Contents

List of Figures xxiii
List of Tables xlvii
Foreword by Irad Ben-Gal liii
Foreword by Ravi Shankar lv
Preface lvii
Acknowledgments lix
1. Overview of Machine Learning 1
1.1 Introduction 1
1.2 Applications of Machine Learning 2
1.3 History of Machine Learning 7
1.3.1 Foundational Phase of Machine Learning (Before 1940) 8
1.3.2 Transitional Phase of Machine Learning (1940–2000) 9
1.3.3 Contemporary Phase of Machine Learning (Post 2000) 10
1.4 Definition of Machine Learning 11
1.5 Traditional Programming vs. Machine Learning 13
1.6 Classification of Machine Learning Algorithms 15
1.6.1 Supervised Learning 16
1.6.2 Unsupervised Learning 23
1.7 Supervised and Unsupervised Learning in a Real-life Scenario 26
1.8 Reinforcement Learning 27
1.8.1 Working Principle of Reinforcement Learning 28
1.8.2 Components of Reinforcement Learning 29
1.8.3 Reward Maximization and Discount 31

1.8.4 Exploration and Exploitation 32
1.8.5 Timing of Execution of Various Steps 33
1.8.6 Markov's Decision Process 34
1.8.7 Understanding Q-Learning 35
1.9 Data Science vs. Data Mining vs. Machine Learning vs. Deep Learning vs. Artificial Intelligence 44
REMIND ME 48
ADDITIONAL RESOURCES 49
BOOKS 49
WEBLINKS TO VIDEOS 50
WEBLINKS TO BLOGS 50
TEST ME (QUESTIONS) 51

2. Introduction to Python 53
2.1 Features of Python 54
2.2 History of Python 55
2.3 Installation of Python 56
2.4 My First Python Program 57
2.4.1 Interactive Mode Programming 58
2.4.2 Script Mode Programming 59
2.4.3 Running Python Script through IDE 60
2.5 Identifiers in Python 63
2.6 Keywords in Python 63
2.7 Indentation in Python 64
2.8 Comments in Python 65
2.9 Variables in Python 65
2.9.1 Declaring Variables and Assignment of Values to the Variables 66
2.10 Data Types in Python 67
2.11 Python Numbers 68
2.12 Python Strings 69
2.13 Python List 73
2.13.1 Main Features of the List Data Type 74
2.13.2 Basic Operations on List in Python 74
2.14 Python Tuples 81
2.14.1 Features of Tuples 82
2.14.2 Operations over the Tuple Data Type 83
2.14.3 Comparison between Lists and Tuples 87

2.15 Python Dictionary 88
2.15.1 Operations over the Dictionary Data Type 89
2.16 Python Set 96
2.16.1 Operations over the Set Data Type 96
2.17 Comparison between List, Tuple, Dictionary, and Set Data Types 101
2.18 If Statement 102
2.19 Looping Statements 105
2.19.1 `While` Loop 105
2.19.2 `For` Loop 107
2.20 Python Functions 110
REMIND ME 115
ADDITIONAL RESOURCES 116
BOOKS 116
WEBLINKS TO VIDEOS 116
WEBLINKS TO BLOGS 118
TEST ME (QUESTIONS) 118

3. Data Pre-processing 125
3.1 Need for Data Pre-processing 125
3.2 Data Pre-processing Techniques 129
3.2.1 Data Cleaning 130
3.2.2 Data Integration 137
3.2.3 Data Transformation 138
3.2.4 Data Reduction 142
3.3 Handling of Categorical Data 148
REMIND ME 151
ADDITIONAL RESOURCES 152
BOOKS 152
WEBLINKS TO BLOGS 152
TEST ME (QUESTIONS) 153

4. Implementing Data Pre-processing in Python 155
4.1 Step 1: Importing Libraries 156
4.2 Step 2: Loading the Dataset 159
4.3 Step 3: Handling of Missing Values 166
4.4 Step 4: Handling of Categorical Data 168
4.4.1 One-hot Encoding 170

4.5 Step 5: Splitting of Dataset into Training and Testing 175
4.6 Step 6: Feature Scaling 177
4.6.1 Normalization 179
4.6.2 Standardization 181
REMIND ME 182
ADDITIONAL RESOURCES 183
WEBLINKS TO VIDEOS 183
WEBLINKS TO BLOGS 183
TEST ME (QUESTIONS) 184
TEST ME (PRACTICAL SKILLS) 186
5. Simple Linear Regression 187
5.1 Introduction to Simple Linear Regression 187
5.2 Hypothesis or Function 194
5.3 Parameters of Simple Linear Regression 195
5.4 Cost Function and Its Minimization 198
5.4.1 Minimization of Cost Function with One Parameter: $\min_{\phi} J(\phi)$ 199
5.4.2 Minimization of the Cost Function with Both Parameters 203
5.5 Gradient Descent Algorithm 208
5.6 Mathematical Understanding of Gradient Descent Algorithm 213
5.6.1 Role of Learning Rate α 224
5.6.2 Calculating the Values of Parameters by Gradient Descent Algorithm 229
REMIND ME 235
ADDITIONAL RESOURCES 236
WEBLINKS TO VIDEOS 236
WEBLINKS TO BLOGS 238
TEST ME (QUESTIONS) 238
6. Implementing Simple Linear Regression 241
6.1 Step 1: Loading Dataset 243
6.2 Step 2: Splitting the Dataset into Training and Testing Datasets 246
6.3 Step 3: Building the Model and Making Predictions 249
6.4 Step 4: Visualizing the Training Set Results 251
6.5 Step 5: Visualizing the Testing Set Results 255
6.6 Step 6: Evaluating the Model Using Regression Metrics 257
REMIND ME 264
ADDITIONAL RESOURCES 264
WEBLINKS TO VIDEOS 264

WEBLINKS TO BLOGS 265
TEST ME (PRACTICAL SKILLS) 265
7. Multiple Linear Regression and Polynomial Linear Regression 267
7.1 Introduction to Multiple Linear Regression 267
7.2 Handling of Nominal Variable and Concept of Dummy Variables 270
7.3 Building the Multiple Regression Model 275
7.4 Concept and Need of P-value 278
7.5 Calculating P-value 280
7.6 Using P-value and Defining Null Hypothesis for Building Multiple Linear Regression Model 292
7.7 Variable Selection Methods 293
7.7.1 All-in (Knowledge-based Method) 293
7.7.2 Backward Elimination Method 293
7.7.3 Forward Selection Method 295
7.7.4 Bi-directional Elimination 296
7.7.5 All Possible Models 297
7.7.6 Choosing the Most Preferred Variable Selection Method 297
7.8 Importance of Adjusted R-squared Metric in Multiple Linear Regression 298
7.9 Polynomial Linear Regression 299
7.9.1 Case Study of Rank–Salary 301
7.9.2 Case Study of Chemistry Students 306
7.9.3 Polynomial Linear Regression in Case of Multiple Input Attributes 308
7.9.4 Exploring the Versatility of Polynomial Regression in Modeling Various Data Shapes 308
REMIND ME 310
ADDITIONAL RESOURCES 311
WEBLINKS TO VIDEOS 311
WEBLINKS TO BLOGS 311
TEST ME (QUESTIONS) 311
8. Implementation of Multiple Linear Regression and Polynomial Linear Regression 313
8.1 Introduction to Multiple Linear Regression 313
8.1.1 Understanding Dataset and Problem Statement 314
8.2 Polynomial Linear Regression 327
8.2.1 Case Study of Rank Salary 327
8.3 What Will Happen if You Increase the Degree Value Too Much? 342
8.4 Comparison of Simple Linear Regression and Polynomial Regression Results 344

8.5 Case Study of Chemistry Lab Experiment 348
REMIND ME 355
ADDITIONAL RESOURCES 356
WEBLINKS TO VIDEOS 356
WEBLINKS TO BLOGS 356
TEST ME (PRACTICAL SKILLS) 357
9. Classification 359
9.1 Introduction to Classification 359
9.2 Types of Classification 361
9.2.1 Posteriori Classification 362
9.2.2 Priori Classification 362
9.3 Input and Output Attributes 362
9.4 Working of Classification 363
9.5 Guidelines for Size and Quality of the Training Dataset 367
9.6 Lazy and Eager Learner Classifiers 367
9.7 Introduction to Decision Tree Classifier 367
9.7.1 Building Decision Tree 370
9.7.2 Concept of Information Theory 370
9.7.3 Defining Information in Terms of Probability 371
9.7.4 Rule for Selection of Root Node and Other Nodes 373
9.7.5 Information Gain 374
9.7.6 Building Decision Tree Using Information Gain 375
9.7.7 Drawbacks of Information Gain Theory 408
9.7.8 Split Algorithm Based on Gini Index 408
9.7.9 Building Decision Tree Using Gini Index 412
9.7.10 Advantages of Decision Tree 445
9.7.11 Disadvantages of Decision Tree 445
9.8 Random Forest Classification 445
9.9 Naive Bayes Method 448
9.9.1 Bayes' Theorem in Action 449
9.9.2 Using Bayes' Theorem for Classification 451
9.9.3 Working of Naive Bayes Classifier Using Laplace Estimator 466
9.10 k-Nearest Neighbor Algorithm 467
9.11 Logistic Regression 470
9.12 Understanding Metrics to Know the Quality of Classifier 478
9.12.1 True Positive 479

9.12.2 True Negative 479
9.12.3 False Positive 480
9.12.4 False Negative 480
9.12.5 Confusion Matrix 480
9.12.6 Precision 481
9.12.7 Recall 482
9.12.8 F-Measure 483
9.12.9 Evaluating a Confusion Matrix Have More Than Two Classes 484
REMIND ME 486
ADDITIONAL RESOURCES 487
BOOKS 487
WEBLINKS TO VIDEOS 487
WEBLINKS TO BLOGS 488
TEST ME (QUESTIONS) 488
10. Support Vector Machine Classifier 495
10.1 Support Vector Machines 495
10.1.1 SVM Working Principle 495
10.2 Rules for Identification of Correct Hyperplane 499
10.3 Some Important Concepts and Terminologies Used in SVM 503
10.3.1 Support Vectors 504
10.3.2 Maximized Margin 504
10.3.3 Positive and Negative Hyperplane 505
10.4 Types of SVM Classifiers 506
10.4.1 Linear SVM Classifier 506
10.4.2 Non-linear SVM Classifier 507
10.5 Deep Dive into Non-linear SVM Classifier 507
10.5.1 Case 1: When Non-linear Data Is in One-Dimension 509
10.5.2 Case 2: When Non-linear Data Is Two-Dimensional 513
10.6 Defining Hyperplane 514
10.7 Limitation of Mapping Data Points to Higher Dimensional Space and Projecting Them Back to Lower Dimensional Space 515
10.8 SVM Kernel 515
10.8.1 Radial Basis Function (RBF) or Gaussian Kernel Function 516
10.8.2 Selecting the Appropriate Value of the Hyperparameter (σ) 519
10.8.3 Understanding RBF Kernel Function from a 3D Perspective 521
10.8.4 Applying Multiple Kernels to a Dataset 525

10.9 Tuning Parameters of SVM 526
10.9.1 Gamma 526
10.9.2 Regularization 527
REMIND ME 528
ADDITIONAL RESOURCES 530
WEBLINKS TO VIDEOS 530
WEBLINKS TO BLOGS 530
TEST ME (QUESTIONS) 531
11. Implementation of Classification Algorithms 533
11.1 Introduction to Classification Algorithms and Steps for Its Implementation 533
11.2 Preparing the Pre-processing Data Template (Step 1 to Step 4)—Common to All the Classifier Models 535
11.3 Implementation of Decision Tree Classifier 544
11.4 Implementation of Random Forest Classifier 554
11.5 Implementation of Naive Bayes Classifier 561
11.6 Implementation of k-NN 567
11.7 Implementation of Logistic Regression Classifier 574
11.8 Implementation of Support Vector Machine Linear Model 581
11.9 Implementation of Non-linear Support Vector Machine Model 587
11.10 Comparative Analysis of Different Classifiers 592
11.11 Text Classification 594
REMIND ME 604
ADDITIONAL RESOURCES 606
WEBLINKS TO VIDEOS 606
WEBLINKS TO BLOGS 607
TEST ME (QUESTIONS) 607
TEST ME (PRACTICAL SKILLS) 609
12. Clustering 611
12.1 Introduction to Clustering 611
12.2 Applications of Clustering 613
12.3 Features of an Ideal Clustering Algorithm 614
12.4 Distance Metrics 615
12.4.1 Euclidean Distance 615
12.4.2 Manhattan Distance 618
12.4.3 Minkowski Distance 619
12.4.4 Chebyshev Distance 620

12.5 Major Clustering Methods/Algorithms 621
12.6 Partitioning Clustering 622
12.6.1 k-means Clustering 622
12.6.2 Finding the Optimal Number of Clusters 646
12.6.3 Issues of k-means Algorithm 647
12.7 Hierarchical Clustering Algorithms (HCA) 648
12.7.1 Agglomerative Clustering 649
12.7.2 Divisive Clustering 675
12.7.3 Density-based Clustering 681
REMIND ME 692
ADDITIONAL RESOURCES 693
BOOKS 693
WEBLINKS TO VIDEOS 693
WEBLINKS TO BLOGS 694
TEST ME (QUESTIONS) 695
13. Implementation of Clustering 699
13.1 Implementation of k-means Clustering and Hierarchical Clustering 699
13.2 Case Study I: Clustering the Customers of a Mall 700
13.2.1 Step 1: Importing Libraries and the Dataset 702
13.2.2 Step 2: Find the Optimal Number of Clusters 704
13.2.3 Step 3: Fitting k-means to the Dataset 707
13.2.4 Step 4: Visualizing the Clusters 709
13.2.5 Step 5: Creating a Dendrogram and Finding the Optimal Number of Clusters 712
13.2.6 Step 6: Fitting the Hierarchical Clustering Model over the Dataset 714
13.2.7 Step 7: Visualizing the Different Clusters 716
13.3 Case Study II: U.S. Arrests Dataset 719
13.3.1 Step 1: Importing the Libraries and Dataset 720
13.3.2 Step 2: Find the Optimal Number of Clusters 721
13.3.3 Step 3: Fitting k-means to the Dataset 722
13.3.4 Step 4: Visualizing the Clusters 724
13.3.5 Step 5: Creating a Dendrogram and Finding the Optimal Number of Clusters 726
13.3.6 Step 6: Fitting the Hierarchical Clustering Model over the Dataset 727
13.3.7 Step 7: Visualizing the Different Clusters 728
13.4 Case Study III: Iris Dataset 729
13.4.1 Step 1: Importing the Libraries and Dataset 730
13.4.2 Step 2: Find the Optimal Number of Clusters 731

13.4.3 Step 3: Fitting k-means to the Dataset 732
13.4.4 Step 4: Visualizing the Clusters 734
REMIND ME 736
ADDITIONAL RESOURCES 737
WEBLINKS TO VIDEOS 737
WEBLINKS TO BLOGS 737
TEST ME (QUESTIONS) 737
14. Association Mining 739
14.1 Introduction to Association Rule Mining 739
14.2 Defining Association Rule Mining 743
14.3 Representations of Items for Association Mining 744
14.4 Metrics for Evaluating the Strength of Association Rules 745
14.4.1 Support 745
14.4.2 Confidence 747
14.4.3 Lift 749
14.5 Naive Algorithm for Finding Association Rules 753
14.5.1 Working of Naive Algorithm 753
14.5.2 Limitations of Naive Algorithm 755
14.5.3 Improved Naive Algorithm to Deal with a Larger Dataset 756
14.6 Approaches for Transaction Database Storage 757
14.6.1 Simple Transaction Storage 758
14.6.2 Horizontal Storage 758
14.6.3 Vertical Representation 759
14.7 Apriori Algorithm 760
14.7.1 A Brief Discussion about the Inventors 760
14.7.2 Working of the Apriori Algorithm 762
14.7.3 Apriori Property 781
14.8 Closed and Maximal Itemsets 791
14.9 Limitations of Apriori Algorithm 795
14.10 Apriori–TID Algorithm for Generating Association Mining Rules 795
14.10.1 Advantages of the Apriori–TID Algorithm 799
REMIND ME 800
ADDITIONAL RESOURCES 801
BOOKS 801
WEBLINKS TO VIDEOS 801

WEBLINKS TO BLOGS 802
TEST ME (QUESTIONS) 802
15. Implementation of Association Mining 807
15.1 Building Association Mining Model 807
15.2 Step 1: Importing Relevant Libraries and Loading the Dataset 810
15.3 Step 2: Making Transactions 812
15.4 Step 3: Building the Model 814
15.5 Step 4: Viewing and Interpreting the Mined Rules 816
REMIND ME 819
ADDITIONAL RESOURCES 819
WEBLINK TO VIDEO 819
WEBLINK TO BLOG 819
TEST ME (QUESTIONS) 820
16. Artificial Neural Network 821
16.1 Introduction to Artificial Neural Network 821
16.2 Human Brain: An Inspiration for the Development of Neural Network 824
16.2.1 Neuron 825
16.3 Mapping of Human Brain Neurons to Artificial Neural Network 827
16.4 Working of Artificial Neural Network 829
16.5 How Does ANN Processes Multiple Input Values? 831
16.6 Role of Weights in Building ANN 833
16.7 Three-Step Learning Process of a Neuron 835
16.7.1 Commonly Used Activation Functions 837
16.8 Training of Neural Network 842
16.9 Finding Optimized Value of Weights for Minimizing the Cost Function 852
16.9.1 Brute-force Approach 852
16.9.2 Gradient Descent Approach 853
16.9.3 Stochastic Gradient Descent Algorithm to Find Global Minima 856
16.9.4 Mini-batch Method 858
16.10 Final Step-by-Step Procedure for Training the Neural Network 859
REMIND ME 860
ADDITIONAL RESOURCES 861
WEBLINKS TO VIDEOS 861
WEBLINKS TO BLOGS 862
TEST ME (QUESTIONS) 862

17. Implementation of the ANN 865
17.1 Building Artificial Neural Network for Cancer Detection 865
17.2 Importing the Dataset and Pre-processing 866
17.2.1 Step 1: Importing the Libraries 866
17.2.2 Step 2: Creating Dependent and Independent Variables 867
17.2.3 Step 3: Encoding Categorical Data 869
17.2.4 Step 4: Splitting the Dataset into Training and Testing 870
17.2.5 Step 5: Feature Scaling 871
17.3 Building the Artificial Neural Network 871
17.3.1 Step 6: Importing the ANN Libraries 871
17.3.2 Step 7: Initializing the ANN 872
17.3.3 Step 8: Adding the Input Layer and the First Hidden Layer 872
17.3.4 Step 9: Adding Second and Third Hidden Layers 873
17.3.5 Step 10: Adding the Output Layer 874
17.3.6 Step 11: Compiling the ANN 875
17.3.7 Step 12: Fitting the ANN into the Training Set 876
17.4 Predicting the Results 878
17.4.1 Step 13: Predicting the Test Set Results 878
17.4.2 Step 14: Making the Confusion Matrix 880
REMIND ME 884
ADDITIONAL RESOURCES 885
WEBLINKS TO VIDEOS 885
WEBLINKS TO BLOGS 885
TEST ME (ASSIGNMENTS) 886
18. Deep Learning and Convolutional Neural Network 889
18.1 Image Recognition 889
18.2 Identifying Digits with Machine Learning 892
18.2.1 Converting Image to Numbers 893
18.2.2 Issues with This Simple Image Classifier 896
18.2.3 Possible Solutions 897
18.3 How Does Human Brain Identify Images? 900
18.4 Working Principle of Convolutional Neural Network (CNN) 903
18.5 Architecture of a Basic CNN Model 905
18.5.1 Convolution Layer 906
18.5.2 Pooling Layer 915
18.5.3 Flattening 920

18.5.4 Full Connection 922
18.6 Training a CNN Model 923
18.7 Deciding the Optimal Number of Epochs to Train a Neural Network 925
REMIND ME 930
ADDITIONAL RESOURCES 931
WEBLINKS TO VIDEOS 931
WEBLINKS TO BLOGS 931
TEST ME (QUESTIONS) 932
19. Implementation of Convolutional Neural Network 937
19.1 Building Image Classifier with CNN 937
19.2 Dog–Cat Classifier 938
19.2.1 Step1: Importing the `keras` Library and Packages 940
19.2.2 Step 2: Initializing the CNN Model 940
19.2.3 Step 3: Add Convolution Layer 941
19.2.4 Step 4: Pooling 943
19.2.5 Step 5: Flattening 944
19.2.6 Step 6: Full Connection 945
19.2.7 Step 7: Compilation of CNN 946
19.3 Phase 2: Image Augmentation and Fitting the Model 947
19.4 Phase 3: Getting Predictions from Our Model and Model Improvement 952
19.5 Further Improving the Accuracy of the Model by Adding Another Convolutional and Max Pooling Layer 954
19.6 Case Study 2: Handwritten Digit Classifier Using MNIST Dataset 958
REMIND ME 965
ADDITIONAL RESOURCES 966
WEBLINKS TO VIDEOS 966
WEBLINKS TO BLOGS AND DATASETS 967
TEST ME (PRACTICAL SKILLS) 967
20. Recurrent Neural Network 969
20.1 Neural Networks and Human Brain 969
20.2 Extending Artificial Neural Network to Recurrent Neural Network 972
20.3 Limitations of Feed-forward Networks 975
20.4 Working Principle of Recurrent Neural Networks 976
20.5 Understanding the Recurrent Neural Networks by Using a Case Study of a Coaching Center 977
20.6 Mathematical Modeling of RNN 980

20.7 Limitations of Recurrent Neural Networks 983
20.7.1 Vanishing and Exploding Gradients 985
20.8 Long Short-Term Memory (LSTM) 989
20.9 Working Principle of LSTM 991
20.9.1 Internals of an RNN Cell 991
20.9.2 Internals of an LSTM Cell 995
20.10 Applications of RNN and LSTM 1001
REMIND ME 1003
ADDITIONAL RESOURCES 1004
WEBLINKS TO VIDEOS 1004
WEBLINKS TO BLOGS 1005
TEST ME (QUESTIONS) 1005
21. Implementation of Recurrent Neural Network 1007
21.1 Introduction 1007
21.2 Implementation of RNN in Python 1008
21.2.1 Data Pre-processing 1008
21.2.2 Building Architecture of RNN 1022
21.2.3 Making Predictions and Visualizing Results 1029
REMIND ME 1038
ADDITIONAL RESOURCES 1039
WEBLINKS TO VIDEOS 1039
WEBLINKS TO BLOGS 1040
TEST ME (PRACTICAL SKILLS) 1040
22. Genetic Algorithm 1041
22.1 Intuition of Genetic Algorithm 1041
22.2 The Inspiration behind Genetic Algorithm 1042
22.2.1 Natural Selection: Survival of the Fittest 1043
22.2.2 Recombination: Crossover 1044
22.2.3 Mutation: Changes in Genes 1046
22.3 Correlation between Nature and Genetic Algorithm 1049
22.4 Mathematical Representation of Genes and Fitness Theory 1050
22.5 Natural Selection Implementation through Roulette Wheel 1054
22.5.1 Working of Roulette Wheel Used in the Casino Game 1054
22.6 Implementation of Recombining or Crossover in Genetic Algorithm 1064
22.6.1 Single-point Crossover 1065
22.6.2 Double-point Crossover 1068

22.7 Implementation of Mutation in Genetic Algorithm 1072
22.7.1 Mathematical Modeling of Mutation 1072
22.7.2 How to Decide the Number of Genes to Be Mutated? 1074
22.8 Elitism 1075
22.8.1 Implementation of Elitism and Concept of Elitism Ratio 1077
22.9 Advantages and Disadvantages of Genetic Algorithm 1079
22.10 Applications of Genetic Algorithm 1080
REMIND ME 1080
ADDITIONAL RESOURCES 1081
WEBLINKS TO VIDEOS 1081
WEBLINKS TO BLOGS 1082
TEST ME (QUESTIONS) 1082
Index 1085