

Contents

List of figures	vii
List of tables	xi
About the author	xiii
Dedication	xv
1 An overview of data mining	1
1.1 What's data mining?	1
1.2 Data mining process models	1
1.3 Data collection	1
1.4 Data preprocessing	2
1.5 Data modeling	3
1.6 Model assessment	8
1.7 Model deployment	9
1.8 Summary	9
References	10
2 Introduction to bioinformatics	11
2.1 A primer to molecular biology	11
2.2 What is bioinformatics?	11
2.3 Data mining issues in bioinformatics	12
2.4 Challenges in biological data mining	17
2.5 Summary	17
References	17
3 Phosphorylation motif discovery	19
3.1 Background and problem description	19
3.2 The nature of the problem	20
3.3 Data collection	20
3.4 Data preprocessing	21
3.5 Modeling: A discriminative pattern mining perspective	21
3.6 Validation: Permutation p -value calculation	24
3.7 Discussion and future perspective	26
References	27
4 Phosphorylation site prediction	29
4.1 Background and problem description	29
4.2 Data collection and data preprocessing	29

4.3	Modeling: Different learning schemes	32
4.4	Validation: Cross-validation and independent test	35
4.5	Discussion and future perspective	35
	References	36
5	Protein inference in shotgun proteomics	39
5.1	Introduction to proteomics	39
5.2	Protein identification in proteomics	40
5.3	Protein inference: Problem formulation	40
5.4	Data collection	41
5.5	Modeling with different data mining techniques	41
5.6	Validation: Target-decoy versus decoy-free	44
5.7	Discussion and future perspective	48
	References	48
6	PPI network inference from AP-MS data	51
6.1	Introduction to protein–protein interactions	51
6.2	AP-MS data generation	51
6.3	Data collection and preprocessing	52
6.4	Modeling with different data mining techniques	52
6.5	Validation	56
6.6	Discussion and future perspective	57
	References	58
7	Protein complex identification from AP-MS data	61
7.1	An introduction to protein complex identification	61
7.2	Data collection and data preprocessing	61
7.3	Modeling: A graph clustering framework	61
7.4	Validation	67
7.5	Discussion and future perspective	68
	References	68
8	Biomarker discovery	69
8.1	An introduction to biomarker discovery	69
8.2	Data preprocessing	69
8.3	Modeling	70
8.4	Validation	75
8.5	Case study	76
8.6	Discussion and future perspective	77
	References	78
	Conclusions	79
	Index	81