



Artificial Intelligence for Humans, Volume 3: Deep Learning and Neural Networks

Jeff Heaton

Heaton Research, Inc.
St. Louis, MO, USA

Publisher: Heaton Research, Inc.

Artificial Intelligence for Humans, Volume 3: Neural Networks and Deep Learning

December, 2015

Author: Jeff Heaton

Editor: Tracy Heaton

ISBN: 978-1505714340

Edition: 1.0

Copyright © 2015 by Heaton Research Inc., 1734 Clarkson Rd. #107, Chesterfield, MO 63017-4976. World rights reserved. The author(s) created reusable code in this publication expressly for reuse by readers. Heaton Research, Inc. grants readers permission to reuse the code found in this publication or downloaded from our website so long as (author(s)) are attributed in any application containing the reusable code and the source code itself is never redistributed, posted online by electronic transmission, sold or commercially exploited as a stand-alone product. Aside from this specific exception concerning reusable code, no part of this publication may be stored in a retrieval system, transmitted, or reproduced in any way, including, but not limited to photo copy, photograph, magnetic, or other record, without prior agreement and written permission of the publisher.

Heaton Research, Encog, the Encog Logo and the Heaton Research logo are all trademarks of Heaton Research, Inc., in the United States and/or other countries.

TRADEMARKS: Heaton Research has attempted throughout this book to distinguish proprietary trademarks from descriptive terms by following the capitalization style used by the manufacturer.

The author and publisher have made their best efforts to prepare this book, so the content is based upon the final release of software whenever possible. Portions of the manuscript may be based upon pre-release versions supplied by software manufacturer(s). The author and the publisher make no representation or warranties of any kind with regard to the completeness or accuracy of the contents herein and accept no liability of any kind including but not limited to performance, merchantability, fitness for any particular

purpose, or any losses or damages of any kind caused or alleged to be caused directly or indirectly from this book.

SOFTWARE LICENSE AGREEMENT: TERMS AND CONDITIONS

The media and/or any online materials accompanying this book that are available now or in the future contain programs and/or text files (the “Software”) to be used in connection with the book. Heaton Research, Inc. hereby grants to you a license to use and distribute software programs that make use of the compiled binary form of this book’s source code. You may not redistribute the source code contained in this book, without the written permission of Heaton Research, Inc. Your purchase, acceptance, or use of the Software will constitute your acceptance of such terms.

The Software compilation is the property of Heaton Research, Inc. unless otherwise indicated and is protected by copyright to Heaton Research, Inc. or other copyright owner(s) as indicated in the media files (the “Owner(s)”). You are hereby granted a license to use and distribute the Software for your personal, noncommercial use only. You may not reproduce, sell, distribute, publish, circulate, or commercially exploit the Software, or any portion thereof, without the written consent of Heaton Research, Inc. and the specific copyright owner(s) of any component software included on this media.

In the event that the Software or components include specific license requirements or end-user agreements, statements of condition, disclaimers, limitations or warranties (“End-User License”), those End-User Licenses supersede the terms and conditions herein as to that particular Software component. Your purchase, acceptance, or use of the Software will constitute your acceptance of such End-User Licenses.

By purchase, use or acceptance of the Software you further agree to comply with all export laws and regulations of the United States as such laws and regulations may exist from time to time.

SOFTWARE SUPPORT

Components of the supplemental Software and any offers associated with them may be supported by the specific Owner(s) of that material but they are

not supported by Heaton Research, Inc.. Information regarding any available support may be obtained from the Owner(s) using the information provided in the appropriate README files or listed elsewhere on the media.

Should the manufacturer(s) or other Owner(s) cease to offer support or decline to honor any offer, Heaton Research, Inc. bears no responsibility. This notice concerning support for the Software is provided for your information only. Heaton Research, Inc. is not the agent or principal of the Owner(s), and Heaton Research, Inc. is in no way responsible for providing any support for the Software, nor is it liable or responsible for any support provided, or not provided, by the Owner(s).

WARRANTY

Heaton Research, Inc. warrants the enclosed media to be free of physical defects for a period of ninety (90) days after purchase. The Software is not available from Heaton Research, Inc. in any other form or media than that enclosed herein or posted to www.heatonresearch.com. If you discover a defect in the media during this warranty period, you may obtain a replacement of identical format at no charge by sending the defective media, postage prepaid, with proof of purchase to:

Heaton Research, Inc.
Customer Support Department
1734 Clarkson Rd #107
Chesterfield, MO 63017-4976
Web: www.heatonresearch.com
E-Mail: support@heatonresearch.com

DISCLAIMER

Heaton Research, Inc. makes no warranty or representation, either expressed or implied, with respect to the Software or its contents, quality, performance, merchantability, or fitness for a particular purpose. In no event will Heaton Research, Inc., its distributors, or dealers be liable to you or any other party for direct, indirect, special, incidental, consequential, or other damages

arising out of the use of or inability to use the Software or its contents even if advised of the possibility of such damage. In the event that the Software includes an online update feature, Heaton Research, Inc. further disclaims any obligation to provide this feature for any specific duration other than the initial posting.

The exclusion of implied warranties is not permitted by some states. Therefore, the above exclusion may not apply to you. This warranty provides you with specific legal rights; there may be other rights that you may have that vary from state to state. The pricing of the book with the Software by Heaton Research, Inc. reflects the allocation of risk and limitations on liability contained in this agreement of Terms and Conditions.

SHAREWARE DISTRIBUTION

This Software may use various programs and libraries that are distributed as shareware. Copyright laws apply to both shareware and ordinary commercial software, and the copyright Owner(s) retains all rights. If you try a shareware program and continue using it, you are expected to register it. Individual programs differ on details of trial periods, registration, and payment. Please observe the requirements stated in appropriate files.

*This book is dedicated to my mom Mary,
thank you for all the love
and encouragement over the years.*

Contents

Introduction	xxi
0.1 Series Introduction	xxi
0.1.1 Programming Languages	xxii
0.1.2 Online Labs	xxiii
0.1.3 Code Repositories	xxiii
0.1.4 Books Planned for the Series	xxiii
0.1.5 Other Resources	xxiv
0.2 Neural Networks Introduction	xxv
0.3 The Kickstarter Campaign	xxvi
0.4 Background Information	xxxiv
0.4.1 Neural Network Structure	xxxv
0.4.2 A Simple Example	xxxvii
0.4.3 Training: Supervised and Unsupervised	xxxviii
0.4.4 Miles per Gallon	xxxix
0.5 A Neural Network Roadmap	xli
0.6 Data Sets Used in this Book	xliii
0.6.1 MNIST Handwritten Digits	xliii
0.6.2 Iris Data Set	xliv
0.6.3 Auto MPG Data Set	xlvi
0.6.4 Sunspots Data Set	xlvii

0.6.5	XOR Operator	xlvi
0.6.6	Kaggle Otto Group Challenge	xlvi
1	Neural Network Basics	1
1.1	Neurons and Layers	2
1.2	Types of Neurons	7
1.2.1	Input and Output Neurons	7
1.2.2	Hidden Neurons	8
1.2.3	Bias Neurons	9
1.2.4	Context Neurons	9
1.2.5	Other Neuron Types	11
1.3	Activation Functions	11
1.3.1	Linear Activation Function	12
1.3.2	Step Activation Function	13
1.3.3	Sigmoid Activation Function	14
1.3.4	Hyperbolic Tangent Activation Function	15
1.4	Rectified Linear Units (ReLU)	16
1.4.1	Softmax Activation Function	17
1.4.2	What Role does Bias Play?	20
1.5	Logic with Neural Networks	22
1.6	Chapter Summary	25
2	Self-Organizing Maps	29
2.1	Self-Organizing Maps	30
2.1.1	Understanding Neighborhood Functions	34
2.1.2	Mexican Hat Neighborhood Function	38
2.1.3	Calculating SOM Error	40
2.2	Chapter Summary	41
3	Hopfield & Boltzmann Machines	43

3.1	Hopfield Neural Networks	44
3.1.1	Training a Hopfield Network	46
3.2	Hopfield-Tank Networks	50
3.3	Boltzmann Machines	51
3.3.1	Boltzmann Machine Probability	53
3.4	Applying the Boltzmann Machine	54
3.4.1	Traveling Salesman Problem	54
3.4.2	Optimization Problems	58
3.4.3	Boltzmann Machine Training	62
3.5	Chapter Summary	62
4	Feedforward Neural Networks	65
4.1	Feedforward Neural Network Structure	66
4.1.1	Single-Output Neural Networks for Regression	66
4.2	Calculating the Output	69
4.3	Initializing Weights	73
4.4	Radial-Basis Function Networks	76
4.4.1	Radial-Basis Functions	77
4.4.2	Radial-Basis Function Networks	80
4.5	Normalizing Data	82
4.5.1	One-of-N Encoding	82
4.5.2	Range Normalization	84
4.5.3	Z-Score Normalization	85
4.5.4	Complex Normalization	88
4.6	Chapter Summary	89
5	Training & Evaluation	93
5.1	Evaluating Classification	94
5.1.1	Binary Classification	95

5.1.2	Multi-Class Classification	102
5.1.3	Log Loss	103
5.1.4	Multi-Class Log Loss	105
5.2	Evaluating Regression	106
5.3	Training with Simulated Annealing	107
5.4	Chapter Summary	110
6	Backpropagation Training	113
6.1	Understanding Gradients	114
6.1.1	What is a Gradient	115
6.1.2	Calculating Gradients	116
6.2	Calculating Output Node Deltas	118
6.2.1	Quadratic Error function	119
6.2.2	Cross Entropy Error Function	120
6.3	Calculating Remaining Node Deltas	120
6.4	Derivatives of the Activation Functions	121
6.4.1	Derivative of the Linear Activation Function	121
6.4.2	Derivative of the Softmax Activation Function	121
6.4.3	Derivative of the Sigmoid Activation Function	122
6.4.4	Derivative of the Hyperbolic Tangent Activation Function	123
6.4.5	Derivative of the ReLU Activation Function	123
6.5	Applying Backpropagation	125
6.5.1	Batch and Online Training	125
6.5.2	Stochastic Gradient Descent	126
6.5.3	Backpropagation Weight Update	127
6.5.4	Choosing Learning Rate and Momentum	128
6.5.5	Nesterov Momentum	129
6.6	Chapter Summary	130

7	Other Propagation Training	133
7.1	Resilient Propagation	134
7.2	RPROP Arguments	134
7.3	Data Structures	136
7.4	Understanding RPROP	137
7.4.1	Determine Sign Change of Gradient	137
7.4.2	Calculate Weight Change	139
7.4.3	Modify Update Values	140
7.5	Levenberg-Marquardt Algorithm	140
7.6	Calculation of the Hessian	143
7.7	LMA with Multiple Outputs	145
7.8	Overview of the LMA Process	146
7.9	Chapter Summary	146
8	NEAT, CPPN & HyperNEAT	149
8.1	NEAT Networks	150
8.1.1	NEAT Mutation	153
8.1.2	NEAT Crossover	154
8.1.3	NEAT Speciation	158
8.2	CPPN Networks	159
8.2.1	CPPN Phenotype	161
8.3	HyperNEAT Networks	165
8.3.1	HyperNEAT Substrate	165
8.3.2	HyperNEAT Computer Vision	166
8.4	Chapter Summary	168
9	Deep Learning	171
9.1	Deep Learning Components	172
9.2	Partially Labeled Data	172

9.3	Rectified Linear Units	173
9.4	Convolutional Neural Networks	174
9.5	Neuron Dropout	175
9.6	GPU Training	176
9.7	Tools for Deep Learning	178
9.7.1	H2O	179
9.7.2	Theano	179
9.7.3	Lasagne and NoLearn	180
9.7.4	ConvNetJS	181
9.8	Deep Belief Neural Networks	181
9.8.1	Restricted Boltzmann Machines	185
9.8.2	Training a DBNN	186
9.8.3	Layer-Wise Sampling	187
9.8.4	Computing Positive Gradients	188
9.8.5	Gibbs Sampling	190
9.8.6	Update Weights & Biases	191
9.8.7	DBNN Backpropagation	192
9.8.8	Deep Belief Application	193
9.9	Chapter Summary	195
10	Convolutional Neural Networks	197
10.1	LeNET-5	198
10.2	Convolutional Layers	200
10.3	Max-Pool Layers	203
10.4	Dense Layers	204
10.5	ConvNets for the MNIST Data Set	205
10.6	Chapter Summary	207
11	Pruning and Model Selection	209

11.1	Understanding Pruning	210
11.1.1	Pruning Connections	210
11.1.2	Pruning Neurons	211
11.1.3	Improving or Degrading Performance	211
11.2	Pruning Algorithm	212
11.3	Model Selection	213
11.3.1	Grid Search Model Selection	214
11.3.2	Random Search Model Selection	217
11.3.3	Other Model Selection Techniques	218
11.4	Chapter Summary	219
12	Dropout and Regularization	223
12.1	L1 and L2 Regularization	224
12.1.1	Understanding L1 Regularization	225
12.1.2	Understanding L2 Regularization	227
12.2	Dropout Layers	228
12.2.1	Dropout Layer	228
12.2.2	Implementing a Dropout Layer	229
12.3	Using Dropout	232
12.4	Chapter Summary	234
13	Time Series and Recurrent Networks	237
13.1	Time Series Encoding	238
13.1.1	Encoding Data for Input and Output Neurons	239
13.1.2	Predicting the Sine Wave	241
13.2	Simple Recurrent Neural Networks	243
13.2.1	Elman Neural Networks	245
13.2.2	Jordan Neural Networks	246
13.2.3	Backpropagation through Time	248

13.2.4	Gated Recurrent Units	252
13.3	Chapter Summary	253
14	Architecting Neural Networks	257
14.1	Evaluating Neural Networks	258
14.2	Training Parameters	259
14.2.1	Learning Rate	259
14.2.2	Momentum	261
14.2.3	Batch Size	263
14.3	General Hyper-Parameters	264
14.3.1	Activation Functions	264
14.3.2	Hidden Neuron Configurations	266
14.4	LeNet-5 Hyper-Parameters	267
14.5	Chapter Summary	268
15	Visualization	271
15.1	Confusion Matrix	272
15.1.1	Reading a Confusion Matrix	272
15.1.2	Generating a Confusion Matrix	273
15.2	t-SNE Dimension Reduction	274
15.2.1	t-SNE as a Visualization	276
15.2.2	t-SNE Beyond Visualization	280
15.3	Chapter Summary	281
16	Modeling with Neural Networks	285
16.0.1	Lessons from the Challenge	289
16.0.2	The Winning Approach to the Challenge	291
16.0.3	Our Approach to the Challenge	293
16.1	Modeling with Deep Learning	294
16.1.1	Neural Network Structure	294

16.1.2 Bagging Multiple Neural Networks	298
16.2 Chapter Summary	299
A Examples	303
A.1 Artificial Intelligence for Humans	303
A.2 Latest Versions	304
A.3 Obtaining the Examples	304
A.3.1 Download ZIP File	304
A.3.2 Clone the Git Repository	305
A.4 Example Contents	306
A.5 Contributing to the Project	309
References	311

