

Contents

About the authors	xiii
Preface	xv
Acknowledgments	xix

PART 1 Artificial intelligence: methods

CHAPTER 1 Knowledge representation and reasoning	3
1.1 Introduction	3
1.2 Knowledge representation	4
1.3 Reasoning	6
1.4 Computer science and knowledge representation and reasoning	6
1.5 Artificial intelligence and knowledge representation and reasoning	8
1.6 Languages for knowledge representation and reasoning	9
1.7 Artificial intelligence and bioinformatics	9
CHAPTER 2 Machine learning	11
2.1 Introduction	12
2.2 Classification	14
2.2.1 Supervised machine learning algorithms	15
2.2.2 Support vector machines	16
2.3 Clustering	17
2.3.1 K-means clustering algorithm	19
2.4 Association learning	21
2.4.1 Association learning algorithms	23
2.5 Reinforcement learning	27
CHAPTER 3 Artificial intelligence	29
3.1 A brief history of artificial intelligence	29
3.2 Artificial intelligence and bioinformatics	31
3.3 Artificial intelligence in medicine: a short tale	32
CHAPTER 4 Data science	35
4.1 Introduction	36
4.2 A quick primer on data	39
4.3 The data science process	41
4.3.1 Problem and data understanding	41
4.3.2 Data processing and feature engineering	43
4.3.3 Modeling	44
4.3.4 Evaluation	45
4.3.5 Reporting and finalize the model	47

4.4	Languages for data science	48
4.4.1	MATLAB®	48
4.4.2	Julia	49
4.4.3	R	49
4.4.4	Python	49
4.5	Low and no coding tools for data science	50
4.5.1	WEKA	50
4.5.2	KNIME	50
4.5.3	PyCaret	51
CHAPTER 5	Deep learning	53
5.1	Introduction	53
5.2	Introducing basic principles behind deep learning	55
5.2.1	Artificial neuron models	55
5.2.2	Representation learning vs feature engineering	57
5.3	Popular deep neural networks architecture	59
5.3.1	Convolutional neural networks	60
CHAPTER 6	Explainability of AI methods	61
6.1	Introduction	61
6.2	Explainable models in machine learning	62
6.2.1	External explainer models	64
6.2.2	Self-explaining models	65
6.3	Application of explainable AI in medicine	66
6.3.1	Explainable AI for text mining applications	66
CHAPTER 7	Intelligent agents	69
7.1	Introduction	69
7.2	Types of intelligent agents	73
7.3	Agent-oriented programming frameworks	79

PART 2 Artificial intelligence: bioinformatics

CHAPTER 8	Sequence analysis	83
8.1	Introduction	84
8.2	String similarity methods	85
8.2.1	Edit distance	86
8.2.2	Recursive formulation	87
8.2.3	Calculating sequences similarity scores	89
8.3	Dynamic programming algorithm for edit distances	89
8.3.1	Needleman and Wunsch algorithm	91
8.4	Multi-parameterized edit distances	92
8.4.1	Background	93
8.4.2	An algorithm for computing the MPED	94
8.5	Alignment free sequence comparison	95

CHAPTER 9	Structure analysis	97
9.1	Introduction	97
9.2	Protein secondary structure prediction	98
9.2.1	Artificial neural networks for structure prediction	99
9.3	Tertiary structure prediction	101
9.3.1	Databases of structural classifications	101
9.3.2	Algorithms for tertiary structure analysis	102
CHAPTER 10	Omics sciences	105
10.1	Introduction	106
10.2	Genomics	106
10.2.1	Next-generation sequencing	107
10.3	Transcriptomics	108
10.3.1	RNA-seq data analysis	109
10.4	Epigenomics	110
10.5	Proteomics	112
10.6	Metabolomics	112
10.7	Interactomics	113
10.8	Gene prioritization	115
10.8.1	Gene-prioritization approaches	117
CHAPTER 11	Ontologies in bioinformatics	119
11.1	Introduction	119
11.2	Biomedical ontologies	120
11.2.1	Gene ontology	121
11.2.2	Human phenotype ontology	122
11.2.3	Disease ontology	123
11.3	Semantic similarity measures	123
11.3.1	Knowledge extraction from ontologies	126
11.4	Functional enrichment analysis	127
CHAPTER 12	Integrative bioinformatics	129
12.1	Introduction	129
12.2	Data integration in bioinformatics	130
12.2.1	Deep learning and data integration in multiomics	130
12.2.2	Integration of transcriptomics data	132
12.3	Databases, tools, and languages	133
CHAPTER 13	Biological networks analysis	137
13.1	Introduction	138
13.2	Networks in biology	139
13.2.1	Random graph model	140
13.2.2	Scale-free model	141
13.2.3	Geometric random graph model	141
13.3	Motif discovery	141
13.4	Network embedding (representation learning)	142

13.5	Networks alignment	144
13.5.1	Global network alignment	146
13.5.2	Local network alignment	146
13.5.3	Pairwise and multiple alignment	147
13.5.4	Quality evaluation of network alignment	148
CHAPTER 14	Biological pathway analysis	151
14.1	Introduction	152
14.2	Biological pathways	153
14.3	Pathway databases	155
14.4	Pathway representation formats	156
14.4.1	BioPAX	157
14.4.2	CellML	158
14.4.3	HUPO	158
14.4.4	SBML	158
14.5	Pathways enrichment analysis methods	158
14.5.1	ORA	159
14.5.2	GSEA	160
14.5.3	TEA	160
14.6	Pathway enrichment analysis tools	162
14.6.1	ORA software tools	162
14.6.2	GSEA software tools	163
14.6.3	TEA software tools	165
CHAPTER 15	Knowledge extraction from biomedical texts	167
15.1	Introduction	167
15.2	A primer on text analysis	168
15.2.1	Data collection	168
15.2.2	Text pre-processing	169
15.2.3	Text cleaning	169
15.2.4	Tokenization	169
15.2.5	Filtering	169
15.2.6	Lemmatization and stemming	169
15.2.7	Text representation	170
15.3	Biomedical text mining tasks	171
15.3.1	Named entity recognition	172
15.3.2	Text summarization	172
15.3.3	Topic modeling	172
15.3.4	Sentiment analysis	173
CHAPTER 16	Artificial intelligence in bioinformatics: issues and challenges	175
16.1	Introduction	176
16.2	Evolution of bioinformatics	176
16.2.1	Origin of bioinformatics	176

16.2.2	NIH-NHGRI definition of bioinformatics	177
16.2.3	SIB definition of bioinformatics	177
16.2.4	EBI definition of bioinformatics	177
16.2.5	Other definitions of bioinformatics	178
16.2.6	Artificial intelligence bioinformatics	178
16.3	Challenges for artificial intelligence in bioinformatics	178
16.3.1	High demand of computing power	179
16.3.2	Lack of trust and explainability	179
16.3.3	Limited awareness of AI potentiality	179
16.3.4	Human vs AI precision	179
16.3.5	Data privacy and security	179
16.3.6	The bias in AI	180
16.3.7	Data scarcity	181
APPENDIX A	Python code examples	183
A.1	Classification of omics data	183
A.2	Cluster analysis of gene expression data	186
A.3	Python agent-oriented programming framework	190
A.4	Sequences similarity score calculation	192
A.5	Dynamic programming	193
A.6	Analysis of FASTQ sequences	196
A.7	Analysis of alignment map in SAM/BAM format	199
A.8	Mass spectrometer data analysis	201
APPENDIX B	Java code examples	205
B.1	Java agent-oriented programming frameworks	205
B.1.1	Java Agent DEvelopment (JADE) framework	205
B.1.2	JADEX Java Agent DEvelopment eXtension framework	209
	Bibliography	213
	Index	237