	168
8.4.3	Tree fitting by least squares	171
8.4.4	Improving a tree: swapping heuristics	173
8.5	Phylogenetic trees based on character information	175
8.5.1	Parsimony	176
8.5.2	General parsimony	177
8.6	Special cases allowing easy tree construction	181
8.6.1	Strict character compatibility	181
8.6.2	Ultrametric trees	184
8.6.3	Additive trees	185
8.7	Evolutionary algorithms or genetic algorithms	185
8.8	Randomness, parallelism and CHE procedures	190
8.8.1	Concluding additional information from different runs	190
8.8.2	Sensitivity analysis suitable for parallel CHE	191
8.8.3	Tradeoff between randomness and quality	194
	Further reading	195

<i>Appendix A</i> Methods for function minimization	196
<i>Appendix B</i> Online resources	228
<i>Index</i>	231