Contents

Preface **vii**

1 The Learning Problem **1**
1.1 Problem Setup . . . 1
1.1.1 Components of Learning . . . 3
1.1.2 A Simple Learning Model . . . 5
1.1.3 Learning versus Design . . . 9
1.2 Types of Learning . . . 11
1.2.1 Supervised Learning . . . 11
1.2.2 Reinforcement Learning . . . 12
1.2.3 Unsupervised Learning . . . 13
1.2.4 Other Views of Learning . . . 14
1.3 Is Learning Feasible? . . . 15
1.3.1 Outside the Data Set . . . 16
1.3.2 Probability to the Rescue . . . 18
1.3.3 Feasibility of Learning . . . 24
1.4 Error and Noise . . . 27
1.4.1 Error Measures . . . 28
1.4.2 Noisy Targets . . . 30
1.5 Problems . . . 33

2 Training versus Testing **39**
2.1 Theory of Generalization . . . 39
2.1.1 Effective Number of Hypotheses . . . 41
2.1.2 Bounding the Growth Function . . . 46
2.1.3 The VC Dimension . . . 50
2.1.4 The VC Generalization Bound . . . 53
2.2 Interpreting the Generalization Bound . . . 55
2.2.1 Sample Complexity . . . 57
2.2.2 Penalty for Model Complexity . . . 58
2.2.3 The Test Set . . . 59
2.2.4 Other Target Types . . . 61
2.3 Approximation-Generalization Tradeoff . . . 62

2.3.1 Bias and Variance 62
2.3.2 The Learning Curve 66
2.4 Problems 69

3 The Linear Model **77**
3.1 Linear Classification 77
3.1.1 Non-Separable Data 79
3.2 Linear Regression 82
3.2.1 The Algorithm 84
3.2.2 Generalization Issues 87
3.3 Logistic Regression 88
3.3.1 Predicting a Probability 89
3.3.2 Gradient Descent 93
3.4 Nonlinear Transformation 99
3.4.1 The $\mathcal{Z}$ Space 99
3.4.2 Computation and Generalization 104
3.5 Problems 109

4 Overfitting **119**
4.1 When Does Overfitting Occur? 119
4.1.1 A Case Study: Overfitting with Polynomials 120
4.1.2 Catalysts for Overfitting 123
4.2 Regularization 126
4.2.1 A Soft Order Constraint 128
4.2.2 Weight Decay and Augmented Error 132
4.2.3 Choosing a Regularizer: Pill or Poison? 134
4.3 Validation 137
4.3.1 The Validation Set 138
4.3.2 Model Selection 141
4.3.3 Cross Validation 145
4.3.4 Theory Versus Practice 151
4.4 Problems 154

5 Three Learning Principles **167**
5.1 Occam's Razor 167
5.2 Sampling Bias 171
5.3 Data Snooping 173
5.4 Problems 178

Epilogue **181**

Further Reading **183**

Appendix Proof of the VC Bound **187**
A.1 Relating Generalization Error to In-Sample Deviations 188
A.2 Bounding Worst Case Deviation Using the Growth Function . . 190
A.3 Bounding the Deviation between In-Sample Errors 191

Notation **193**

Index **197**