

Contents

List of Figures	x
List of Tables	xiv
1 Introduction	1
1.1 Aims of Thesis	1
1.2 Contributions of Thesis	1
1.3 Thesis Outline	2
2 Background Theory	4
2.1 Basic Concepts	4
2.1.1 Terminology	4
2.1.2 Temporal Probabilistic Models	8
2.1.3 Gene Expression and Microarray Data	10
2.2 Learning of Bayesian Models: A Survey of the Literature . . .	14
2.2.1 On Learning the Factorized Joint Probability Distribution	15
2.2.2 Some Popular Markov Blanket Induction Algorithms .	19
2.2.3 Bayesian Learning and Gene Expression	24
2.3 Complexity and Learning	25
2.3.1 Noisy-OR Model	26
2.3.2 Naive Bayes Model	27

2.3.3	Extensions of the Naive Bayes Model	28
2.4	Philosophical Underpinnings	29
2.5	Conclusion	32
3	A Time-Based Probabilistic Model	34
3.1	Introduction	34
3.2	A Possible Model	34
3.3	Simplifying Assumptions for the Compression Function $\mathcal{F}$. .	38
3.3.1	Assumptions	38
3.3.2	Justifications for the Assumptions	38
3.4	A Temporal Model	39
3.5	Modelling Time Series Data: Microarray Data	41
3.5.1	Quantization	43
3.5.2	Time Shifts	44
3.5.3	A Compression Function	47
3.5.4	Cyclic Time Period: λ	49
3.6	On the $\mathcal{F}$ -Model and Complexity	49
3.7	On the $\mathcal{F}$ -model and Causality	50
3.8	Conclusion	51
4	Learning the $\mathcal{F}$-model from Data	53
4.1	A Learning Approach	53
4.1.1	Two Schools of Thought	53
4.1.2	A Search Strategy	55
4.2	An Parent Induction Algorithm: PIA	56
4.2.1	Local Neighborhood	56
4.2.2	Selecting Candidate Causes: Phase 1	57
4.2.3	Learning Local Neighborhoods over the Pruned Space .	61
4.2.4	A Local Learning Algorithm: PIA	63

4.3	A Score	63
4.3.1	Mutual Information	63
4.3.2	The Compression Function and a Problem	66
4.3.3	Refinement of the PIA	68
4.4	A Network Induction Algorithm	70
4.5	Summary	71
5	A Time Series Microarray Data Simulator	72
5.1	Introduction	72
5.1.1	Background	72
5.1.2	Motivation	72
5.1.3	Previous and Related Work	73
5.2	Modelling Time Series Microarray Data	74
5.2.1	Dependencies	74
5.2.2	Temporal Features	75
5.2.3	Noise	76
5.2.4	Summary	77
5.3	A Model and a Simulator for Time Series Microarray Data	77
5.3.1	A Model	78
5.3.2	An Implementation of the Model	80
5.3.3	Multiple Causes	92
5.4	Critique and Extensions	92
5.4.1	Biological Plausibility	92
5.4.2	Extensions	94
5.5	Conclusion	95
6	Experimentation	96
6.1	Introduction	96
6.1.1	Background	96

