Contents

Foreword xv

Preface xvii

Who This Book is For . . . xvii

How to Use This Book . . . xviii

Should You Buy This Book? . . . xviii

1 Introduction 1

1.1 What is Machine Learning . . . 1

1.2 Types of Learning . . . 1

1.2.1 Supervised Learning . . . 1

1.2.2 Unsupervised Learning . . . 2

1.2.3 Semi-Supervised Learning . . . 2

1.2.4 Reinforcement Learning . . . 3

1.3 How Supervised Learning Works . . . 3

1.4 Why the Model Works on New Data . . . 7

2 Notation and Definitions 9

2.1 Notation . . . 9

2.1.1 Data Structures . . . 9

2.1.2 Capital Sigma Notation . . . 10

2.1.3 Capital Pi Notation . . . 11

2.1.4 Operations on Sets . . . 11
2.1.5 Operations on Vectors . . . 11
2.1.6 Functions . . . 12
2.1.7 Max and Arg Max . . . 13
2.1.8 Assignment Operator . . . 14
2.1.9 Derivative and Gradient . . . 14
2.2 Random Variable . . . 15
2.3 Unbiased Estimators . . . 17
2.4 Bayes' Rule . . . 17
2.5 Parameter Estimation . . . 17
2.6 Parameters vs. Hyperparameters . . . 18
2.7 Classification vs. Regression . . . 19
2.8 Model-Based vs. Instance-Based Learning . . . 19
2.9 Shallow vs. Deep Learning . . . 20

3 Fundamental Algorithms **21**
3.1 Linear Regression . . . 21
3.1.1 Problem Statement . . . 21
3.1.2 Solution . . . 23
3.2 Logistic Regression . . . 25
3.2.1 Problem Statement . . . 25
3.2.2 Solution . . . 26
3.3 Decision Tree Learning . . . 27
3.3.1 Problem Statement . . . 27
3.3.2 Solution . . . 28
3.4 Support Vector Machine . . . 30
3.4.1 Dealing with Noise . . . 31
3.4.2 Dealing with Inherent Non-Linearity . . . 32
3.5 k-Nearest Neighbors . . . 34

4 Anatomy of a Learning Algorithm 35

4.1 Building Blocks of a Learning Algorithm 35

4.2 Gradient Descent 36

4.3 How Machine Learning Engineers Work 41

4.4 Learning Algorithms' Particularities 41

5 Basic Practice 43

5.1 Feature Engineering 43

5.1.1 One-Hot Encoding 44

5.1.2 Binning 44

5.1.3 Normalization 45

5.1.4 Standardization 45

5.1.5 Dealing with Missing Features 46

5.1.6 Data Imputation Techniques 47

5.2 Learning Algorithm Selection 47

5.3 Three Sets 49

5.4 Underfitting and Overfitting 51

5.5 Regularization 52

5.6 Model Performance Assessment 54

5.6.1 Confusion Matrix 55

5.6.2 Precision/Recall 55

5.6.3 Accuracy 56

5.6.4 Cost-Sensitive Accuracy 57

5.6.5 Area under the ROC Curve (AUC) 58

5.7 Hyperparameter Tuning 58

5.7.1 Cross-Validation 60

6 Neural Networks and Deep Learning **61**

6.1 Neural Networks 61

6.1.1 Multilayer Perceptron Example 62

6.1.2 Feed-Forward Neural Network Architecture 64

6.2 Deep Learning 65

6.2.1 Convolutional Neural Network 65

6.2.2 Recurrent Neural Network 72

7 Problems and Solutions **77**

7.1 Kernel Regression 77

7.2 Multiclass Classification 78

7.3 One-Class Classification 79

7.4 Multi-Label Classification 81

7.5 Ensemble Learning 83

7.5.1 Boosting and Bagging 83

7.5.2 Random Forest 84

7.5.3 Gradient Boosting 85

7.6 Learning to Label Sequences 87

7.7 Sequence-to-Sequence Learning 89

7.8 Active Learning 90

7.9 Semi-Supervised Learning 91

7.10 One-Shot Learning 94

7.11 Zero-Shot Learning 95

8 Advanced Practice **97**

8.1 Handling Imbalanced Datasets 97

8.2 Combining Models 99

8.3 Training Neural Networks 99

8.4 Advanced Regularization 100

8.5 Handling Multiple Inputs 101

8.6 Handling Multiple Outputs . . . 102
8.7 Transfer Learning . . . 102
8.8 Algorithmic Efficiency . . . 104

9 Unsupervised Learning **107**
9.1 Density Estimation . . . 107
9.2 Clustering . . . 109
9.2.1 K-Means . . . 110
9.2.2 DBSCAN and HDBSCAN . . . 111
9.2.3 Determining the Number of Clusters . . . 112
9.2.4 Other Clustering Algorithms . . . 114
9.3 Dimensionality Reduction . . . 118
9.3.1 Principal Component Analysis . . . 118
9.3.2 UMAP . . . 119

10 Other Forms of Learning **121**
10.1 Metric Learning . . . 121
10.2 Learning to Rank . . . 123
10.3 Learning to Recommend . . . 125
10.3.1 Factorization Machines . . . 126
10.3.2 Denoising Autoencoders . . . 128
10.4 Self-Supervised Learning: Word Embeddings . . . 128
10.5 Outlier Detection . . . 130

11 Conclusion **131**
11.1 What Wasn't Covered . . . 131
11.1.1 Topic Modeling . . . 131
11.1.2 Gaussian Processes . . . 132
11.1.3 Generalized Linear Models . . . 132
11.1.4 Probabilistic Graphical Models . . . 132
11.1.5 Markov Chain Monte Carlo . . . 132

11.1.6 Generative Adversarial Networks 133
11.1.7 Genetic Algorithms 133
11.1.8 Reinforcement Learning 134
11.2 Acknowledgements 134
11.3 What to Read Next 135

Index **137**