Contents

Prej	face		page	e xv
1	Introd	uction gliftcanact energy		1
	1.1	What Is Learning?		1
	1.2	When Do We Need Machine Learning?		3
	1.3	Types of Learning		4
	1.4	Relations to Other Fields		6
	1.5	How to Read This Book		7
	1.6	Notation Notation		8
	1.0			O
Par	t 1 Fo	undations		
2	A Gent	tle Start		13
	2.1	A Formal Model – The Statistical Learning Framework		13
	2.2	Empirical Risk Minimization		15
	2.3	Empirical Risk Minimization with Inductive Bias		16
	2.4	Exercises ** ** ** ** ** ** ** ** ** ** ** ** **		20
3	A Forn	nal Learning Model		22
	3.1	PAC Learning		22
	3.2	A More General Learning Model		23
	3.3	Summary		28
	3.4	Bibliographic Remarks		28
	3.5	Exercises		28
	0.0	Professor - Profes		20
4	Learni	ng via Uniform Convergence		31
	4.1	Uniform Convergence Is Sufficient for Learnability		31
	4.2	Finite Classes Are Agnostic PAC Learnable		32
	4.3	Summary		34
	4.4	Bibliographic Remarks		35
	4.5	Exercises		35

			36	10	Boosting	15.4 Duality*	1	.01
5	The Bias-Complexity Trade-off			176	10.1 V	Weak Learnability 122 and MVZ-1103 animomologic 2.21	1	.02
	5.1 The No-Free-Lunch Theorem		37	177				.05
	5.2 Error Decomposition		10	177	10.3 L	in an Combinations of Base Hypotheses	1	108
	5.3 Summary		11	8791	10.4 A	AdaBoost for Face Recognition		110
	5.4 Bibliographic Remarks		41	22		summary sporties len		111
	5.5 Exercises	- Carrier of	41	179	40 C T	Chlis agentic Domorks	1	111
	J.J Excluses			179	10.0 E	vercises 250EqK SHILAST OHIT SHILDUSONIA 1.0.	1	112
	Ti VO Dimension		43	181				257
6	The VC-Dimension		43	11	Model S	election and Validation WMVZ-Relegatine malgin [8.8]		114
	6.1 Infinite-Size Classes Can Be Learnable		44	187		Model Selection Using SRM	-	115
	6.2 The VC-Dimension		46	188	11.1 T	Validation Semants		116
	6.3 Examples		48	188	11.2	What to Do If Learning Fails	00 1	120
	6.4 The Fundamental Theorem of PAC Lea	HIIIIE	49	001			1	123
	6.5 Proof of Theorem 6.7		53	190		Summary Exercises Random Andrews Research Processor Research	UM.	123
	6.6 Summary		53	190	11.5			
	6.7 Bibliographic Remarks		54	12	Convex	Learning Problems		124
	6.8 Exercises		34		10.1	Convexity, Lipschitzness, and Smoothness		124
			50	198	100	Garage I coming Problems		130
7	Nonuniform Learnability		58	201	12.2	Surrogate Loss Functions which have garden by the page 2.4		134
7	T 1114-		58	206				135
			60	209		Summary		136
	Tangth and Occa	am's Razor	63	210		Biolographic Terraina		136
	CY Lilitar Conciet	tency	66	210	12.6	Exercises		30.0
	D'CC Nations of Les	arnability	67	12	Podular	ization and Stability		137
			70	13	Regulai	= 1 1 1 X = Minimization reference \ Signature \ 1.8		137
	7.6 Summary		70	213	13.1	Regularized Loss Minimization (01651que) significant		139
	7.7 Bibliographic Remarks		71	214	13.2	Stable Rules Do Not Overfit and many A good and good as		140
	7.8 Exercises	t i Foundations		217	13.3	Tikhonov Regularization as a Stabilizer and Implant Est		144
			73	217		Controlling the Fitting-Statisticy France of		146
8	The Runtime of Learning			218		Summary		146
	8.1 Computational Complexity of Learning	2.2 Tempings 1 C.S	74	218	13.6	Bibliographic Remarks		147
			76		13.7	Exercises rodrigion (20)	веИ	14/
	8.3 Efficiently Learnable, but Not by a Pro		80	219				150
	8.4 Hardness of Learning*		81	14	Stocha	stic Gradient Descent		151
	8.5 Summary		82	220	The state of the s	Gradient Descent		154
	8.6 Bibliographic Remarks		82	225	14.2	Subgradients *noitainemelaint ineichtele.		156
			83	25	14.3	Stochastic Gradient Descent (SGD)		
	8.7 Exercises labol grammal			225	14.4	Variants arables and the EM Assamed Singularity States of the Variants arables and the EM Assamed Singularity Sing		159
				25		Learning with SGD		162
P	Part 2 From Theory to Algorithms	3.4 Bibliographic Res			14.6	Summary		165
8			89	28		Bibliographic Remarks		166
9	Linear Predictors			29		Exercises showed Neural Networks		166
		Learning via Uniform Co	90	30				167
	0.2 Linear Regression	4.1 Uniform Converg	94	15	Suppor	rt Vector Machines A The Power of Machines		
	9.3 Logistic Regression		97	34	15.1	Margin and Hard-SVM of No wireless of The		167
	9.4 Summary	4-5 Summary	99	35	152	Soft-SVM and Norm Regularization		171
	9.5 Bibliographic Remarks		99	36	15.3	Optimality Conditions and "Support Vectors"*		175
	9.6 Exercises	4.5 Exercises	99		13,5			

	15.4	Duality*	175	321	20.7 Summary 24.25 24	
	15.4	Implementing Soft-SVM Using SGD	176	321	20.8 Bibliographic Remarks alterna Sidographic Remarks 24	10
			177	322	20.9 Exercises 24	10
		Summary Bibliographic Remarks Heart To an American Deposit Scott	177			
			178	Par	t 3 Additional Learning Models vnoent hoomsybA A his	
	15.8	Exercises and modification of the Lace Recognition and the Lot		21	Online Learning and a solitoral among a solitora	15
	Kornel	10.5 Sammary 10.6 Bibliographic Remarks	179	-	21.1 Online Classification in the Realizable Case 24	16
			179	325	21.2 Online Classification in the Unrealizable Case 25	
	16.1	Embeddings into Feature Spaces seeisrox 3 7.01	181	332	21.2 Online Classification in the Officialization Case 25.3 Online Convex Optimization 25.	
	16.2	The Kernel Trick	186	333	21.3 Online Convex Optimization 25	
		Implementing Soft-SVM with Kernels	187	335	21.4 The Online Letter and Tage	
	16.4	Summary Model Selection Using Stellar World Summary	188	336	21.5 Summer	
	16.5	Bibliographic Remarks goodsbile V 2.11	188		21.0 Biolographie remarks	
1	16.6	Exercises alis animas. Fit oft or radio 4.11	100	337	21.7 Exercises 26	02
I	Multicl	ass, Ranking, and Complex Prediction Problems	190	22	Clustering 26	64
		One-versus-All and All-Pairs	190	0.00	22.1 Linkage-Based Clustering Algorithms 26	66
			193	340	22.2 k-Means and Other Cost Minimization Clusterings 26	68
	17.2	Linear Mutuclass redictors	198	175	22.3 Spectral Clustering to more affiliation and to look 27	
		Structured Output Prediction	201		22.4 Information Bottleneck*	73
	17.4	Ranking and the analysis Resident grimman Massures	206	11-8	ZZ. I MIOIMATON ZOVIZZA	74
		Bipartite Ranking and Multivariate Performance Measures	209	342	22.5 Tringh Devel tel of States	76
1	17.6	Summary Has Sammaras with Market Summary Has Sammary H	210	745	22.0 Summary 325 3 Blockship 211 10 Block 10 10 10	76
		Bibliographic Remarks	210	177	ZZ.7 Biolographie Remarks	76
	17.8	Exercises	210	351	22.8 Exercises State of the Sta	70
3	Decisio	on Trees Villidate bas acutest singak	212	23	Dimensionality Reduction 2	78
	181	Sample Complexity	213	253	23.1 Principal Component Analysis (PCA)	19
	18.1	Sample Complexity Decision Tree Algorithms	213 214	763	25.1 Timepar Component Timeryon (2 City)	79 83
	18.2	Decision Tree Algorithms May 20 Jold oct 2018 21 std 182		355	23.2 Random Projections	
	18.2 18.3	Decision Tree Algorithms (1976) and and analysis of the Random Forests (1976) and analysis of the Random Forest (1976) and analysis of th	214	355 357	23.2 Random Projections 24 23.3 Compressed Sensing 24	83
	18.2 18.3 18.4	Decision Tree Algorithms Random Forests Summary Random Forests Summary	214 217 217	355	23.2 Random Projections 23.3 Compressed Sensing 24.4 PCA or Compressed Sensing? 25.5 PCA or Compressed Sensing?	83 85
	18.2 18.3 18.4 18.5	Decision Tree Algorithms Random Forests Summary Bibliographic Remarks	214 217 217 218	355 357 357	23.2 Random Projections 24 23.3 Compressed Sensing 24 23.4 PCA or Compressed Sensing? 25 23.5 Summary 26	83 85 92
	18.2 18.3 18.4 18.5	Decision Tree Algorithms Random Forests Summary Random Forests Summary	214 217 217	355 357 357 359	23.2 Random Projections 24 23.3 Compressed Sensing 24 23.4 PCA or Compressed Sensing? 25 23.5 Summary 26 23.6 Bibliographic Remarks 26	83 85 92 92 92
111111111111111111111111111111111111111	18.2 18.3 18.4 18.5 18.6	Decision Tree Algorithms Random Forests Summary Bibliographic Remarks Exercises	214 217 217 218	355 357 357	23.2 Random Projections 23 23.3 Compressed Sensing 23 23.4 PCA or Compressed Sensing? 25 23.5 Summary 26 23.6 Bibliographic Remarks 29 23.7 Exercises 29	83 85 92 92 92 93
12 14 14 14 14 14 14 14 14 14 14 14 14 14	18.2 18.3 18.4 18.5 18.6	Decision Tree Algorithms Random Forests Summary Bibliographic Remarks Exercises St Neighbor	214 217 217 218 218 219	355 357 357 359	23.2 Random Projections 23 23.3 Compressed Sensing 23 23.4 PCA or Compressed Sensing? 25 23.5 Summary 26 23.6 Bibliographic Remarks 29 23.7 Exercises 29	83 85 92 92 92
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	18.2 18.3 18.4 18.5 18.6 Neare	Decision Tree Algorithms Random Forests Summary Bibliographic Remarks Exercises St Neighbor k Nearest Neighbors	214 217 217 218 218 219 219	355 357 357 359 359	23.1 Random Projections 28 23.2 Random Projections 28 23.3 Compressed Sensing 28 23.4 PCA or Compressed Sensing? 29 23.5 Summary 29 23.6 Bibliographic Remarks 29 23.7 Exercises 29 Generative Models 29	83 85 92 92 92 93
15 16 16 16 16 16 16 16 16 16 16 16 16 16	18.2 18.3 18.4 18.5 18.6 Neare : 19.1 19.2	Decision Tree Algorithms Random Forests Summary Bibliographic Remarks Exercises St Neighbor k Nearest Neighbors Analysis	214 217 217 218 218 219 219 220	226 728 728 928 24	23.2 Random Projections 24 23.3 Compressed Sensing 25 23.4 PCA or Compressed Sensing? 25 23.5 Summary 25 23.6 Bibliographic Remarks 25 23.7 Exercises 25 Generative Models 26 24.1 Maximum Likelihood Estimator 26	83 85 92 92 92 93
12 14 14 14 14 14 15 16 16 16 16 16 16 16 16 16 16 16 16 16	18.2 18.3 18.4 18.5 18.6 Neare 19.1 19.2 19.3	Decision Tree Algorithms Random Forests Summary Bibliographic Remarks Exercises St Neighbor k Nearest Neighbors Analysis Efficient Implementation*	214 217 217 218 218 219 219 220 225	228 728 728 928 928 24	23.2 Random Projections 22 23.3 Compressed Sensing 22 23.4 PCA or Compressed Sensing? 22 23.5 Summary 22 23.6 Bibliographic Remarks 22 23.7 Exercises 22 Generative Models 24 24.1 Maximum Likelihood Estimator 24 25.3 Naive Bayes 25 26.3 Compressed Sensing? 25 27 28 29 20 20 20 20 20 20 20 20 20	83 85 92 92 92 93 95 95
12 14 14 14 14 14 14 14 14 14 14 14 14 14	18.2 18.3 18.4 18.5 18.6 Neare 19.1 19.2 19.3 19.4	Decision Tree Algorithms Random Forests Summary Bibliographic Remarks Exercises St Neighbor k Nearest Neighbors Analysis Efficient Implementation* Summary (IDE) Inspect Institute State In	214 217 217 218 218 219 219 220 225 225	226 726 726 926 24 106 106	23.2 Random Projections 23.3 Compressed Sensing 23.4 PCA or Compressed Sensing? 23.5 Summary 23.6 Bibliographic Remarks 23.7 Exercises 22 Generative Models 24.1 Maximum Likelihood Estimator 24.2 Naive Bayes 24.3 Linear Discriminant Analysis 25 26 27 28 29 20 20 21 22 23 24 25 26 27 27 28 28 29 20 20 20 20 20 20 20 20 20	83 85 92 92 92 93 95 95 99 90 90
15 16 16 16 16 16 16 16 16 16 16 16 16 16	18.2 18.3 18.4 18.5 18.6 Neare 19.1 19.2 19.3 19.4 19.5	Decision Tree Algorithms Random Forests Summary Bibliographic Remarks Exercises St Neighbor k Nearest Neighbors Analysis Efficient Implementation* Summary Bibliographic Remarks	214 217 217 218 218 219 219 220 225 225 225	226 226 226 24 24 26	23.2 Random Projections 23 23.3 Compressed Sensing 23 23.4 PCA or Compressed Sensing? 25 23.5 Summary 26 23.6 Bibliographic Remarks 27 23.7 Exercises 28 Generative Models 29 24.1 Maximum Likelihood Estimator 29 24.2 Naive Bayes 24 24.3 Linear Discriminant Analysis 33 24.4 Latent Variables and the EM Algorithm 3	83 85 92 92 93 93 95 95 99 600 601
15 15 15 15 15 15 15 15 15 15 15 15 15 1	18.2 18.3 18.4 18.5 18.6 Neare 19.1 19.2 19.3 19.4 19.5	Decision Tree Algorithms Random Forests Summary Bibliographic Remarks Exercises St Neighbor k Nearest Neighbors Analysis Efficient Implementation* Summary Bibliographic Remarks Exercises GOZ distribution Analysis Exercises GOZ distribution Analysis Exercises	214 217 217 218 218 219 219 220 225 225	226 726 726 926 24 106 106	23.2 Random Projections 23.3 Compressed Sensing 23.4 PCA or Compressed Sensing? 23.5 Summary 23.6 Bibliographic Remarks 23.7 Exercises 24.1 Maximum Likelihood Estimator 24.2 Naive Bayes 24.3 Linear Discriminant Analysis 24.4 Latent Variables and the EM Algorithm 24.5 Bayesian Reasoning	83 85 92 92 92 93 95 95 95 96 90 801 805
15 15 15 15 15 15 15 15 15 15 15 15 15 1	18.2 18.3 18.4 18.5 18.6 Neare 19.1 19.2 19.3 19.4 19.5	Decision Tree Algorithms Random Forests Random Forests Summary Bibliographic Remarks Exercises St Neighbor k Nearest Neighbors Analysis Efficient Implementation* Summary Bibliographic Remarks CALL Summary CALL S	214 217 217 218 218 219 219 220 225 225 225 225	226 226 226 226 24 26 26 26 26 26 26 26 26 26 26 26 26 26	23.2 Random Projections 23.3 Compressed Sensing 23.4 PCA or Compressed Sensing? 23.5 Summary 23.6 Bibliographic Remarks 23.7 Exercises 24.1 Maximum Likelihood Estimator 24.2 Naive Bayes 24.3 Linear Discriminant Analysis 24.4 Latent Variables and the EM Algorithm 24.5 Bayesian Reasoning 24.6 Summary 22.2 Sandard Policy Sensitive Summary 23.3 Compressed Sensing 24.4 Summary 25.5 Summary 26.5 Summary 27.5 Sandard Policy Sensitive Summary 28.5 Summary 29.5 Su	83 85 92 92 92 93 95 95 96 90 801 805
15	18.2 18.3 18.4 18.5 18.6 Neare 19.1 19.2 19.3 19.4 19.5 19.6 Neura	Pecision Tree Algorithms Random Forests Summary Bibliographic Remarks Exercises **Neighbor** **Nearest Neighbors** Analysis Efficient Implementation** Summary Bibliographic Remarks **Exercises** **Constant of the first of	214 217 217 218 218 219 219 220 225 225 225 225 225 225	226 226 226 24 24 26	23.2 Random Projections 23.3 Compressed Sensing 23.4 PCA or Compressed Sensing? 23.5 Summary 23.6 Bibliographic Remarks 23.7 Exercises 24.1 Maximum Likelihood Estimator 24.2 Naive Bayes 24.3 Linear Discriminant Analysis 24.4 Latent Variables and the EM Algorithm 24.5 Bayesian Reasoning 24.6 Summary 24.7 Bibliographic Remarks	83 85 92 92 92 93 95 95 99 600 601 605 607
15 15 15 15 15 15 15 15 15 15 15 15 15 1	18.2 18.3 18.4 18.5 18.6 Neare 19.1 19.2 19.3 19.4 19.5 19.6 Neura	Pecision Tree Algorithms Random Forests Summary Bibliographic Remarks Exercises **Neighbor** **Nearest Neighbors** Analysis Efficient Implementation** Summary Bibliographic Remarks Exercises **Commany Bibliographic Remarks **Exercises **Commany Bibliographic Remarks **Exercises **Incommany Bibliographic Remark	214 217 217 218 218 219 219 220 225 225 225 225 225 225 225 225	226 226 226 226 24 26 26 26 26 26 26 26 26 26 26 26 26 26	23.2 Random Projections 23.3 Compressed Sensing 23.4 PCA or Compressed Sensing? 23.5 Summary 23.6 Bibliographic Remarks 23.7 Exercises 24.1 Maximum Likelihood Estimator 24.2 Naive Bayes 24.3 Linear Discriminant Analysis 24.4 Latent Variables and the EM Algorithm 24.5 Bayesian Reasoning 24.6 Summary 24.7 Bibliographic Remarks 24.8 Exercises 25.	83 85 92 92 93 95 95 96 90 801 805 807 808
15	18.2 18.3 18.4 18.5 18.6 Neare 19.1 19.2 19.3 19.4 19.5 19.6 Neura	Decision Tree Algorithms Random Forests Random Forests Summary Bibliographic Remarks Exercises St Neighbor k Nearest Neighbors Analysis Efficient Implementation* Summary Bibliographic Remarks Exercises Commany Close Institute Close Ins	214 217 217 218 218 219 219 220 225 225 225 225 225 228 229 230	226 726 726 726 726 726 726 726 726 726	23.2 Random Projections 23.3 Compressed Sensing 23.4 PCA or Compressed Sensing? 23.5 Summary 23.6 Bibliographic Remarks 23.7 Exercises 24.1 Maximum Likelihood Estimator 24.2 Naive Bayes 24.3 Linear Discriminant Analysis 24.4 Latent Variables and the EM Algorithm 24.5 Bayesian Reasoning 24.6 Summary 24.7 Bibliographic Remarks 24.8 Exercises 25. Summary 26.8 Exercises 26. Summary 27. Bibliographic Remarks 28. Summary 29. S	83 85 92 92 93 95 95 96 90 801 805 807 808
15	18.2 18.3 18.4 18.5 18.6 Neare 19.1 19.2 19.3 19.4 19.5 19.6 Neura 20.1 20.2	Decision Tree Algorithms Random Forests Summary Bibliographic Remarks Exercises St Neighbor k Nearest Neighbors Analysis Efficient Implementation* Summary Bibliographic Remarks Exercises I Networks Feedforward Neural Networks Learning Neural Networks The Expressive Power of Neural Networks	214 217 217 218 218 219 219 220 225 225 225 225 225 228 229 230 231	24 1026 24 1036 1036 1036 1036 1036 1036 1036 1036	23.2 Random Projections 23.3 Compressed Sensing 23.4 PCA or Compressed Sensing? 23.5 Summary 23.6 Bibliographic Remarks 23.7 Exercises 24.1 Maximum Likelihood Estimator 24.2 Naive Bayes 24.3 Linear Discriminant Analysis 24.4 Latent Variables and the EM Algorithm 24.5 Bayesian Reasoning 24.6 Summary 24.7 Bibliographic Remarks 24.8 Exercises 3 Feature Selection and Generation 22.2 Sandon Projections 24.2 Paragraphic Remarks 25.3 Summary 26.4 Summary 27.4 Bibliographic Remarks 28.4 Exercises 3 Feature Selection and Generation 3 3 3 3 3 4 4 5 6 6 6 7 7 7 7 7 7 7 7 7 7	83 85 92 92 92 93 95 95 96 90 805 807 807 808
15	18.2 18.3 18.4 18.5 18.6 Neare 19.1 19.2 19.3 19.4 19.5 19.6 Neura 20.1 20.2	Decision Tree Algorithms Random Forests Summary Bibliographic Remarks Exercises St Neighbor k Nearest Neighbors Analysis Efficient Implementation* Summary Bibliographic Remarks Exercises I Networks Feedforward Neural Networks The Expressive Power of Neural Networks The Sample Complexity of Neural Networks	214 217 217 218 218 219 219 220 225 225 225 225 225 228 229 230 231 234	278 728 728 728 728 728 728 728 728 728	23.2 Random Projections 23.3 Compressed Sensing 23.4 PCA or Compressed Sensing? 23.5 Summary 23.6 Bibliographic Remarks 23.7 Exercises 22.6 Generative Models 24.1 Maximum Likelihood Estimator 24.2 Naive Bayes 24.3 Linear Discriminant Analysis 24.4 Latent Variables and the EM Algorithm 24.5 Bayesian Reasoning 24.6 Summary 24.7 Bibliographic Remarks 24.8 Exercises Feature Selection and Generation 25.1 Feature Selection 32.6 Sensing 24.7 Bibliographic Remarks 25.1 Feature Selection 35.1 Feature Selection 36.2 Sensing 37.2 Sensing 38.2 Sensing 39.3 Sensing 39.4 Sensing 39	83 85 92 92 93 93 95 95 96 805 807 807 808
15	18.2 18.3 18.4 18.5 18.6 Neare: 19.1 19.2 19.3 19.4 19.5 19.6 Neura: 20.1 20.2 20.3 20.4	Decision Tree Algorithms Random Forests Summary Bibliographic Remarks Exercises st Neighbor k Nearest Neighbors Analysis Efficient Implementation* Summary Bibliographic Remarks Exercises I Networks Feedforward Neural Networks Learning Neural Networks The Expressive Power of Neural Networks The Sample Complexity of Neural Networks The Sample Complexity of Neural Networks	214 217 217 218 218 219 219 220 225 225 225 225 225 228 229 230 231	24 1026 24 1036 1036 1036 1036 1036 1036 1036 1036	23.2 Random Projections 23.3 Compressed Sensing 23.4 PCA or Compressed Sensing? 23.5 Summary 23.6 Bibliographic Remarks 23.7 Exercises 24.1 Maximum Likelihood Estimator 24.2 Naive Bayes 24.3 Linear Discriminant Analysis 24.4 Latent Variables and the EM Algorithm 24.5 Bayesian Reasoning 24.6 Summary 24.7 Bibliographic Remarks 24.8 Exercises Feature Selection and Generation 25.1 Feature Selection 25.2 Feature Manipulation and Normalization 3	83 85 92 92 92 93 95 95 96 90 805 807 807 808

ents	A STATE OF THE STA		Contents xiii
25.4	Summary Vienness Vien	321	B.5 Bennet's and Bernstein's Inequalities 376
	Bibliographic Remarks	321	B.6 Slud's Inequality
	Exercises Exercises	322	B.7 Concentration of χ^2 Variables 378
4 Adva	anced Theory alphoM gnimes LibrolibaA &	Part	Appendix C Linear Algebra 380
Radema	acher Complexities	325	C.1 Basic Definitions 380
26.1	The Rademacher Complexity	325	C.2 Eigenvalues and Eigenvectors 381
26.2	Rademacher Complexity of Linear Classes	332	C.3 Positive Definite Matrices 381
26.3	Generalization Bounds for SVM	333	C.4 Singular Value Decomposition (SVD) 381
26.4	Generalization Bounds for Predictors with Low ℓ_1 Norm	335	References 385
	Bibliographic Remarks	336	Index 395
100	21.6 Bibliographic Remarks	337	
	g Humbers	337	
27.1	Covering From Covering to Rademacher Complexity via Chaining	338	
27.2	Bibliographic Remarks	340	
A.Fra	22.2 & Means and Other Cost Minimization Chistennes		The Prince of th
Proof o	f the Fundamental Theorem of Learning Theory	341	in dis
28.1	The Upper Bound for the Agnostic Case	341	1853 17
28.2	The Lower Bound for the Agnostic Case	342	by 1
28.3	The Upper Bound for the Realizable Case	347	multi-man and a second
Multicla	ass Learnability socious 1 8.00	351	
	The Natarajan Dimension	351	
	The Multiclass Fundamental Theorem	352	
	Calculating the Natarajan Dimension	353	algorithms of the action and the action of the action of the action of the state of the
	On Good and Bad ERMs	355	Moralista de la casa de
29.5	Bibliographic Remarks	357	Attended to the second of the
29.6	Exercises Sensing? Exercises	357	MACHINE CONTROL OF THE CONTROL OF TH
18.6	23.5 Summary consumary address Summary	359	energian as
	23.6 Bibliographic Remarks 23.7 Exercises		gent brings
ALTONOMIS SERVICES	Compression Bounds	359	experience year's a second
	Examples sleboli evits and	361 363	ing tools are the second of th
30.3	Bibliographic Remarks	303	adapt.
PAC-Ba		364	The first gual or the balance
	PAC-Bayes Bounds alaylan A manimina of Tabul 1.4.2	364	10 the main concept graps +
	Bibliographic Remarks	366	machine learn? How do we go said have a second as
5.700.00	Exercises pamoese A gamoese A gamoes	366	18 learning always possible? Converting the state of the
51.5		208	falled?
endix A	Technical Lemmas all and Policies Remains	369	The second goal of this book is to present several key macrass, in across any
endix R	24.8 Exercises Neasure Concentration	372	in practice and on the other hand give a wife constraint of defining the d
Page 1		372	techniques. Additionally, we may specific attention to algorithms appropriate for
B.1 B.2	Markov's Inequality Chebyshev's Inequality	373	large-scale learning (a.k.u. "Big Data"), since in second vexes, our world has become
B.2 B.3	Chernoff's Bounds	373	increasingly "digitized" and the amount of data available for learning is dramati-
	Hoeffding's Inequality	375	cally increasing. As a result, in many applications data is plentiful and computation