

An Introduction to Support Vector Machines
and other kernel-based learning methods

NELLO CRISTIANINI AND JOHN SHAWE-TAYLOR

CAMBRIDGE
UNIVERSITY PRESS

Contents

<i>Preface</i>	ix
<i>Notation</i>	xiii
1 The Learning Methodology	1
1.1 Supervised Learning	1
1.2 Learning and Generalisation	3
1.3 Improving Generalisation	4
1.4 Attractions and Drawbacks of Learning	6
1.5 Support Vector Machines for Learning	7
1.6 Exercises	7
1.7 Further Reading and Advanced Topics	8
2 Linear Learning Machines	9
2.1 Linear Classification	9
2.1.1 Rosenblatt's Perceptron	11
2.1.2 Other Linear Classifiers	19
2.1.3 Multi-class Discrimination	20
2.2 Linear Regression	20
2.2.1 Least Squares	21
2.2.2 Ridge Regression	22
2.3 Dual Representation of Linear Machines	24
2.4 Exercises	25
2.5 Further Reading and Advanced Topics	25
3 Kernel-Induced Feature Spaces	26
3.1 Learning in Feature Space	27
3.2 The Implicit Mapping into Feature Space	30
3.3 Making Kernels	32
3.3.1 Characterisation of Kernels	33
3.3.2 Making Kernels from Kernels	42
3.3.3 Making Kernels from Features	44
3.4 Working in Feature Space	46

3.5	Kernels and Gaussian Processes	48
3.6	Exercises	49
3.7	Further Reading and Advanced Topics	50
4	Generalisation Theory	52
4.1	Probably Approximately Correct Learning	52
4.2	Vapnik Chervonenkis (VC) Theory	54
4.3	Margin-Based Bounds on Generalisation	59
4.3.1	Maximal Margin Bounds	59
4.3.2	Margin Percentile Bounds	64
4.3.3	Soft Margin Bounds	65
4.4	Other Bounds on Generalisation and Luckiness	69
4.5	Generalisation for Regression	70
4.6	Bayesian Analysis of Learning	74
4.7	Exercises	76
4.8	Further Reading and Advanced Topics	76
5	Optimisation Theory	79
5.1	Problem Formulation	79
5.2	Lagrangian Theory	81
5.3	Duality	87
5.4	Exercises	89
5.5	Further Reading and Advanced Topics	90
6	Support Vector Machines	93
6.1	Support Vector Classification	93
6.1.1	The Maximal Margin Classifier	94
6.1.2	Soft Margin Optimisation	103
6.1.3	Linear Programming Support Vector Machines	112
6.2	Support Vector Regression	112
6.2.1	ϵ -Insensitive Loss Regression	114
6.2.2	Kernel Ridge Regression	118
6.2.3	Gaussian Processes	120
6.3	Discussion	121
6.4	Exercises	121
6.5	Further Reading and Advanced Topics	122
7	Implementation Techniques	125
7.1	General Issues	125
7.2	The Naive Solution: Gradient Ascent	129
7.3	General Techniques and Packages	135
7.4	Chunking and Decomposition	136
7.5	Sequential Minimal Optimisation (SMO)	137
7.5.1	Analytical Solution for Two Points	138
7.5.2	Selection Heuristics	140
7.6	Techniques for Gaussian Processes	144

7.7	Exercises145
7.8	Further Reading and Advanced Topics146
8	Applications of Support Vector Machines	149
8.1	Text Categorisation150
8.1.1	A Kernel from IR Applied to Information Filtering150
8.2	Image Recognition152
8.2.1	Aspect Independent Classification153
8.2.2	Colour-Based Classification154
8.3	Hand-written Digit Recognition156
8.4	Bioinformatics157
8.4.1	Protein Homology Detection157
8.4.2	Gene Expression159
8.5	Further Reading and Advanced Topics160
A	Pseudocode for the SMO Algorithm	162
B	Background Mathematics	165
B.1	Vector Spaces165
B.2	Inner Product Spaces167
B.3	Hilbert Spaces169
B.4	Operators, Eigenvalues and Eigenvectors171
	<i>References</i>	173
	<i>Index</i>	187