

Table of Contents

Preface	xi
Chapter 1: Giving Computers the Ability to Learn from Data	1
Building intelligent machines to transform data into knowledge	2
The three different types of machine learning	2
Making predictions about the future with supervised learning	3
Classification for predicting class labels	3
Regression for predicting continuous outcomes	5
Solving interactive problems with reinforcement learning	6
Discovering hidden structures with unsupervised learning	7
Finding subgroups with clustering	7
Dimensionality reduction for data compression	8
Introduction to the basic terminology and notations	8
A roadmap for building machine learning systems	11
Preprocessing – getting data into shape	12
Training and selecting a predictive model	12
Evaluating models and predicting unseen data instances	13
Using Python for machine learning	13
Installing Python and packages from the Python Package Index	14
Using the Anaconda Python distribution and package manager	14
Packages for scientific computing, data science, and machine learning	15
Summary	15
Chapter 2: Training Simple Machine Learning Algorithms for Classification	17
Artificial neurons – a brief glimpse into the early history of machine learning	18
The formal definition of an artificial neuron	19
The perceptron learning rule	21

Implementing a perceptron learning algorithm in Python	24
An object-oriented perceptron API	24
Training a perceptron model on the Iris dataset	28
Adaptive linear neurons and the convergence of learning	34
Minimizing cost functions with gradient descent	35
Implementing Adaline in Python	38
Improving gradient descent through feature scaling	42
Large-scale machine learning and stochastic gradient descent	44
Summary	50
Chapter 3: A Tour of Machine Learning Classifiers	
Using scikit-learn	51
Choosing a classification algorithm	52
First steps with scikit-learn – training a perceptron	52
Modeling class probabilities via logistic regression	59
Logistic regression intuition and conditional probabilities	59
Learning the weights of the logistic cost function	63
Converting an Adaline implementation into an algorithm for logistic regression	66
Training a logistic regression model with scikit-learn	71
Tackling overfitting via regularization	73
Maximum margin classification with support vector machines	76
Maximum margin intuition	77
Dealing with a nonlinearly separable case using slack variables	79
Alternative implementations in scikit-learn	81
Solving nonlinear problems using a kernel SVM	82
Kernel methods for linearly inseparable data	82
Using the kernel trick to find separating hyperplanes in high-dimensional space	84
Decision tree learning	88
Maximizing information gain – getting the most bang for your buck	90
Building a decision tree	95
Combining multiple decision trees via random forests	98
K-nearest neighbors – a lazy learning algorithm	101
Summary	105
Chapter 4: Building Good Training Sets – Data Preprocessing	107
Dealing with missing data	107
Identifying missing values in tabular data	108
Eliminating samples or features with missing values	109
Imputing missing values	110
Understanding the scikit-learn estimator API	111

Handling categorical data	112
Nominal and ordinal features	113
Creating an example dataset	113
Mapping ordinal features	113
Encoding class labels	114
Performing one-hot encoding on nominal features	116
Partitioning a dataset into separate training and test sets	118
Bringing features onto the same scale	120
Selecting meaningful features	123
L1 and L2 regularization as penalties against model complexity	124
A geometric interpretation of L2 regularization	124
Sparse solutions with L1 regularization	126
Sequential feature selection algorithms	130
Assessing feature importance with random forests	136
Summary	139
Chapter 5: Compressing Data via Dimensionality Reduction	141
Unsupervised dimensionality reduction via principal component analysis	142
The main steps behind principal component analysis	142
Extracting the principal components step by step	144
Total and explained variance	147
Feature transformation	148
Principal component analysis in scikit-learn	151
Supervised data compression via linear discriminant analysis	155
Principal component analysis versus linear discriminant analysis	155
The inner workings of linear discriminant analysis	156
Computing the scatter matrices	157
Selecting linear discriminants for the new feature subspace	160
Projecting samples onto the new feature space	162
LDA via scikit-learn	163
Using kernel principal component analysis for nonlinear mappings	165
Kernel functions and the kernel trick	166
Implementing a kernel principal component analysis in Python	172
Example 1 – separating half-moon shapes	173
Example 2 – separating concentric circles	176
Projecting new data points	179
Kernel principal component analysis in scikit-learn	183
Summary	184

Chapter 6: Learning Best Practices for Model Evaluation and Hyperparameter Tuning	185
Streamlining workflows with pipelines	185
Loading the Breast Cancer Wisconsin dataset	186
Combining transformers and estimators in a pipeline	187
Using k-fold cross-validation to assess model performance	189
The holdout method	190
K-fold cross-validation	191
Debugging algorithms with learning and validation curves	195
Diagnosing bias and variance problems with learning curves	196
Addressing over- and underfitting with validation curves	199
Fine-tuning machine learning models via grid search	201
Tuning hyperparameters via grid search	201
Algorithm selection with nested cross-validation	203
Looking at different performance evaluation metrics	205
Reading a confusion matrix	206
Optimizing the precision and recall of a classification model	207
Plotting a receiver operating characteristic	210
Scoring metrics for multiclass classification	213
Dealing with class imbalance	214
Summary	216
Chapter 7: Combining Different Models for Ensemble Learning	219
Learning with ensembles	219
Combining classifiers via majority vote	224
Implementing a simple majority vote classifier	224
Using the majority voting principle to make predictions	231
Evaluating and tuning the ensemble classifier	234
Bagging – building an ensemble of classifiers from bootstrap samples	240
Bagging in a nutshell	240
Applying bagging to classify samples in the Wine dataset	242
Leveraging weak learners via adaptive boosting	246
How boosting works	246
Applying AdaBoost using scikit-learn	251
Summary	254
Chapter 8: Applying Machine Learning to Sentiment Analysis	255
Preparing the IMDb movie review data for text processing	256
Obtaining the movie review dataset	256
Preprocessing the movie dataset into more convenient format	257

Introducing the bag-of-words model	259
Transforming words into feature vectors	259
Assessing word relevancy via term frequency-inverse document frequency	261
Cleaning text data	264
Processing documents into tokens	266
Training a logistic regression model for document classification	268
Working with bigger data – online algorithms and out-of-core learning	270
Topic modeling with Latent Dirichlet Allocation	274
Decomposing text documents with LDA	275
LDA with scikit-learn	275
Summary	279
Chapter 9: Embedding a Machine Learning Model into a Web Application	281
Serializing fitted scikit-learn estimators	282
Setting up an SQLite database for data storage	285
Developing a web application with Flask	287
Our first Flask web application	288
Form validation and rendering	290
Setting up the directory structure	291
Implementing a macro using the Jinja2 templating engine	292
Adding style via CSS	293
Creating the result page	294
Turning the movie review classifier into a web application	294
Files and folders – looking at the directory tree	296
Implementing the main application as app.py	298
Setting up the review form	300
Creating a results page template	302
Deploying the web application to a public server	304
Creating a PythonAnywhere account	304
Uploading the movie classifier application	305
Updating the movie classifier	306
Summary	308
Chapter 10: Predicting Continuous Target Variables with Regression Analysis	309
Introducing linear regression	310
Simple linear regression	310
Multiple linear regression	311
Exploring the Housing dataset	312
Loading the Housing dataset into a data frame	313

Visualizing the important characteristics of a dataset	314
Looking at relationships using a correlation matrix	316
Implementing an ordinary least squares linear regression model	319
Solving regression for regression parameters with gradient descent	319
Estimating coefficient of a regression model via scikit-learn	324
Fitting a robust regression model using RANSAC	325
Evaluating the performance of linear regression models	328
Using regularized methods for regression	332
Turning a linear regression model into a curve – polynomial regression	334
Adding polynomial terms using scikit-learn	334
Modeling nonlinear relationships in the Housing dataset	336
Dealing with nonlinear relationships using random forests	339
Decision tree regression	340
Random forest regression	342
Summary	345
Chapter 11: Working with Unlabeled Data – Clustering Analysis	347
Grouping objects by similarity using k-means	348
K-means clustering using scikit-learn	348
A smarter way of placing the initial cluster centroids using k-means++	353
Hard versus soft clustering	354
Using the elbow method to find the optimal number of clusters	357
Quantifying the quality of clustering via silhouette plots	358
Organizing clusters as a hierarchical tree	363
Grouping clusters in bottom-up fashion	364
Performing hierarchical clustering on a distance matrix	365
Attaching dendrograms to a heat map	369
Applying agglomerative clustering via scikit-learn	371
Locating regions of high density via DBSCAN	372
Summary	378
Chapter 12: Implementing a Multilayer Artificial Neural Network from Scratch	379
Modeling complex functions with artificial neural networks	380
Single-layer neural network recap	382
Introducing the multilayer neural network architecture	384
Activating a neural network via forward propagation	387
Classifying handwritten digits	389
Obtaining the MNIST dataset	390
Implementing a multilayer perceptron	396

Training an artificial neural network	407
Computing the logistic cost function	408
Developing your intuition for backpropagation	411
Training neural networks via backpropagation	412
About the convergence in neural networks	417
A few last words about the neural network implementation	418
Summary	419
Chapter 13: Parallelizing Neural Network Training with TensorFlow	421
TensorFlow and training performance	421
What is TensorFlow?	423
How we will learn TensorFlow	424
First steps with TensorFlow	424
Working with array structures	427
Developing a simple model with the low-level TensorFlow API	428
Training neural networks efficiently with high-level TensorFlow APIs	433
Building multilayer neural networks using TensorFlow's Layers API	434
Developing a multilayer neural network with Keras	438
Choosing activation functions for multilayer networks	443
Logistic function recap	444
Estimating class probabilities in multiclass classification via the softmax function	446
Broadening the output spectrum using a hyperbolic tangent	447
Rectified linear unit activation	449
Summary	451
Chapter 14: Going Deeper – The Mechanics of TensorFlow	453
Key features of TensorFlow	454
TensorFlow ranks and tensors	454
How to get the rank and shape of a tensor	455
Understanding TensorFlow's computation graphs	456
Placeholders in TensorFlow	459
Defining placeholders	459
Feeding placeholders with data	460
Defining placeholders for data arrays with varying batchsizes	461
Variables in TensorFlow	462
Defining variables	463
Initializing variables	465
Variable scope	466
Reusing variables	468

Building a regression model	471
Executing objects in a TensorFlow graph using their names	475
Saving and restoring a model in TensorFlow	476
Transforming Tensors as multidimensional data arrays	479
Utilizing control flow mechanics in building graphs	483
Visualizing the graph with TensorBoard	487
Extending your TensorBoard experience	490
Summary	491
Chapter 15: Classifying Images with Deep Convolutional Neural Networks	493
Building blocks of convolutional neural networks	494
Understanding CNNs and learning feature hierarchies	494
Performing discrete convolutions	496
Performing a discrete convolution in one dimension	496
The effect of zero-padding in a convolution	499
Determining the size of the convolution output	501
Performing a discrete convolution in 2D	502
Subsampling	506
Putting everything together to build a CNN	508
Working with multiple input or color channels	508
Regularizing a neural network with dropout	512
Implementing a deep convolutional neural network using TensorFlow	514
The multilayer CNN architecture	514
Loading and preprocessing the data	516
Implementing a CNN in the TensorFlow low-level API	517
Implementing a CNN in the TensorFlow Layers API	530
Summary	536
Chapter 16: Modeling Sequential Data Using Recurrent Neural Networks	537
Introducing sequential data	538
Modeling sequential data – order matters	538
Representing sequences	539
The different categories of sequence modeling	540
RNNs for modeling sequences	541
Understanding the structure and flow of an RNN	541
Computing activations in an RNN	543
The challenges of learning long-range interactions	546
LSTM units	548

Implementing a multilayer RNN for sequence modeling in TensorFlow	550
Project one – performing sentiment analysis of IMDb movie reviews using multilayer RNNs	551
Preparing the data	552
Embedding	556
Building an RNN model	558
The SentimentRNN class constructor	559
The build method	560
Step 1 – defining multilayer RNN cells	562
Step 2 – defining the initial states for the RNN cells	562
Step 3 – creating the RNN using the RNN cells and their states	563
The train method	563
The predict method	565
Instantiating the SentimentRNN class	565
Training and optimizing the sentiment analysis RNN model	566
Project two – implementing an RNN for character-level language modeling in TensorFlow	567
Preparing the data	568
Building a character-level RNN model	572
The constructor	573
The build method	574
The train method	576
The sample method	578
Creating and training the CharRNN Model	579
The CharRNN model in the sampling mode	580
Chapter and book summary	580
Index	583