

Table of Contents

Preface	1
Chapter 1: Journey from Statistics to Machine Learning	7
Statistical terminology for model building and validation	8
Machine learning	8
Major differences between statistical modeling and machine learning	10
Steps in machine learning model development and deployment	11
Statistical fundamentals and terminology for model building and validation	12
Bias versus variance trade-off	32
Train and test data	34
Machine learning terminology for model building and validation	35
Linear regression versus gradient descent	38
Machine learning losses	41
When to stop tuning machine learning models	43
Train, validation, and test data	44
Cross-validation	46
Grid search	46
Machine learning model overview	50
Summary	54
Chapter 2: Parallelism of Statistics and Machine Learning	55
Comparison between regression and machine learning models	55
Compensating factors in machine learning models	57
Assumptions of linear regression	58
Steps applied in linear regression modeling	61
Example of simple linear regression from first principles	61
Example of simple linear regression using the wine quality data	64
Example of multilinear regression - step-by-step methodology of model building	66
Backward and forward selection	69
Machine learning models - ridge and lasso regression	75
Example of ridge regression machine learning	77
Example of lasso regression machine learning model	80
Regularization parameters in linear regression and ridge/lasso regression	82
Summary	82

Chapter 3: Logistic Regression Versus Random Forest	83
Maximum likelihood estimation	83
Logistic regression – introduction and advantages	85
Terminology involved in logistic regression	87
Applying steps in logistic regression modeling	94
Example of logistic regression using German credit data	94
Random forest	111
Example of random forest using German credit data	113
Grid search on random forest	117
Variable importance plot	120
Comparison of logistic regression with random forest	122
Summary	124
Chapter 4: Tree-Based Machine Learning Models	125
Introducing decision tree classifiers	126
Terminology used in decision trees	127
Decision tree working methodology from first principles	128
Comparison between logistic regression and decision trees	134
Comparison of error components across various styles of models	135
Remedial actions to push the model towards the ideal region	136
HR attrition data example	137
Decision tree classifier	140
Tuning class weights in decision tree classifier	143
Bagging classifier	145
Random forest classifier	149
Random forest classifier - grid search	155
AdaBoost classifier	158
Gradient boosting classifier	163
Comparison between AdaBoosting versus gradient boosting	166
Extreme gradient boosting - XGBoost classifier	169
Ensemble of ensembles - model stacking	174
Ensemble of ensembles with different types of classifiers	174
Ensemble of ensembles with bootstrap samples using a single type of classifier	182
Summary	185
Chapter 5: K-Nearest Neighbors and Naive Bayes	187
K-nearest neighbors	188
KNN voter example	188
Curse of dimensionality	189

Curse of dimensionality with 1D, 2D, and 3D example	192
KNN classifier with breast cancer Wisconsin data example	195
Tuning of k-value in KNN classifier	200
Naive Bayes	203
Probability fundamentals	204
Joint probability	205
Understanding Bayes theorem with conditional probability	206
Naive Bayes classification	208
Laplace estimator	209
Naive Bayes SMS spam classification example	210
Summary	220
Chapter 6: Support Vector Machines and Neural Networks	221
Support vector machines working principles	221
Maximum margin classifier	222
Support vector classifier	224
Support vector machines	225
Kernel functions	227
SVM multilabel classifier with letter recognition data example	228
Maximum margin classifier - linear kernel	229
Polynomial kernel	232
RBF kernel	234
Artificial neural networks - ANN	241
Activation functions	244
Forward propagation and backpropagation	245
Optimization of neural networks	254
Stochastic gradient descent - SGD	255
Momentum	256
Nesterov accelerated gradient - NAG	257
Adagrad	258
Adadelta	258
RMSprop	258
Adaptive moment estimation - Adam	259
Limited-memory broyden-fletcher-goldfarb-shanno - L-BFGS optimization algorithm	259
Dropout in neural networks	261
ANN classifier applied on handwritten digits using scikit-learn	262
Introduction to deep learning	268
Solving methodology	270
Deep learning software	271

Deep neural network classifier applied on handwritten digits using Keras	272
Summary	280
Chapter 7: Recommendation Engines	281
<hr/>	
Content-based filtering	281
Cosine similarity	282
Collaborative filtering	283
Advantages of collaborative filtering over content-based filtering	284
Matrix factorization using the alternating least squares algorithm for collaborative filtering	284
Evaluation of recommendation engine model	287
Hyperparameter selection in recommendation engines using grid search	287
Recommendation engine application on movie lens data	288
User-user similarity matrix	291
Movie-movie similarity matrix	293
Collaborative filtering using ALS	295
Grid search on collaborative filtering	300
Summary	304
Chapter 8: Unsupervised Learning	305
<hr/>	
K-means clustering	306
K-means working methodology from first principles	307
Optimal number of clusters and cluster evaluation	314
The elbow method	314
K-means clustering with the iris data example	315
Principal component analysis - PCA	321
PCA working methodology from first principles	326
PCA applied on handwritten digits using scikit-learn	329
Singular value decomposition - SVD	340
SVD applied on handwritten digits using scikit-learn	341
Deep auto encoders	344
Model building technique using encoder-decoder architecture	345
Deep auto encoders applied on handwritten digits using Keras	347
Summary	358
Chapter 9: Reinforcement Learning	359
<hr/>	
Introduction to reinforcement learning	360
Comparing supervised, unsupervised, and reinforcement learning in detail	360
Characteristics of reinforcement learning	361
Reinforcement learning basics	362
Category 1 - value based	366

Category 2 - policy based	367
Category 3 - actor-critic	367
Category 4 - model-free	367
Category 5 - model-based	368
Fundamental categories in sequential decision making	369
Markov decision processes and Bellman equations	369
Dynamic programming	377
Algorithms to compute optimal policy using dynamic programming	378
Grid world example using value and policy iteration algorithms with basic Python	382
Monte Carlo methods	389
Comparison between dynamic programming and Monte Carlo methods	389
Key advantages of MC over DP methods	389
Monte Carlo prediction	391
The suitability of Monte Carlo prediction on grid-world problems	392
Modeling Blackjack example of Monte Carlo methods using Python	393
Temporal difference learning	403
Comparison between Monte Carlo methods and temporal difference learning	404
TD prediction	404
Driving office example for TD learning	406
SARSA on-policy TD control	407
Q-learning - off-policy TD control	409
Cliff walking example of on-policy and off-policy of TD control	410
Applications of reinforcement learning with integration of machine learning and deep learning	416
Automotive vehicle control - self-driving cars	416
Google DeepMind's AlphaGo	417
Robo soccer	418
Further reading	419
Summary	419
Index	421
