

Contents

Prologue

xi

1 Introduction

1

1.1	The Topic At Hand	2
1.2	Learning Is Key	3
1.3	A Little Bit Of History	4
1.4	Key Methodologies Covered In This Book	6
1.5	Classical Mathematical Modelling	9
1.6	Machine Learning Is Different	11
1.7	Simplicity Leading To Complexity	12

2 General Matters

17

2.1	Jargon And Notation	17
2.2	Scaling	18
2.3	Measuring Distances	19
2.4	Curse Of Dimensionality	20
2.5	Principal Components Analysis	21
2.6	Maximum Likelihood Estimation	22
2.7	Confusion Matrix	26
2.8	Cost Functions	28
2.9	Gradient Descent	33
2.10	Training, Testing And Validation	35
2.11	Bias And Variance	37
2.12	Lagrange Multipliers	44
2.13	Multiple Classes	45
2.14	Information Theory And Entropy	47
2.15	Natural Language Processing	50
2.16	Bayes Theorem	51
2.17	What Follows	53

3	<i>K</i> Nearest Neighbours	55
3.1	Executive Summary	55
3.2	What Is It Used For?	55
3.3	How It Works	56
3.4	The Algorithm	58
3.5	Problems With KNN	58
3.6	Example: Heights and weights	59
3.7	Regression	62
4	<i>K</i> Means Clustering	65
4.1	Executive Summary	65
4.2	What Is It Used For?	65
4.3	What Does <i>K</i> Means Clustering Do?	67
4.4	Scree Plots	71
4.5	Example: Crime in England, 13 dimensions	72
4.6	Example: Volatility	75
4.7	Example: Interest rates and inflation	77
4.8	Example: Rates, inflation and GDP	80
4.9	A Few Comments	81
5	Naïve Bayes Classifier	83
5.1	Executive Summary	83
5.2	What Is It Used For?	83
5.3	Using Bayes Theorem	84
5.4	Application Of NBC	84
5.5	In Symbols	85
5.6	Example: Political speeches	86
6	Regression Methods	91
6.1	Executive Summary	91
6.2	What Is It Used For?	91
6.3	Linear Regression In Many Dimensions	92
6.4	Logistic Regression	93
6.5	Example: Political speeches again	95
6.6	Other Regression Methods	96
7	Support Vector Machines	99
7.1	Executive Summary	99
7.2	What Is It Used For?	99
7.3	Hard Margins	100
7.4	Example: Irises	102
7.5	Lagrange Multiplier Version	104
7.6	Soft Margins	106
7.7	Kernel Trick	108

8	Self-Organizing Maps	113
8.1	Executive Summary	113
8.2	What Is It Used For?	113
8.3	The Method	114
8.4	The Learning Algorithm	116
8.5	Example: Grouping shares	119
8.6	Example: Voting in the House of Commons	124
9	Decision Trees	127
9.1	Executive Summary	127
9.2	What Is It Used For?	127
9.3	Example: Magazine subscription	129
9.4	Entropy	134
9.5	Overfitting And Stopping Rules	137
9.6	Pruning	137
9.7	Numerical Features	138
9.8	Regression	139
9.9	Looking Ahead	144
9.10	Bagging And Random Forests	145
10	Neural Networks	147
10.1	Executive Summary	147
10.2	What Is It Used For?	147
10.3	A Very Simple Network	147
10.4	Universal Approximation Theorem	149
10.5	An Even Simpler Network	150
10.6	The Mathematical Manipulations In Detail	151
10.7	Common Activation Functions	154
10.8	The Goal	156
10.9	Example: Approximating a function	157
10.10	Cost Function	158
10.11	Backpropagation	159
10.12	Example: Character recognition	162
10.13	Training And Testing	164
10.14	More Architectures	168
10.15	Deep Learning	170
11	Reinforcement Learning	173
11.1	Executive Summary	173
11.2	What Is It Used For?	173
11.3	Going Offroad In Your Lamborghini 400 GT	174
11.4	Jargon	175
11.5	A First Look At Blackjack	176
11.6	The Classical MDP Approach In O&Xs	177

11.7	More Jargon	179
11.8	Example: The multi-armed bandit	180
11.9	Getting More Sophisticated 1	183
11.10	Example: A maze	186
11.11	Value Notation	190
11.12	The Bellman Equations	192
11.13	Optimal Policy	193
11.14	The Role Of Probability	194
11.15	Getting More Sophisticated 2	195
11.16	Monte Carlo Policy Evaluation	195
11.17	Temporal Difference Learning	198
11.18	Pros And Cons: MC v TD	200
11.19	Finding The Optimal Policy	200
11.20	Sarsa	201
11.21	Q Learning	203
11.22	Example: Blackjack	204
11.23	Large State Spaces	215
Datasets		217
Epilogue		221
Index		223