Contents

List of Pseudocode Examples i

Online Code Examples ii

Preface iii

1 Artificial Neural Networks 1

1.1 Introduction 1
1.2 What is an Artificial Neural Network? 2
1.3 The Origins of Neural Networks 5
1.4 From Backprop to Deep Learning 8
1.5 An Overview of Chapters 9

2 Linear Associative Networks 11

2.1 Introduction 11
2.2 Setting One Connection Weight 12
2.3 Learning One Association 14
2.4 Gradient Descent 16
2.5 Learning Two Associations 18
2.6 Learning Many Associations 23
2.7 Learning Photographs 25
2.8 Summary 26

3 Perceptrons 27

3.1 Introduction 27
3.2 The Perceptron Learning Algorithm 28
3.3 The Exclusive OR Problem 32
3.4 Why Exclusive OR Matters 34
3.5 Summary 36

4 The Backpropagation Algorithm 37

4.1 Introduction 37
4.2 The Backpropagation Algorithm 39
4.3 Why Use Sigmoidal Hidden Units? 48
4.4 Generalisation and Over-fitting 49
4.5 Vanishing Gradients 52
4.6 Speeding Up Backprop 55
4.7 Local and Global Mimima 57
4.8 Temporal Backprop 59
4.9 Early Backprop Achievements 62
4.10 Summary 62

5 Hopfield Nets 63

5.1 Introduction . . . 63
5.2 The Hopfield Net . . . 64
5.3 Learning One Network State . . . 65
5.4 Content Addressable Memory . . . 66
5.5 Tolerance to Damage . . . 68
5.6 The Energy Function . . . 68
5.7 Summary . . . 70

6 Boltzmann Machines 71

6.1 Introduction . . . 71
6.2 Learning in Generative Models . . . 72
6.3 The Boltzmann Machine Energy Function . . . 74
6.4 Simulated Annealing . . . 76
6.5 Learning by Sculpting Distributions . . . 77
6.6 Learning in Boltzmann Machines . . . 78
6.7 Learning by Maximising Likelihood . . . 80
6.8 Autoencoder Networks . . . 84
6.9 Summary . . . 84

7 Deep RBMs 85

7.1 Introduction . . . 85
7.2 Restricted Boltzmann Machines . . . 87
7.3 Training Restricted Boltzmann Machines . . . 87
7.4 Deep Autoencoder Networks . . . 93
7.5 Summary . . . 96

8 Variational Autoencoders 97

8.1 Introduction . . . 97
8.2 Why Favour Independent Features? . . . 98
8.3 Overview of Variational Autoencoders . . . 100
8.4 Latent Variables and Manifolds . . . 103
8.5 Key Quantities . . . 104
8.6 How Variational Autoencoders Work . . . 106
8.7 The Evidence Lower Bound . . . 110
8.8 An Alternative Derivation . . . 115
8.9 Maximising the Lower Bound . . . 115
8.10 Conditional Variational Autoencoders . . . 119
8.11 Applications . . . 119
8.12 Summary . . . 120

9 Deep Backprop Networks 121

9.1 Introduction . . . 121
9.2 Convolutional Neural Networks . . . 122
9.3 LeNet1 . . . 125
9.4 LeNet5 . . . 127
9.5 AlexNet . . . 129
9.6 GoogLeNet . . . 130
9.7 ResNet . . . 130
9.8 Ladder Autoencoder Networks . . . 132
9.9 Denoising Autoencoders . . . 135
9.10 Fooling Neural Networks . . . 136
9.11 Generative Adversarial Networks . . . 137
9.12 Temporal Deep Neural Networks . . . 140
9.13 Capsule Networks . . . 141
9.14 Summary . . . 142

10 Reinforcement Learning 143

10.1 Introduction . . . 143
10.2 What's the Problem? . . . 146
10.3 Key Quantities . . . 147
10.4 Markov Decision Processes . . . 148
10.5 Formalising the Problem . . . 149
10.6 The Bellman Equation . . . 150
10.7 Learning State-Value Functions . . . 152
10.8 Eligibility Traces . . . 155
10.9 Learning Action-Value Functions . . . 157
10.10 Balancing a Pole . . . 164
10.11 Applications . . . 166
10.12 Summary . . . 168

11 The Emperor's New AI? 169

11.1 Artificial Intelligence . . . 169
11.2 Yet Another Revolution? . . . 170

Further Reading 173

Appendices

A Glossary 175
B Mathematical Symbols 181
C A Vector and Matrix Tutorial 183
D Maximum Likelihood Estimation 187
E Bayes' Theorem 189
References 191
Index 199