

Contents

Preface	ix
Acknowledgement	xi
PART 1 INTRODUCTION	1
1 Introduction to the Basics of Molecular Biology	3
1.1 Basic cell architecture	3
1.2 The structure, content and scale of deoxyribonucleic acid (DNA)	4
1.3 History of the human genome	9
1.4 Genes and proteins	10
1.5 Current knowledge and the 'central dogma'	21
1.6 Why proteins are important	23
1.7 Gene and cell regulation	24
1.8 When cell regulation goes wrong	26
1.9 So, what is bioinformatics?	27
1.10 Summary of chapter	28
1.11 Further reading	29
2 Introduction to Problems and Challenges in Bioinformatics	31
2.1 Introduction	31
2.2 Genome	31
2.3 Transcriptome	40
2.4 Proteome	50
2.5 Interference technology, viruses and the immune system	57
2.6 Summary of chapter	63
2.7 Further reading	64

3	Introduction to Artificial Intelligence and Computer Science	65
3.1	Introduction to search	65
3.2	Search algorithms	66
3.3	Heuristic search methods	72
3.4	Optimal search strategies	76
3.5	Problems with search techniques	83
3.6	Complexity of search	84
3.7	Use of graphs in bioinformatics	86
3.8	Grammars, languages and automata	90
3.9	Classes of problems	96
3.10	Summary of chapter	98
3.11	Further reading	99
PART 2	CURRENT TECHNIQUES	101
4	Probabilistic Approaches	103
4.1	Introduction to probability	103
4.2	Bayes' Theorem	105
4.3	Bayesian networks	111
4.4	Markov networks	116
4.5	Summary of chapter	125
4.6	References	126
5	Nearest Neighbour and Clustering Approaches	127
5.1	Introduction	127
5.2	Nearest neighbour method	130
5.3	Nearest neighbour approach for secondary structure protein folding prediction	132
5.4	Clustering	135
5.5	Advanced clustering techniques	138
5.6	Application guidelines	144
5.7	Summary of chapter	145
5.8	References	146
6	Identification (Decision) Trees	147
6.1	Method	147
6.2	Gain criterion	152
6.3	Over fitting and pruning	157
6.4	Application guidelines	160
6.5	Bioinformatics applications	163
6.6	Background	169

6.7	Summary of chapter	170
6.8	References	170
7	Neural Networks	173
7.1	Method	173
7.2	Application guidelines	185
7.3	Bioinformatics applications	187
7.4	Background	192
7.5	Summary of chapter	193
7.6	References	193
8	Genetic Algorithms	195
8.1	Single-objective genetic algorithms – method	195
8.2	Single-objective genetic algorithms – example	202
8.3	Multi-objective genetic algorithms – method	205
8.4	Application guidelines	207
8.5	Genetic algorithms – bioinformatics applications	210
8.6	Summary of chapter	217
8.7	References and further reading	217
PART 3	FUTURE TECHNIQUES	219
9	Genetic Programming	221
9.1	Method	221
9.2	Application guidelines	230
9.3	Bioinformatics applications	232
9.4	Background	236
9.5	Summary of chapter	236
9.6	References	237
10	Cellular Automata	239
10.1	Method	239
10.2	Application guidelines	245
10.3	Bioinformatics applications	247
10.4	Background	251
10.5	Summary of chapter	252
10.6	References and further reading	252
11	Hybrid Methods	255
11.1	Method	255
11.2	Neural-genetic algorithm for analysing gene expression data	256
11.3	Genetic algorithm and k nearest neighbour hybrid for biochemistry solvation	262

11.4	Genetic programming neural networks for determining gene – gene interactions in epidemiology	265
11.5	Application guidelines	268
11.6	Conclusions	268
11.7	Summary of chapter	269
11.8	References and further reading	269
Index		271
PART 2 CURRENT TECHNIQUES		
2.1	Introduction	101
2.2	Genetic algorithms – bioinformatics applications	101
2.3	Summary of chapter	103
2.4	References and further reading	103
2.5	Bayesian networks	111
2.6	Markov networks	111
2.7	Summary of chapter	111
2.8	References	111
PART 3 FUTURE TECHNIQUES		
3.1	Genetic Programming	121
3.2	Application guidelines	127
3.3	Background	127
3.4	Summary of chapter	130
3.5	References	131
3.6	Clustering	131
3.7	Advanced clustering techniques	131
3.8	Application guidelines	141
3.9	Summary of chapter	141
3.10	References	141
3.11	Identification (Decision) Trees	141
3.12	Method	141
3.13	Gain criterion	141
3.14	Summary of chapter	141
3.15	References	141
3.16	Hybrid Methods	141
3.17	Method	141
3.18	Neural-genetic algorithm for analysing gene expression data	141
3.19	Genetic algorithm and a nearest neighbour hybrid for biochemistry solution	141