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

NELLO CRISTIANINI AND JOHN SHAWE-TAYLOR

CAMBRIDGE W UNIVERSITY PRESS

Contents

Pr	eface	i	х
No	tation	n xii	ii
1	Learning Methodology	1	
•	1.1	Supervised Learning	_
,	1.2	Learning and Generalisation	
	1.3	Improving Generalisation.	
	1.4	Attractions and Drawbacks of Learning.	
	1.5	Support Vector Machines for Learning	
	1.6	Exercises.	
	1.7	Further Reading and Advanced Topics	
2	Line	ar Learning Machines	9
	2.1	Linear Classification.	9
		2.1.1 Rosenblatt's Perceptron	1
		2.1.2 Other Linear Classifiers	9
		2.1.3 Multi-class Discrimination	0
	2.2	Linear Regression 20	0
		2.2.1 Least Squares	1
		2.2.2 Ridge Regression	2
	2.3	Dual Representation of Linear Machines	4
	2.4	Exercises	5
	2.5	Further Reading and Advanced Topics	5
3	Kerr	nel-Induced Feature Spaces 20	6
	3.1	Learning in Feature Space	7
	3.2	The Implicit Mapping into Feature Space. 30	0
	3.3	Making Kernels	2
		3.3.1 Characterisation of Kernels	3
		3.3.2 Making Kernels from Kernels	2
		3.3.3 Making Kernels from Features	4
	3.4	Working in Feature Space	5

Contents

	3.5	Kernels and Gaussian Processes	48
	3.6	Exercises	. 49
	3.7	Further Reading and Advanced Topics	. 50
4	Gen	neralisation 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	
		4.3.2 Margin Percentile Bounds	
		4.3.3 Soft Margin Bounds	
	4.4	Other Bounds on Generalisation and Luckiness.	
	4.5	Generalisation for Regression	
	4.6	Bayesian Analysis of Learning.	
	4.7	Exercises	
	4.8	Further Reading and Advanced Topics	
5	Opt	imisation Theory	79
	5.1	Problem Formulation.	
	5.2	Lagrangian Theory.	
	5.3	Duality.	
	5.4	Exercises	
	5.5	Further Reading and Advanced Topics	
6	Sup	port Vector Machines	93
6	Sup 6.1	port Vector Machines Support Vector Classification	
6			93
6		Support Vector Classification	93 94
6		Support Vector Classification 6.1.1 The Maximal Margin Classifier	93 94 103
6		Support Vector Classification 6.1.1 The Maximal Margin Classifier 6.1.2 Soft Margin Optimisation	93 94 103 112
6	6.1	Support Vector Classification6.1.1The Maximal Margin Classifier6.1.2Soft Margin Optimisation6.1.3Linear Programming Support Vector Machines	93 94 103 112 112
6	6.1	Support Vector Classification6.1.1The Maximal Margin Classifier.6.1.2Soft Margin Optimisation.6.1.3Linear Programming Support Vector Machines.Support Vector Regression.	93 94 103 112 112 114
6	6.1	Support Vector Classification 6.1.1 The Maximal Margin Classifier 6.1.2 Soft Margin Optimisation 6.1.3 Linear Programming Support Vector Machines Support Vector Regression	93 94 103 112 112 114 118 120
6	6.1	Support Vector Classification6.1.1The Maximal Margin Classifier6.1.2Soft Margin Optimisation6.1.3Linear Programming Support Vector MachinesSupport Vector Regression6.2.1e-Insensitive Loss Regression6.2.2Kernel Ridge Regression	93 94 103 112 112 114 118 120
6	6.1 6.2	Support Vector Classification 6.1.1 The Maximal Margin Classifier. 6.1.2 Soft Margin Optimisation. 6.1.3 Linear Programming Support Vector Machines. Support Vector Regression. 6.2.1 6.2.2 Kernel Ridge Regression. 6.2.3 \ Gaussian Processes. x Discussion. x Exercises. x	93 94 103 112 112 114 118 120 121 121
6	6.1 6.2 6.3	Support Vector Classification 6.1.1 The Maximal Margin Classifier 6.1.2 Soft Margin Optimisation 6.1.3 Linear Programming Support Vector Machines Support Vector Regression 6.2.1 6.2.2 Kernel Ridge Regression 6.2.3 \ Gaussian Processes x	93 94 103 112 112 114 118 120 121 121
6 7	6.1 6.2 6.3 6.4 6.5 Imp	Support Vector Classification 6.1.1 The Maximal Margin Classifier 6.1.2 Soft Margin Optimisation. 6.1.3 Linear Programming Support Vector Machines. Support Vector Regression. 6.2.1 6.2.1 e-Insensitive Loss Regression. 6.2.2 Kernel Ridge Regression. 6.2.3 \ Gaussian Processes. x Discussion x Exercises. Further Reading and Advanced Topics. lementation Techniques Image: Construct Science S	93 94 103 112 112 114 118 120 121 121 122 125
	6.1 6.2 6.3 6.4 6.5 Imp 7.1	Support Vector Classification 6.1.1 The Maximal Margin Classifier. 6.1.2 Soft Margin Optimisation. 6.1.3 Linear Programming Support Vector Machines. Support Vector Regression. . 6.2.1 e-Insensitive Loss Regression. 6.2.2 Kernel Ridge Regression. 6.2.3 \ Gaussian Processes. . Discussion x Exercises. . Further Reading and Advanced Topics. lementation Techniques General Issues.	93 94 103 112 112 114 118 120 121 121 121 122 125
	6.1 6.2 6.3 6.4 6.5 Imp 7.1 7.2	Support Vector Classification 6.1.1 The Maximal Margin Classifier. 6.1.2 Soft Margin Optimisation. 6.1.3 Linear Programming Support Vector Machines. Support Vector Regression. 6.2.1 6.2.1 e-Insensitive Loss Regression. 6.2.2 Kernel Ridge Regression. 6.2.3 \ Gaussian Processes. * Discussion * Exercises. * Further Reading and Advanced Topics. Iementation Techniques General Issues. The Naive Solution: Gradient Ascent.	93 94 103 112 112 114 118 120 121 121 122 125 125 129
	6.1 6.2 6.3 6.4 6.5 Imp 7.1 7.2 7.3	Support Vector Classification 6.1.1 The Maximal Margin Classifier. 6.1.2 Soft Margin Optimisation. 6.1.3 Linear Programming Support Vector Machines. Support Vector Regression. 6.2.1 6.2.1 e-Insensitive Loss Regression. 6.2.2 Kernel Ridge Regression. 6.2.3 Gaussian Processes. Discussion. * Exercises. * Further Reading and Advanced Topics. Iementation Techniques General Issues. The Naive Solution: Gradient Ascent. General Techniques and Packages.	93 94 103 112 112 114 118 120 121 121 121 122 125 129 135
	 6.1 6.2 6.3 6.4 6.5 Imp 7.1 7.2 7.3 7.4 	Support Vector Classification 6.1.1 The Maximal Margin Classifier 6.1.2 Soft Margin Optimisation 6.1.3 Linear Programming Support Vector Machines Support Vector Regression . 6.2.1 e-Insensitive Loss Regression 6.2.2 Kernel Ridge Regression 6.2.3 Gaussian Processes Discussion x Exercises . Further Reading and Advanced Topics . Imentation Techniques . General Issues . The Naive Solution: Gradient Ascent . General Techniques and Packages . Chunking and Decomposition .	93 94 103 112 112 114 118 120 121 121 121 122 125 125 129 135 136
	6.1 6.2 6.3 6.4 6.5 Imp 7.1 7.2 7.3	Support Vector Classification 6.1.1 The Maximal Margin Classifier 6.1.2 Soft Margin Optimisation 6.1.3 Linear Programming Support Vector Machines Support Vector Regression 6.2.1 6.2.1 e-Insensitive Loss Regression 6.2.2 Kernel Ridge Regression 6.2.3 \ Gaussian Processes 5 Discussion x Exercises 5 Further Reading and Advanced Topics 5 Imentation Techniques 5 General Issues 5 The Naive Solution: Gradient Ascent 5 General Techniques and Packages 5 Chunking and Decomposition 5 Sequential Minimal Optimisation (SMO) 5	93 94 103 112 112 114 118 120 121 121 121 122 125 125 129 135 136 137
	 6.1 6.2 6.3 6.4 6.5 Imp 7.1 7.2 7.3 7.4 	Support Vector Classification 6.1.1 The Maximal Margin Classifier 6.1.2 Soft Margin Optimisation 6.1.3 Linear Programming Support Vector Machines. Support Vector Regression. 6.2.1 6.2.1 e-Insensitive Loss Regression. 6.2.2 Kernel Ridge Regression. 6.2.3 \ Gaussian Processes. . Discussion * Exercises. . Further Reading and Advanced Topics. . Iementation Techniques . General Issues. . The Naive Solution: Gradient Ascent. . General Techniques and Packages. . Chunking and Decomposition. . Sequential Minimal Optimisation (SMO). . 7.5.1 Analytical Solution for Two Points.	93 94 103 112 112 114 118 120 121 121 121 122 125 125 129 135 136 137 138
	 6.1 6.2 6.3 6.4 6.5 Imp 7.1 7.2 7.3 7.4 	Support Vector Classification 6.1.1 The Maximal Margin Classifier 6.1.2 Soft Margin Optimisation 6.1.3 Linear Programming Support Vector Machines Support Vector Regression 6.2.1 6.2.1 e-Insensitive Loss Regression 6.2.2 Kernel Ridge Regression 6.2.3 \ Gaussian Processes 5 Discussion x Exercises 5 Further Reading and Advanced Topics 5 Imentation Techniques 5 General Issues 5 The Naive Solution: Gradient Ascent 5 General Techniques and Packages 5 Chunking and Decomposition 5 Sequential Minimal Optimisation (SMO) 5	93 94 103 112 112 114 118 120 121 121 121 122 125 129 135 136 137 138 140

vi

Contents

7.7	Exercises	145
7.8	Further Reading and Advanced Topics	.146
Appl	lications of Support Vector Machines	149
8.1	Text Categorisation	.150
	8.1.1 A Kernel from IR Applied to Information Filtering	150
8.2	Image Recognition	.152
	8.2.1 Aspect Independent Classification	.153
	8.2.2 Colour-Based Classification	.154
8.3	Hand-written Digit Recognition	.156
8.4		
8.5		
Pseu	docode for the SMO Algorithm	162
Back	ground Mathematics	165
B.I	Vector Spaces	165
B.2		
B.3	-	
B.4	-	
erenc	ces	173
ex		187
	7.8 App) 8.1 8.2 8.3 8.4 8.5 Pseu B.3 B.1 B.2 B.3 B.4 erence	 Further Reading and Advanced Topics. Applications of Support Vector Machines 8.1 Text Categorisation 8.1 Text Categorisation 8.1.1 A Kernel from IR Applied to Information Filtering 8.2 Image Recognition. 8.2.1 Aspect Independent Classification 8.2.2 Colour-Based Classification 8.3 Hand-written Digit Recognition. 8.3 Hand-written Digit Recognition. 8.4 Bioinformatics. 8.4.1 Protein Homology^Detection. 8.4.2 Gene Expression. 8.5 Further Reading and Advanced Topics. Pseudocode for the SMO Algorithm Background Mathematics B.1 Vector Spaces. B.2 Inner Product Spaces. B.3 Hilbert Spaces. B.4 Operators, Eigenvalues and Eigenvectors.

vn