

Table of Contents

Abstract	ii
Table of Contents	iii
List of Figures	xii
List of Tables	xvi
List of Algorithms	xvii
Acknowledgements	xviii
Notation	xx
Introduction	1
Machine Learning and Artificial Intelligence	2
Taxonomy of Machine Learning Algorithms	2
Structured Data	4
Kernel Machines	5
Overview of the Thesis	6
Sources of the Thesis	8
Other Projects	9
I Learning Structured Data	10
1 Statistical Learning Theory	11

TABLE OF CONTENTS

1.1	Statistical Learning Theory	11
1.1.1	The Supervised Learning Problem	12
1.1.2	Loss Function	13
1.1.3	Risk Functionals	14
1.1.4	Empirical Risk Minimization	15
1.1.5	Regularization Theory	16
1.1.6	Sample and Approximation Error	16
1.2	Mathematical Foundations of Kernels	17
1.2.1	Euclidean and Hilbert Spaces	18
1.2.2	Mercer's Theorem	19
1.2.3	Reproducing Kernel Hilbert Spaces	22
1.2.4	Representer Theorem	25
1.3	Support Vector Machines	28
1.3.1	SVMs and the Regularization Theory	29
1.3.2	Primal and Dual Formulations	31
1.3.3	SVMs for Regression	34
1.3.4	Support Vector Clustering	35
1.3.5	Complexity	36
1.3.6	Multiclass Classification	39
2	Voted Perceptron Algorithm	41
2.1	Voted Perceptron	41
2.1.1	Training Algorithm	42
2.1.2	Prediction Function	44
2.2	Dual VP	48
2.3	Regularization	54
2.4	Complexity	57
2.5	Loss Function	58
3	Processing Structured Data	61
3.1	Basic Kernels	61
3.2	Constructing Kernels	62
3.3	Normalizing the Kernel	65
3.4	Kernels for Discrete Objects	66

TABLE OF CONTENTS

3.5	Kernels for Strings	67
3.5.1	Spectrum Kernel	68
3.5.2	Mismatch String Kernel	69
3.5.3	String Subsequence Kernel	70
3.5.4	Weighted String Kernel	72
3.5.5	Dynamic Time–Alignment Kernel	73
3.5.6	Dynamic Alignment Kernel	74
3.5.7	Marginalized Kernel	77
3.5.8	Marginalized Count Kernel	77
3.5.9	The Fisher Kernel	78
3.6	Kernels for Trees	80
3.6.1	Parse Tree Kernel	81
3.6.2	String Tree Kernel	85
3.6.3	Label Mutation Elastic Structure Tree Kernel	87
3.7	Kernels for Graphs	90
3.7.1	Subgraph Kernel	91
3.7.2	Frequent Subgraphs Kernel	91
3.7.3	Cyclic Pattern Kernel	93
3.7.4	Marginalized Graph Kernel	94
3.7.5	Extended Marginalized Graph Kernel	95
3.7.6	A Family of Kernels for Small Molecules	97
3.7.7	Synchronized Random Walks Kernel	99
3.7.8	Walk Based Kernels	101
3.7.9	Tree–structured Pattern Kernel	101
3.7.10	Basic Terms Kernel	102
3.8	Recursive Neural Networks	104
II	Preference Learning	108
4	Preference Learning in Natural Language Processing	109
4.1	Introduction	110
4.2	An Introduction to the Parsing Problem	111
4.3	Ranking and Preference Problems	113
4.3.1	The Utility Function Approach	116

TABLE OF CONTENTS

4.3.2	Recursive Neural Networks Preference Model	116
4.3.3	Kernel Ranking and Preference Model	117
4.3.4	Cancelling Out Effect	120
4.4	Preference Model for SVMs and VP	121
4.4.1	SVMs and Preference Model	121
4.4.2	VP and Preference Model	122
4.4.2.1	Dual VP and Preference Model	123
4.5	Applications to Natural Language	125
4.5.1	The First Pass Attachment	125
4.5.1.1	Tree Reduction and Specialization	129
4.5.2	The Reranking Task	130
4.6	Experimental Results	131
4.6.1	First Pass Attachment	131
4.6.2	Reranking Task	135
4.6.3	The Role of Representation	136
4.6.4	Comparing Different Preference Loss Functions	138
4.7	Conclusions	140
5	On the Consistency of Preference Learning	141
5.1	Introduction	141
5.2	The Bayes Function	142
5.3	A New Model of Preference and Ranking	147
5.3.1	The Partial Order Model	148
5.3.2	The 0–1 Loss Function	149
5.3.3	Three Approaches for the Partial Order Model	149
5.4	A Comparison of the Three Approaches	150
5.4.1	The Direct Model	151
5.4.2	The Utility Function Model	152
5.4.3	The Pairwise Model	154
5.5	Dependence on Size of Set of Alternatives	156
5.5.1	Ranking Two Alternatives	157
5.5.2	Ranking k Alternatives	158
5.6	Conclusions	161

III Structured Kernels for Molecular Biology	163
6 Weighted Decomposition Kernel	164
6.1 Introduction	164
6.2 Decomposition Kernels	167
6.2.1 Equivalence of Tensor Product and Direct Sum	168
6.2.2 All-Substructures Kernels	173
6.3 Weighted Decomposition Kernels	174
6.3.1 Data Types	174
6.3.2 Graph Probability Distribution Kernels	175
6.3.3 General Form of WDKs	177
6.3.4 A WDK for Biological Sequences	178
6.3.5 A WDK for Molecules	179
6.4 Algorithms and Complexity	181
6.4.1 Indexing and Sorting Selectors	182
6.4.2 Computing Histograms	183
6.4.2.1 Sequences Histograms Computation	183
6.4.2.2 Trees Histograms Computation	184
6.4.2.3 DAGs Histograms Computation	186
6.4.3 Reducing HIK Complexity	189
6.4.4 Optimizing Histogram Inner Product	191
6.5 Experimental Results	192
6.5.1 Protein Subcellular Localization	192
6.5.2 Protein Family Classification	197
6.5.3 HIV Dataset	199
6.5.4 Predictive Toxicology Challenge	204
6.6 Conclusions	208
7 Prediction of Zinc Binding Sites	209
7.1 Introduction	209
7.2 Dataset Description and Statistics	212
7.2.1 Data preparation	212
7.2.2 A Taxonomy of Zinc Sites and Sequences	213
7.2.3 Bonding State Autocorrelation	214

TABLE OF CONTENTS

7.2.3.1	Patterns of Binding Sites	216
7.3	Methods	217
7.3.1	Standard Window Based Local Predictor	219
7.3.2	Semipattern Based Predictor	219
7.3.3	Gating Network	220
7.4	Experimental Results	221
7.5	Discussion and Conclusions	226
Conclusions		229
Theoretical Contributions		229
Experimental Contributions		230
A Ranking and Preference Error Probability		232
A.1	A Detailed Solution of the Integral	232
A.2	Another Method to Solve the Integral	234
Index		257