

Contents

<i>Preface</i>	vii
<i>Notational Conventions</i>	xv
1. Why Kernels for Structured Data?	1
1.1 Supervised Machine Learning	2
1.2 Kernel Methods	4
1.3 Representing Structured Data	6
1.4 Goals and Contributions	7
1.5 Outline	8
1.6 Bibliographical Notes	9
2. Kernel Methods in a Nutshell	13
2.1 Mathematical Foundations	14
2.1.1 From Sets to Functions	14
2.1.2 Measures and Integrals	15
2.1.3 Metric Spaces	17
2.1.4 Linear Spaces and Banach Spaces	18
2.1.5 Inner Product Spaces and Hilbert Spaces	20
2.1.6 Reproducing Kernels and Positive-Definite Functions	22
2.1.7 Matrix Computations	24
2.1.8 Partitioned Inverse Equations	26
2.2 Recognising Patterns with Kernels	27
2.2.1 Supervised Learning	27
2.2.2 Empirical Risk Minimisation	29
2.2.3 Assessing Predictive Performance	30
2.3 Foundations of Kernel Methods	33

2.3.1	Model Fitting and Linear Inverse Equations . . .	33
2.3.2	Common Grounds of Kernel Methods	34
2.3.3	Representer Theorem	36
2.4	Kernel Machines	38
2.4.1	Regularised Least Squares	38
2.4.2	Support Vector Machines	41
2.4.3	Gaussian Processes	45
2.4.4	Kernel Perceptron	47
2.4.5	Kernel Principal Component Analysis	48
2.4.6	Distance-Based Algorithms	51
2.5	Summary	52
3.	Kernel Design	55
3.1	General Remarks on Kernels and Examples	56
3.1.1	Classes of Kernels	56
3.1.2	Good Kernels	57
3.1.3	Kernels on Inner Product Spaces	58
3.1.4	Some Illustrations	60
3.2	Kernel Functions	61
3.2.1	Closure Properties	62
3.2.2	Kernel Modifiers	65
3.2.3	Minimal and Maximal Functions	66
3.2.4	Soft-Maximal Kernels	67
3.3	Introduction to Kernels for Structured Data	68
3.3.1	Intersection and Crossproduct Kernels on Sets . .	68
3.3.2	Minimal and Maximal Functions on Sets	71
3.3.3	Kernels on Multisets	72
3.3.4	Convolution Kernels	73
3.4	Prior Work	73
3.4.1	Kernels from Generative Models	74
3.4.2	Kernels from Instance Space Graphs	76
3.4.3	String Kernels	79
3.4.4	Tree Kernels	82
3.5	Summary	83
4.	Basic Term Kernels	85
4.1	Logics for Learning	86
4.1.1	Propositional Logic for Learning	86

4.1.2	First-Order Logic for Learning	87
4.1.3	Lambda Calculus	88
4.1.4	Lambda Calculus with Polymorphic Types	90
4.1.5	Basic Terms for Learning	91
4.2	Kernels for Basic Terms	94
4.2.1	Default Kernels for Basic Terms	95
4.2.2	Positive Definiteness of the Default Kernel	98
4.2.3	Specifying Kernels	101
4.3	Multi-Instance Learning	103
4.3.1	The Multi-Instance Setting	104
4.3.2	Separating MI Problems	105
4.3.3	Convergence of the MI Kernel Perceptron	108
4.3.4	Alternative MI Kernels	111
4.3.5	Learning MI Ray Concepts	113
4.4	Related Work	114
4.4.1	Kernels for General Data Structures	114
4.4.2	Multi-Instance Learning	115
4.5	Applications and Experiments	116
4.5.1	East/West Challenge	116
4.5.2	Drug Activity Prediction	117
4.5.3	Structure Elucidation from Spectroscopic Analyses	120
4.5.4	Spatial Clustering	123
4.6	Summary	125
5.	Graph Kernels	127
5.1	Motivation and Approach	128
5.2	Labelled Directed Graphs	130
5.2.1	Basic Terminology and Notation	131
5.2.2	Matrix Notation and some Functions	132
5.2.3	Product Graphs	133
5.2.4	Limits of Matrix Power Series	133
5.3	Complete Graph Kernels	135
5.4	Walk Kernels	139
5.4.1	Kernels Based on Label Pairs	140
5.4.2	Kernels Based on Contiguous Label Sequences	142
5.4.3	Transition Graphs	143
5.4.4	Non-Contiguous Label Sequences	144
5.5	Cyclic Pattern Kernels	146
5.5.1	Undirected Graphs	146

5.5.2	Kernel Definition	147
5.5.3	Kernel Computation	148
5.6	Related Work	151
5.7	Relational Reinforcement Learning	154
5.7.1	Relational Reinforcement Learning	155
5.7.2	Kernels for Graphs with Parallel Edges	157
5.7.3	Kernel Based RRL in the Blocks World	158
5.7.4	Experiments	161
5.7.5	Future Work	165
5.8	Molecule Classification	166
5.8.1	Mutagenicity	166
5.8.2	HIV Data	168
5.9	Summary	172
6.	Conclusions	175
	<i>Bibliography</i>	179
	<i>Index</i>	191