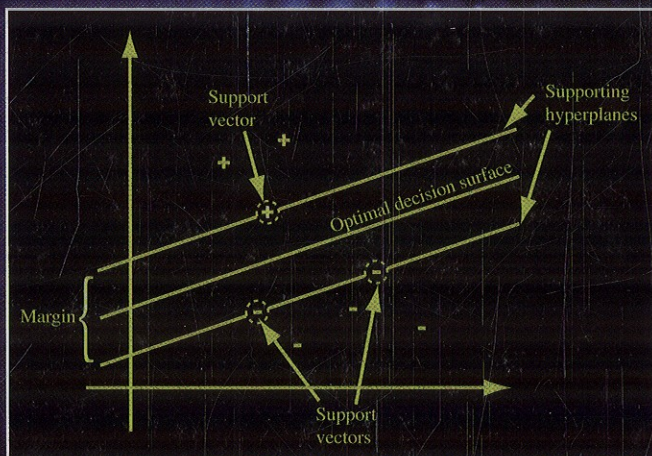


Wiley Series on Methods and Applications in Data Mining
Daniel T. Larose, Series Editor

KNOWLEDGE DISCOVERY WITH SUPPORT VECTOR MACHINES



LUTZ HAMEL

CONTENTS

PREFACE	xiii
PART I	1
1 WHAT IS KNOWLEDGE DISCOVERY?	3
1.1 Machine Learning	4
1.2 Structure of the Universe X	6
1.3 Inductive Learning	8
1.4 Model Representations	9
<i>Exercises</i>	<i>11</i>
Bibliographic Notes	11
2 KNOWLEDGE DISCOVERY ENVIRONMENTS	13
2.1 Computational Aspects of Knowledge Discovery	13
2.1.1 Data Access	14
2.1.2 Visualization	17
2.1.3 Data Manipulation	20
2.1.4 Model Building and Evaluation	23
2.1.5 Model Deployment	26

2.2	Other Tool Sets	27
	Exercises	27
	Bibliographic Notes	28
3	DESCRIBING DATA MATHEMATICALLY	31
3.1	From Data Sets to Vector Spaces	31
3.1.1	Vectors	35
3.1.2	Vector Spaces	40
3.2	The Dot Product as a Similarity Score	41
3.3	Lines, Planes, and Hyperplanes	44
	Exercises	47
	Bibliographic Notes	48
4	LINEAR DECISION SURFACES AND FUNCTIONS	49
4.1	From Data Sets to Decision Functions	49
4.1.1	Linear Decision Surfaces Through the Origin	50
4.1.2	Decision Surfaces with an Offset Term	51
4.2	Simple Learning Algorithm	54
4.3	Discussion	57
	Exercises	58
	Bibliographic Notes	59
5	PERCEPTRON LEARNING	61
5.1	Perceptron Architecture and Training	62
5.2	Duality	67
5.3	Discussion	70
	Exercises	71
	Bibliographic Notes	72
6	MAXIMUM-MARGIN CLASSIFIERS	73
6.1	Optimization Problems	74
6.2	Maximum Margins	75
6.3	Optimizing the Margin	77
6.4	Quadratic Programming	82
6.5	Discussion	86
	Exercises	87
	Bibliographic Notes	88

7	SUPPORT VECTOR MACHINES	91
7.1	The Lagrangian Dual	92
7.2	Dual Maximum-Margin Optimization	97
7.2.1	The Dual Decision Function	101
7.3	Linear Support Vector Machines	102
7.4	Nonlinear Support Vector Machines	103
7.4.1	The Kernel Trick	106
7.4.2	Feature Search	109
7.4.3	A Closer Look at Kernels	109
7.5	Soft-Margin Classifiers	114
7.5.1	The Dual Setting for Soft-Margin Classifiers	118
7.6	Tool Support	122
7.6.1	WEKA	123
7.6.2	R	126
7.7	Discussion	128
	Exercises	130
	Bibliographic Notes	131
8	IMPLEMENTATION	133
8.1	Gradient Ascent	134
8.1.1	The Kernel-Adatron Algorithm	136
8.2	Quadratic Programming	138
8.2.1	Chunking	139
8.3	Sequential Minimal Optimization	142
8.4	Discussion	144
	Exercises	144
	Bibliographic Notes	145
9	EVALUATING WHAT HAS BEEN LEARNED	147
9.1	Performance Metrics	148
9.1.1	The Confusion Matrix	149
9.2	Model Evaluation	152
9.2.1	The Hold-Out Method	155
9.2.2	The Leave-One-Out Method	157
9.2.3	N -Fold Cross-Validation	158

9.3	Error Confidence Intervals	160
9.3.1	Comparison of Models	162
9.4	Model Evaluation in Practice	163
9.4.1	WEKA	163
9.4.2	R	167
	Exercises	169
	Bibliographic Notes	170
10	ELEMENTS OF STATISTICAL LEARNING THEORY	171
10.1	The VC-Dimension and Model Complexity	172
10.2	A Theoretical Setting for Machine Learning	175
10.3	Empirical Risk Minimization	176
10.4	VC-Confidence	177
10.5	Structural Risk Minimization	179
10.6	Discussion	180
	Exercises	180
	Bibliographic Notes	181
PART III		183
11	MULTICLASS CLASSIFICATION	185
11.1	One-Versus-the-Rest Classification	185
11.2	Pairwise Classification	189
11.3	Discussion	192
	Exercises	192
	Bibliographic Notes	192
12	REGRESSION WITH SUPPORT VECTOR MACHINES	193
12.1	Regression as Machine Learning	193
12.2	Simple and Multiple Linear Regression	194
12.3	Regression with Maximum-Margin Machines	197
12.4	Regression with Support Vector Machines	200
12.5	Model Evaluation	202
12.6	Tool Support	203
12.6.1	WEKA	204
12.6.2	R	205
	Exercises	207
	Bibliographic Notes	208

13	NOVELTY DETECTION	209
13.1	Maximum-Margin Machines	210
13.2	The Dual Setting	212
13.3	Novelty Detection in R	214
	Exercises	217
	Bibliographic Notes	217
APPENDIX A	NOTATION	219
APPENDIX B	TUTORIAL INTRODUCTION TO R	221
B.1	Programming Constructs	222
B.2	Data Constructs	224
B.3	Basic Data Analysis	227
	Bibliographic Notes	230
REFERENCES		231
INDEX		237