

Table of Contents

Preface	v
Chapter 1: Unsupervised Machine Learning	1
Principal component analysis	2
PCA – a primer	2
Employing PCA	4
Introducing k-means clustering	7
Clustering – a primer	8
Kick-starting clustering analysis	8
Tuning your clustering configurations	13
Self-organizing maps	18
SOM – a primer	18
Employing SOM	20
Further reading	24
Summary	25
Chapter 2: Deep Belief Networks	27
Neural networks – a primer	28
The composition of a neural network	28
Network topologies	29
Restricted Boltzmann Machine	33
Introducing the RBM	33
Topology	34
Training	35
Applications of the RBM	37
Further applications of the RBM	49
Deep belief networks	49
Training a DBN	50
Applying the DBN	50
Validating the DBN	54

Further reading	55
Summary	56
Chapter 3: Stacked Denoising Autoencoders	57
Autoencoders	57
Introducing the autoencoder	58
Topology	58
Training	59
Denoising autoencoders	60
Applying a dA	62
Stacked Denoising Autoencoders	66
Applying the SdA	67
Assessing SdA performance	74
Further reading	75
Summary	75
Chapter 4: Convolutional Neural Networks	77
Introducing the CNN	77
Understanding the convnet topology	79
Understanding convolution layers	81
Understanding pooling layers	85
Training a convnet	88
Putting it all together	88
Applying a CNN	92
Further Reading	99
Summary	100
Chapter 5: Semi-Supervised Learning	101
Introduction	101
Understanding semi-supervised learning	102
Semi-supervised algorithms in action	103
Self-training	103
Implementing self-training	105
Finessing your self-training implementation	110
Contrastive Pessimistic Likelihood Estimation	114
Further reading	126
Summary	127
Chapter 6: Text Feature Engineering	129
Introduction	129
Text feature engineering	130
Cleaning text data	131
Text cleaning with BeautifulSoup	131
Managing punctuation and tokenizing	132
Tagging and categorising words	136

Creating features from text data	141
Stemming	141
Bagging and random forests	143
Testing our prepared data	146
Further reading	153
Summary	154
Chapter 7: Feature Engineering Part II	155
Introduction	155
Creating a feature set	156
Engineering features for ML applications	157
Using rescaling techniques to improve the learnability of features	157
Creating effective derived variables	160
Reinterpreting non-numeric features	162
Using feature selection techniques	165
Performing feature selection	167
Feature engineering in practice	175
Acquiring data via RESTful APIs	176
Testing the performance of our model	177
Twitter	180
Deriving and selecting variables using feature engineering techniques	187
Further reading	199
Summary	200
Chapter 8: Ensemble Methods	201
Introducing ensembles	202
Understanding averaging ensembles	203
Using bagging algorithms	203
Using random forests	205
Applying boosting methods	209
Using XGBoost	212
Using stacking ensembles	215
Applying ensembles in practice	218
Using models in dynamic applications	221
Understanding model robustness	222
Identifying modeling risk factors	228
Strategies to managing model robustness	230
Further reading	233
Summary	234
Chapter 9: Additional Python Machine Learning Tools	235
Alternative development tools	236
Introduction to Lasagne	236
Getting to know Lasagne	236

Introduction to TensorFlow	239
Getting to know TensorFlow	239
Using TensorFlow to iteratively improve our models	241
Knowing when to use these libraries	244
Further reading	245
Summary	245
Appendix: Chapter Code Requirements	249
Index	251
