

Contents

Preface — V

Part I: Introduction and preliminaries

- 1 Introduction to machine learning — 3**
 - 1.1 Different paradigms — 3
 - 1.2 Supervised learning — 3
 - 1.3 Model selection — 6
 - 1.4 No-free-lunch theorem — 11
 - 1.5 ML pipeline in practice — 12
 - 1.6 Implementation details — 12
- 2 Probability review — 14**
 - 2.1 Motivation — 14
 - 2.2 Probability space — 15
 - 2.3 Independence and conditioning — 16
 - 2.4 Random variables — 17
 - 2.5 Discrete random variables — 18
 - 2.6 Continuous random variables — 21
 - 2.7 Moments — 24
 - 2.8 Samples from an unknown distribution — 29
- 3 Optimization — 34**
 - 3.1 Definitions — 34
 - 3.2 Gradient-based algorithms — 40
 - 3.3 Nonconvex case — 48
 - 3.4 Constrained optimization: duality theory — 49

Part II: Supervised learning

- 4 Statistical learning theory — 57**
 - 4.1 Learnability — 57
 - 4.2 Finite-sized hypothesis classes — 60
 - 4.3 Bias-complexity tradeoff and bias-variance tradeoff — 64
- 5 Linear models — 71**
 - 5.1 Linear regression — 71

5.2	Classification — 78
5.3	Implementation details — 86
6	Kernel methods — 88
6.1	Inner products and kernels — 88
6.2	Reproducible kernel Hilbert spaces — 91
6.3	Mercer's theorem — 97
6.4	Representer theorem — 99
6.5	Kernel (ridge) regression — 100
7	Gaussian processes — 103
7.1	Formal definition — 104
7.2	Smoothness properties — 107
7.3	Gaussian processes and kernel methods — 112
7.4	Implementation details — 115
8	Deep learning — 117
8.1	Fully connected dense neural networks — 117
8.2	Back propagation — 120
8.3	Approximation theorems — 124
8.4	Beyond feed forward neural networks — 126
8.5	Some elements of modern theory of neural networks — 130
8.6	Practical implementation considerations — 139
8.7	Implementation details — 144
9	Ensemble methods — 147
9.1	Weak learner — 148
9.2	AdaBoost — 149
9.3	Gradient boosting — 158
9.4	Implementation details — 160

Part III: **Beyond supervised learning**

10	Topics in unsupervised learning — 163
10.1	Clustering — 163
10.2	Dimensionality reduction — 170
10.3	Implementation details — 176
11	Reinforcement learning — 177
11.1	Markov decision processes — 177
11.2	Optimality — 181

11.3	Iterative methods —	182
11.4	Policy gradient method —	185

Bibliography —	193
-----------------------	------------

Index —	197
----------------	------------