

Contents

<i>List of Figures</i>	xi
<i>List of Tables</i>	xiii
<i>Preface</i>	xv
<i>Acknowledgments</i>	xix
1 Introduction	1
Ambitions of the twentieth century	2
Pattern classification	4
Prediction and action	7
Chapter notes	9
2 Fundamentals of Prediction	11
Modeling knowledge	13
Prediction via optimization	16
Types of errors and successes	20
The Neyman-Pearson Lemma	22
Decisions that discriminate	26
Chapter notes	30
3 Supervised Learning	33
Sample versus population	33
Supervised learning	34
A first learning algorithm: The perceptron	37
Connection to empirical risk minimization	38
Formal guarantees for the perceptron	41
Chapter notes	46
4 Representations and Features	49
Measurement	50
Quantization	51
Template matching	52
Summarization and histograms	53
Nonlinear predictors	54
Chapter notes	65

5	Optimization	69
	Optimization basics	70
	Gradient descent	71
	Applications to empirical risk minimization	74
	Insights from quadratic functions	76
	Stochastic gradient descent	78
	Analysis of the stochastic gradient method	84
	Implicit convexity	87
	Regularization	90
	Squared loss methods and other optimization tools	94
	Chapter notes	96
6	Generalization	99
	Generalization gap	99
	Overparameterization: Empirical phenomena	100
	Theories of generalization	105
	Algorithmic stability	109
	Model complexity and uniform convergence	115
	Generalization from algorithms	118
	Looking ahead	123
	Chapter notes	123
7	Deep Learning	125
	Deep models and feature representation	126
	Optimization of deep nets	128
	Vanishing gradients	134
	Generalization in deep learning	137
	Chapter notes	141
8	Datasets	143
	The scientific basis of machine learning benchmarks	144
	A tour of datasets in different domains	145
	Longevity of benchmarks	156
	Harms associated with data	164
	Toward better data practices	169
	Limits of data and prediction	172
	Chapter notes	173
9	Causality	175
	The limitations of observation	176
	Causal models	178
	Causal graphs	182
	Interventions and causal effects	184

Confounding	186
Experimentation, randomization, potential outcomes	189
Counterfactuals	192
Chapter notes	197
10 Causal Inference in Practice	199
Design and inference	200
The observational basics: Adjustment and controls	201
Reductions to model fitting	202
Quasi-experiments	206
Limitations of causal inference in practice	209
Chapter notes	211
11 Sequential Decision Making and Dynamic Programming	213
From predictions to actions	214
Dynamical systems	214
Optimal sequential decision making	216
Dynamic programming	217
Computation	220
Partial observation and the separation heuristic	225
Chapter notes	230
12 Reinforcement Learning	231
Exploration-exploitation trade-offs: Regret and PAC-error	232
Unknown models and approximate dynamic programming	241
Certainty equivalence is often optimal	248
The limits of learning in feedback loops	253
Chapter notes	259
13 Epilogue	261
Beyond pattern classification?	264
14 Mathematical Background	265
Common notation	265
Multivariable calculus and linear algebra	265
Probability	267
Estimation	272
<i>Bibliography</i>	275
<i>Index</i>	295