

Contents at a Glance

About the Author	xiii
About the Technical Reviewer	xv
Acknowledgments	xvii
Introduction	xix
Chapter 1: Introduction to Deep Learning	1
Chapter 2: Mathematical Review	11
Chapter 3: A Review of Optimization and Machine Learning	45
Chapter 4: Single and Multilayer Perceptron Models	89
Chapter 5: Convolutional Neural Networks (CNNs).....	101
Chapter 6: Recurrent Neural Networks (RNNs).....	113
Chapter 7: Autoencoders, Restricted Boltzmann Machines, and Deep Belief Networks	125
Chapter 8: Experimental Design and Heuristics	137
Chapter 9: Hardware and Software Suggestions.....	167
Chapter 10: Machine Learning Example Problems.....	171
Chapter 11: Deep Learning and Other Example Problems.....	195
Chapter 12: Closing Statements	219
Index.....	221

Contents

About the Author	xiii
About the Technical Reviewer	xv
Acknowledgments	xvii
Introduction	xix
Chapter 1: Introduction to Deep Learning	1
Deep Learning Models	3
Single Layer Perceptron Model (SLP)	3
Multilayer Perceptron Model (MLP)	4
Convolutional Neural Networks (CNNs)	5
Recurrent Neural Networks (RNNs)	5
Restricted Boltzmann Machines (RBMs)	6
Deep Belief Networks (DBNs)	6
Other Topics Discussed	7
Experimental Design.....	7
Feature Selection.....	7
Applied Machine Learning and Deep Learning.....	7
History of Deep Learning	7
Summary.....	9
Chapter 2: Mathematical Review	11
Statistical Concepts.....	11
Probability.....	11
And vs. Or	12

CONTENTS

Bayes' Theorem	14
Random Variables	14
Variance	15
Standard Deviation	16
Coefficient of Determination (R Squared)	17
Mean Squared Error (MSE)	17
Linear Algebra	17
Scalars and Vectors	17
Properties of Vectors	18
Axioms	19
Subspaces	20
Matrices	20
Summary	43
Chapter 3: A Review of Optimization and Machine Learning	45
Unconstrained Optimization	45
Local Minimizers	47
Global Minimizers	47
Conditions for Local Minimizers	48
Neighborhoods	49
Interior and Boundary Points	50
Machine Learning Methods: Supervised Learning	50
History of Machine Learning	50
What Is an Algorithm?	51
Regression Models	51
Linear Regression	51
Choosing An Appropriate Learning Rate	55
Newton's Method	60
Levenberg-Marquardt Heuristic	61

What Is Multicollinearity?	62
Testing for Multicollinearity	62
Variance Inflation Factor (VIF)	62
Ridge Regression	62
Least Absolute Shrinkage and Selection Operator (LASSO)	63
Comparing Ridge Regression and LASSO	64
Evaluating Regression Models	64
Receiver Operating Characteristic (ROC) Curve	67
Confusion Matrix	68
Limitations to Logistic Regression	69
Support Vector Machine (SVM)	70
Sub-Gradient Method Applied to SVMs	72
Extensions of Support Vector Machines	73
Limitations Associated with SVMs	73
Machine Learning Methods: Unsupervised Learning	74
K-Means Clustering	74
Assignment Step	74
Update Step	75
Limitations of K-Means Clustering	75
Expectation Maximization (EM) Algorithm	76
Expectation Step	77
Maximization Step	77
Decision Tree Learning	78
Classification Trees	79
Regression Trees	80
Limitations of Decision Trees	81
Ensemble Methods and Other Heuristics	82
Gradient Boosting	82
Gradient Boosting Algorithm	82

Random Forest	83
Limitations to Random Forests	83
Bayesian Learning	83
Naïve Bayes Classifier	84
Limitations Associated with Bayesian Classifiers	84
Final Comments on Tuning Machine Learning Algorithms	85
Reinforcement Learning	86
Summary	87
Chapter 4: Single and Multilayer Perceptron Models	89
Single Layer Perceptron (SLP) Model	89
Training the Perceptron Model	90
Widrow-Hoff (WH) Algorithm	90
Limitations of Single Perceptron Models	91
Summary Statistics	94
Multi-Layer Perceptron (MLP) Model	94
Converging upon a Global Optimum	95
Back-propagation Algorithm for MLP Models:	95
Limitations and Considerations for MLP Models	97
How Many Hidden Layers to Use and How Many Neurons Are in It	99
Summary	100
Chapter 5: Convolutional Neural Networks (CNNs)	101
Structure and Properties of CNNs	101
Components of CNN Architectures	103
Convolutional Layer	103
Pooling Layer	105
Rectified Linear Units (ReLU) Layer	106
Fully Connected (FC) Layer	106
Loss Layer	107

Tuning Parameters	108
Notable CNN Architectures	108
Regularization	111
Summary	112
Chapter 6: Recurrent Neural Networks (RNNs).....	113
Fully Recurrent Networks	113
Training RNNs with Back-Propagation Through Time (BPPT)	114
Elman Neural Networks	115
Neural History Compressor	116
Long Short-Term Memory (LSTM)	116
Traditional LSTM	118
Training LSTMs	118
Structural Damping Within RNNs	119
Tuning Parameter Update Algorithm	119
Practical Example of RNN: Pattern Detection	120
Summary	124
Chapter 7: Autoencoders, Restricted Boltzmann Machines, and Deep Belief Networks	125
Autoencoders	125
Linear Autoencoders vs. Principal Components Analysis (PCA)	126
Restricted Boltzmann Machines	127
Contrastive Divergence (CD) Learning	129
Momentum Within RBMs	132
Weight Decay	133
Sparsity	133
No. and Type Hidden Units	133

Deep Belief Networks (DBNs)	134
Fast Learning Algorithm (Hinton and Osindero 2006)	135
Algorithm Steps	136
Summary	136
Chapter 8: Experimental Design and Heuristics	137
Analysis of Variance (ANOVA)	137
One-Way ANOVA	137
Two-Way (Multiple-Way) ANOVA	137
Mixed-Design ANOVA	138
Multivariate ANOVA (MANOVA)	138
F-Statistic and F-Distribution	138
Fisher's Principles	144
Plackett-Burman Designs	146
Space Filling	147
Full Factorial	147
Halton, Faure, and Sobol Sequences	148
A/B Testing	148
Simple Two-Sample A/B Test	149
Beta-Binomial Hierarchical Model for A/B Testing	149
Feature/Variable Selection Techniques	151
Backwards and Forward Selection	151
Principal Component Analysis (PCA)	152
Factor Analysis	154
Limitations of Factor Analysis	155
Handling Categorical Data	155
Encoding Factor Levels	156
Categorical Label Problems: Too Numerous Levels	156
Canonical Correlation Analysis (CCA)	156

Wrappers, Filters, and Embedded (WFE) Algorithms	157
Relief Algorithm	157
Other Local Search Methods	157
Hill Climbing Search Methods	158
Genetic Algorithms (GAs)	158
Simulated Annealing (SA)	159
Ant Colony Optimization (ACO)	159
Variable Neighborhood Search (VNS)	160
Reactive Search Optimization (RSO)	161
Reactive Prohibitions	162
Fixed Tabu Search	163
Reactive Tabu Search (RTS)	164
WalkSAT Algorithm	165
K-Nearest Neighbors (KNN)	165
Summary	166
Chapter 9: Hardware and Software Suggestions.....	167
Processing Data with Standard Hardware	167
Solid State Drives and Hard Drive Disks (HDD)	167
Graphics Processing Unit (GPU)	168
Central Processing Unit (CPU)	169
Random Access Memory (RAM)	169
Motherboard	169
Power Supply Unit (PSU)	170
Optimizing Machine Learning Software	170
Summary	170

Chapter 10: Machine Learning Example Problems.....	171
Problem 1: Asset Price Prediction	171
Problem Type: Supervised Learning—Regression	172
Description of the Experiment	173
Feature Selection.....	175
Model Evaluation	176
Ridge Regression.....	176
Support Vector Regression (SVR).....	178
Problem 2: Speed Dating	180
Problem Type: Classification.....	181
Preprocessing: Data Cleaning and Imputation.....	182
Feature Selection	185
Model Training and Evaluation	186
Method 1: Logistic Regression	186
Method 3: K-Nearest Neighbors (KNN)	189
Method 2: Bayesian Classifier	191
Summary.....	194
Chapter 11: Deep Learning and Other Example Problems.....	195
Autoencoders	195
Convolutional Neural Networks.....	202
Preprocessing.....	204
Model Building and Training	206
Collaborative Filtering.....	214
Summary.....	218
Chapter 12: Closing Statements	219
Index.....	221