

Table of Contents

PREFACE	V
WHY THIS BOOK.....	V
WHOM THIS BOOK IS FOR	VI
HOW THIS BOOK IS ORGANIZED	VI
COLORED FIGURES	VI
EXTRAS	VI
SOFTWARE AND HARDWARE	VI
HOW TO USE THIS BOOK	VII
TYPOGRAPHIC CONVENTIONS	VII
HOW TO REACH THE AUTHOR	VIII
THE BOOK'S WEB SITE	VIII
ACKNOWLEDGEMENTS	IX
PART I MACHINE LEARNING BASICS	1
1 INTRODUCTION TO MACHINE LEARNING	3
1.1 WHAT IS ARTIFICIAL INTELLIGENCE?	3
1.1.1 <i>Definition of Artificial Intelligence</i>	3
1.1.2 <i>Some of the Applied Artificial Intelligence Application Examples</i>	4
1.1.3 <i>Comparison between Human Intelligence and Artificial Intelligence</i>	5
1.1.4 <i>Expert Systems and Modern Machine Learning Models</i>	8

1.1.5	<i>Affective Computing</i>	9
1.1.6	<i>Sentiment Analysis</i>	10
1.1.7	<i>Recommender Systems</i>	10
1.2	WHAT IS MACHINE LEARNING?.....	11
1.2.1	<i>Machine learning evolutions</i>	12
1.2.2	<i>Machine learning tasks</i>	14
1.2.3	<i>Machine learning application example 1: facial recognition</i>	15
1.2.4	<i>Machine learning application example 2: Netflix prize</i>	16
1.2.5	<i>Machine learning application example 3: anomaly detection</i>	18
1.3	MACHINE LEARNING PARADIGMS.....	20
1.4	MACHINE LEARNING SOFTWARE.....	21
1.5	MACHINE LEARNING HARDWARE.....	24
1.6	SUMMARY.....	24
	RECOMMENDED READING.....	25
	EXERCISES.....	25
2	MACHINE LEARNING FUNDAMENTALS ILLUSTRATED WITH REGRESSION	27
2.1	LEARNING THE CONTEXT FIRST.....	27
2.2	EXAMINE THE DATA.....	28
2.2.1	<i>Retrieving data</i>	28
2.2.2	<i>Data semantics</i>	29
2.3	LOADING, FILTERING AND EXPLORING DATA.....	30
2.3.1	<i>Loading data</i>	30
2.3.2	<i>Filtering data</i>	30
2.3.3	<i>Exploring data numerically</i>	31
2.3.4	<i>Exploring data graphically</i>	33
2.4	SELECTING AND TRAINING MODELS.....	37
2.4.1	<i>Labeling data</i>	37
2.4.2	<i>Coping with outliers</i>	38

2.4.3	<i>Fitting with a linear model</i>	39
2.5	EVALUATING A MACHINE LEARNING MODEL	41
2.5.1	<i>Mathematical representation of the linear regression model</i>	41
2.5.2	<i>MSE, RMSE and R^2 (coefficient of determination) as performance metrics</i>	42
2.6	TUNING MACHINE LEARNING MODELS.....	45
2.6.1	<i>Underfitting versus overfitting with polynomial features</i>	45
2.6.2	<i>Cross-Validation</i>	48
2.6.3	<i>L_1 and L_2 Norms</i>	56
2.6.4	<i>Ridge regularization</i>	58
2.6.5	<i>LASSO regularization</i>	60
2.6.6	<i>Elastic Net regularization</i>	62
2.7	LEARNING CURVES	64
2.8	BIAS-VARIANCE TRADE-OFFS	68
2.8.1	<i>Theoretical formulation</i>	68
2.8.2	<i>Illustration of the bias-variance trade-off triage with experiments</i>	70
2.8.3	<i>Implementation details</i>	77
2.9	NO FREE LUNCH (NFL) THEOREM FOR MACHINE LEARNING	82
2.10	SUMMARY	83
	RECOMMENDED READING	84
	EXERCISES.....	84
3	PATTERN RECOGNITION WITH CLASSIFICATION	87
3.1	THE MNIST DATASET	87
3.2	DATA SPLITTING WITH STRATIFIED K-FOLD.....	89
3.3	BINARY CLASSIFIERS.....	90
3.3.1	<i>Test error rate versus total computation time</i>	90
3.3.2	<i>Confusion Matrix</i>	93
3.3.3	<i>Precision, recall and F1 score</i>	99
3.3.4	<i>Decision function and threshold</i>	100

3.3.5	<i>Precision and recall trade-off</i>	103
3.3.6	<i>The ROC curve and AUC score</i>	107
3.4	CLASSIFICATIONS BEYOND BINARY CLASSIFIERS	110
3.4.1	<i>Multiclass Classification</i>	110
3.4.2	<i>Multiclass classification confusion matrix</i>	112
3.4.3	<i>3D plot of confusion matrix</i>	115
3.4.4	<i>Multilabel classification</i>	117
3.4.5	<i>Multioutput classification</i>	118
3.5	SUMMARY	118
	RECOMMENDED READING	119
	EXERCISES	119
4	OPTIMIZATION AND SEARCH ILLUSTRATED WITH LOGISTIC REGRESSION	121
4.1	OPTIMIZATION	121
4.1.1	<i>Optimization technique I: Batch gradient descent</i>	122
4.1.2	<i>Optimization technique II: Stochastic gradient descent (SGD)</i>	123
4.1.3	<i>Learning scheduler</i>	125
4.1.4	<i>Heavy ball method</i>	126
4.1.5	<i>Conjugate gradient and Newton-Raphson iteration</i>	128
4.2	SEARCH	131
4.3	LOGISTIC REGRESSION	132
4.3.1	<i>Logistic function</i>	132
4.3.2	<i>Cost function for logistic function</i>	133
4.3.3	<i>The Iris flower machine learning dataset</i>	134
4.3.4	<i>Logistic Regression Model</i>	137
4.3.5	<i>Softmax regression</i>	139
4.4	KULLBACK-LEIBLER DIVERGENCE (CROSS ENTROPY)	141
4.5	SUMMARY	141
	RECOMMENDED READING	142

EXERCISES.....	142
5 RULE-BASED LEARNING: DECISION TREES.....	143
5.1 THE ORIGIN OF DECISION TREES	143
5.1.1 <i>The earliest version of ID3</i>	143
5.1.2 <i>ID3 and Occam's razor principle</i>	145
5.1.3 <i>C4.5 and C5.0</i>	145
5.2 CLASSIFICATION AND REGRESSION TREES (CART)	146
5.2.1 <i>Gini impurity</i>	146
5.2.2 <i>Role of Gini impurity in splitting nodes</i>	146
5.2.3 <i>Stopping criterion</i>	147
5.2.4 <i>Pruning trees</i>	147
5.3 CLASSIFICATION EXAMPLES WITH TREE MODELS.....	148
5.3.1 <i>Iris dataset: petal width versus petal length</i>	148
5.3.2 <i>Iris dataset: sepal width versus sepal length</i>	152
5.4 REGRESSION	154
5.5 SUMMARY	155
RECOMMENDED READING	156
EXERCISES.....	156
6 INSTANCE-BASED LEARNING: SUPPORT VECTOR MACHINES	157
6.1 SVM FUNDAMENTALS.....	157
6.1.1 <i>The concept of support vectors</i>	158
6.1.2 <i>Linear SVMs</i>	158
6.1.3 <i>Quadratic programming</i>	159
6.1.4 <i>Hard margins versus soft margins: The primal problem</i>	160
6.1.5 <i>Hinge loss functions</i>	160
6.2 LINEARSVC MODELS.....	161
6.3 NONLINEAR SVMs.....	163
6.3.1 <i>Lagrangian dual problem</i>	163

6.3.2	<i>Kernels</i>	164
6.3.3	<i>Mercer’s theorem</i>	165
6.3.4	<i>Nonlinear SVM examples</i>	166
6.4	SUPPORT VECTOR REGRESSORS (SVRS).....	166
6.5	ISSUES WITH SVMs.....	169
6.6	SUMMARY.....	171
	RECOMMENDED READING.....	171
	EXERCISES.....	172
7	RANDOM FORESTS AND ENSEMBLE LEARNING	173
7.1	RANDOM FORESTS.....	173
7.1.1	<i>Concept of random forests</i>	173
7.1.2	<i>Instance sampling: Bagging and pasting</i>	174
7.1.3	<i>Feature sampling: Random patches and random subspaces</i>	175
7.1.4	<i>Baseline with decision trees</i>	175
7.1.5	<i>Random Forests</i>	176
7.1.6	<i>Extra Trees</i>	177
7.1.7	<i>AdaBoost</i>	179
7.2	FEATURE IMPORTANCE.....	179
7.2.1	<i>Iris dataset</i>	180
7.2.2	<i>MNIST dataset</i>	180
7.3	GRADIENT BOOSTING REGRESSION TREES (GBRT).....	182
7.4	ENSEMBLE LEARNING.....	184
7.4.1	<i>Hard voting versus soft voting</i>	184
7.4.2	<i>A hard voting example</i>	184
7.4.3	<i>A soft voting example</i>	186
7.4.4	<i>Bagging and pasting</i>	188
7.4.5	<i>Out-of-bag (OOB) evaluation</i>	189
7.5	SUMMARY.....	191

RECOMMENDED READING	191
EXERCISES.....	191
8 DIMENSIONALITY REDUCTION.....	193
8.1 LINEAR DIMENSIONALITY REDUCTION	194
8.1.1 <i>Eigenvalues and eigenvectors</i>	194
8.1.2 <i>Matrix diagonalization</i>	195
8.1.3 <i>Singular value decomposition (SVD) theorem</i>	196
8.1.4 <i>Principal component analysis (PCA) by SVD</i>	196
8.1.5 <i>Principal components analysis (PCA) versus covariance matrix</i>	197
8.1.6 <i>Principal component analysis (PCA) by maximizing variance</i>	197
8.2 EXAMPLES OF PCA.....	198
8.2.1 <i>PCA applied to the Iris dataset</i>	199
8.2.2 <i>SVD applied to the Iris dataset</i>	201
8.3 INCREMENTAL PCA	202
8.4 NONLINEAR DIMENSIONALITY REDUCTION	202
8.4.1 <i>Kernel PCA</i>	202
8.4.2 <i>Nonlinear Dimensionality Reduction by Locally Linear Embedding (LLE)</i>	203
8.5 SUMMARY	206
RECOMMENDED READING	206
EXERCISES.....	206
PART II ARTIFICIAL NEURAL NETWORKS	207
9 INTRODUCTION TO ARTIFICIAL NEURAL NETWORKS	209
9.1 HISTORY OF ARTIFICIAL NEURAL NETWORKS.....	209
9.1.1 <i>Neuron-based logical calculus by McCulloch and Pitts</i>	209
9.1.2 <i>Hebbian theory or Hebbian learning</i>	212
9.1.3 <i>Rosenblatt's Perceptron</i>	213
9.1.4 <i>Limitations of single layer perceptrons</i>	216
9.2 FEED-FORWARD NEURAL NETWORKS.....	218

9.2.1	<i>single layer perceptron network model</i>	218
9.2.2	<i>Feed-forward neural network architecture</i>	219
9.2.3	<i>Error backpropagation algorithm</i>	221
9.3	EARLIER ARTIFICIAL NEURAL NETWORK MODELS	224
9.3.1	<i>Hopfield networks</i>	224
9.3.2	<i>Boltzmann machines</i>	226
9.3.3	<i>Restricted Boltzmann machines</i>	231
9.4	REFERENCE IMPLEMENTATION	233
9.5	SUMMARY.....	234
	RECOMMENDED READING	234
	EXERCISES	234
10	CONVOLUTIONAL NEURAL NETWORKS	237
10.1	HOW HUMAN VISUAL SYSTEMS WORK.....	237
10.1.1	<i>Visual field</i>	238
10.1.2	<i>Laminar organization of the retina</i>	239
10.1.3	<i>The lateral geniculate nucleus (LGN)</i>	240
10.1.4	<i>The striate vortex layer</i>	241
10.1.5	<i>The receptive field</i>	243
10.1.6	<i>Neocognitron</i>	248
10.2	GENERIC ARCHITECTURE OF CONVOLUTIONAL NEURAL NETWORKS	249
10.2.1	<i>Convolution versus cross-correlation</i>	249
10.2.2	<i>Weight-sharing and backpropagation</i>	250
10.2.3	<i>Zero-Padding</i>	256
10.2.4	<i>Convolution and pooling layers</i>	257
10.2.5	<i>A generic CNN architecture</i>	258
10.3	LENET-5.....	259
10.3.1	<i>Overview</i>	260
10.3.2	<i>Configurations</i>	262

10.3.3	<i>Results</i>	263
10.4	ALEXNET.....	264
10.4.1	<i>Overview</i>	264
10.4.2	<i>Data augmentation for combating overfitting</i>	266
10.4.3	<i>Dropout combating overfitting</i>	267
10.4.4	<i>Results</i>	267
10.5	GOOGLNET.....	268
10.5.1	<i>1×1 kernels</i>	268
10.5.2	<i>Inception module</i>	269
10.5.3	<i>Global averaging pooling</i>	271
10.5.4	<i>Results</i>	271
10.6	RESNET.....	272
10.6.1	<i>The deep residual learning framework</i>	272
10.6.2	<i>Batch normalization</i>	273
10.6.3	<i>He Initialization</i>	274
10.6.4	<i>Architectures</i>	275
10.6.5	<i>Implementation and results</i>	277
10.7	2016 AND 2017 ILSVRC WINNERS	278
10.8	REFERENCE IMPLEMENTATIONS AND OBJECT IDENTIFICATION	278
10.9	SUMMARY	278
	RECOMMENDED READING	278
	EXERCISES.....	279
11	RECURRENT NEURAL NETWORKS	281
11.1	INTRODUCTION TO RNNs.....	281
11.1.1	<i>Simple RNN cells</i>	282
11.1.2	<i>Classification of RNNs</i>	283
11.2	LONG SHORT-TERM MEMORY (LSTM)	284
11.2.1	<i>Back-propagation through time (BPTT)</i>	285

11.2.2	<i>Vanishing/exploding gradients</i>	286
11.2.3	<i>Background information about the arrival of LSTM</i>	287
11.2.4	<i>Long short-term memory (LSTM)</i>	289
11.2.5	<i>LSTM extended with forget-gate</i>	292
11.2.6	<i>Peephole connections</i>	293
11.3	GATED RECURRENT UNIT (GRU)	293
11.4	NATURAL LANGUAGE PROCESSING (NLP)	294
11.4.1	<i>Sub-fields of NLP</i>	294
11.4.2	<i>Connectionist temporal classification (CTC)</i>	296
11.4.3	<i>LSTM applied to speech recognition</i>	297
11.5	REFERENCE IMPLEMENTATIONS	300
11.6	SUMMARY	300
	RECOMMENDED READING	301
	EXERCISES	301
12	AUTOENCODERS	303
12.1	HOW AUTOENCODERS WORK	303
12.2	DENOISING AUTOENCODERS	304
12.3	VARIATIONAL AUTOENCODERS	307
12.3.1	<i>A review of the Bayes' theorem</i>	307
12.3.2	<i>Variational inference</i>	308
12.3.3	<i>Auto-encoding variational Bayes (AEVB)</i>	309
12.4	ADVERSARIAL AUTOENCODERS	312
12.4.1	<i>Generative Adversarial Nets (GANs)</i>	312
12.4.2	<i>Adversarial Autoencoders (AAE)</i>	313
12.5	SUMMARY	314
	RECOMMENDED READING	315
	EXERCISES	315
	APPENDIX A MACHINE LEARNING DEVELOPMENT AND FRAMEWORKS	317

A.1 SET UP PYDEV ON ECLIPSE	317
A.2 OTHER IDES FOR PYTHON	318
A.3 PYTHON PACKAGES FOR MACHINE LEARNING	318
A.3.1 <i>pandas</i>	318
A.3.2 <i>NumPy</i>	319
A.3.3 <i>matplotlib</i>	319
A.3.4 <i>sklearn</i>	319
APPENDIX B MULTI-LAYER PERCEPTRON APPLIED TO THE FUEL ECONOMY USE CASE	321
B.1 THE BASELINE.....	321
B.2 THE EFFECTS OF THE NUMBER OF NEURONS.....	322
B.3 THE EFFECTS OF THE NUMBER OF LAYERS.....	324
B.4 THE EFFECTS THE ACTIVATION FUNCTIONS.....	325
B.5 THE EFFECTS THE SOLVERS.....	326
B.6 THE MLP MODEL FOR CLASSIFICATION TASKS.....	327
APPENDIX C CNN EXAMPLES WITH CAFFE, YOLOV3 AND PYTORCH	329
C.1 THE CAFFE FRAMEWORK	329
C.1.1 <i>Building the Caffe Framework from the Source</i>	329
C.1.2 <i>The LeNet CNN Model for The MNIST Dataset with Caffe</i>	332
C.1.3 <i>Alex’s CIFAR-10 with Caffe</i>	340
C.1.4 <i>The ImageNet Example with Caffe</i>	345
C.2 THE YOLOV3 FRAMEWORK	349
C.2.1 <i>Building the YOLOv3 Framework from the Source</i>	349
C.2.2 <i>ALEX’S CIFAR-10 WITH YOLOv3</i>	353
C.2.3 <i>Use YOLOv3 to detect multi-objects in an image or bounding boxes</i>	360
C.2.4 <i>Training YOLO on the COCO dataset</i>	364
C.2.5 <i>Profiling YOLO</i>	370
C.2.6 <i>The gemm and axpy functions</i>	378
C.2.7 <i>General Matrix Multiplication (GEMM) Examples</i>	380

C.2.8 Accelerating YOLOv3's GEMM on Mac OS with Apple's Accelerate Framework.....	386
C.2.9 COCO training on Mac OS by taking advantage of Apple's Accelerate Framework.....	388
C.2.10 How YOLOv3 training is kicked off.....	393
C.2.11 Object Detection with under-trained YOLOv3	397
C.3 IMAGE PROCESSING BASICS	400
C.3.1 How OpenCV processes images in C.....	400
C.3.2. Converting image matrices to column matrices with the <i>im2col</i> function.....	403
C.3.3 A more advanced image processing example in Java	406
C.3.4 RGB versus HSV image formats.....	409
C.4 THE ART AND SCIENCE OF DEEP LEARNING PERFORMANCE	409
C.5 TRAINING YOLOV3 ON GPUS	412
C5.1 YOLOv3 COCO training on GPUs.....	412
C.5.2 Creating a Ubuntu-18.04 VM	414
C.5.3 Installing cuda-9.1.....	415
C.5.4 Making YOLOv3 on the Ubuntu-18.04-cuda-9.1 VM.....	417
C5.5 Installing OpenCV on Ubuntu-18.04.....	418
C.6 PYTORCH	419
APPENDIX D RNN/LSTM EXAMPLE IMPLEMENTATIONS WITH KERAS/TENSORFLOW	425
D.1 INSTALLING KERAS/TENSORFLOW.....	425
D.2 THE KERAS LSTM TIME SERIES EXAMPLE.....	426
D.2.1 <i>Panads DataFrame</i>	426
D.2.2 <i>NumPy's repeat function</i>	427
D.3 AN LSTM EXAMPLE THAT MODELS A TIME SERIES SEQUENCE OF LIMITED LENGTH.....	428
D.4 MORE EXPERIMENTS WITH THE KERAS LSTM TIME SERIES EXAMPLE	434
D.4.1 <i>Stateful vs stateless LSTM models with $tsteps = 2$ and $lahead = 3$</i>	435
D.4.2 <i>Stateful vs stateless LSTM models with $tsteps = 3$ and $lahead = 2$</i>	437
D.4.3 <i>Stateful vs stateless LSTM models with $tsteps = 2$ and $lahead = 2$</i>	437
REFERENCES	441

INDEX445