

contents

<i>preface</i>	<i>xiii</i>
<i>acknowledgments</i>	<i>xv</i>
<i>about this book</i>	<i>xvi</i>
<i>about the author</i>	<i>xx</i>
<i>about the cover</i>	<i>xxi</i>

PART 1 FUNDAMENTALS OF DEEP LEARNING1

1 What is deep learning? 3

1.1 Artificial intelligence, machine learning, and deep learning 4

Artificial intelligence 4 • *Machine learning 4* • *Learning representations from data 6* • *The “deep” in deep learning 8*
Understanding how deep learning works, in three figures 9
What deep learning has achieved so far 11 • *Don’t believe the short-term hype 12* • *The promise of AI 13*

1.2 Before deep learning: a brief history of machine learning 14

Probabilistic modeling 14 • *Early neural networks 14*
Kernel methods 15 • *Decision trees, random forests,
and gradient boosting machines 16* • *Back to neural
networks 17* • *What makes deep learning different 17*
The modern machine-learning landscape 18

1.3 Why deep learning? Why now? 20

Hardware 20 ▪ *Data* 21 ▪ *Algorithms* 21 ▪ *A new wave of investment* 22 ▪ *The democratization of deep learning* 23 ▪ *Will it last?* 23

2 Before we begin: the mathematical building blocks of neural networks 25

2.1 A first look at a neural network 27

2.2 Data representations for neural networks 31

Scalars (0D tensors) 31 ▪ *Vectors (1D tensors)* 31
Matrices (2D tensors) 31 ▪ *3D tensors and higher-dimensional tensors* 32 ▪ *Key attributes* 32
Manipulating tensors in Numpy 34 ▪ *The notion of data batches* 34 ▪ *Real-world examples of data tensors* 35 ▪ *Vector data* 35 ▪ *Timeseries data or sequence data* 35 ▪ *Image data* 36 ▪ *Video data* 37

2.3 The gears of neural networks: tensor operations — 38

Element-wise operations 38 ▪ *Broadcasting* 39 ▪ *Tensor dot* 40 ▪ *Tensor reshaping* 42 ▪ *Geometric interpretation of tensor operations* 43 ▪ *A geometric interpretation of deep learning* 44

2.4 The engine of neural networks: gradient-based optimization 46

What's a derivative? 47 ▪ *Derivative of a tensor operation: the gradient* 48 ▪ *Stochastic gradient descent* 48
Chaining derivatives: the Backpropagation algorithm 51

2.5 Looking back at our first example 53

2.6 Chapter summary 55

3 Getting started with neural networks 56

3.1 Anatomy of a neural network 58

Layers: the building blocks of deep learning 58 ▪ *Models: networks of layers* 59 ▪ *Loss functions and optimizers: keys to configuring the learning process* 60

3.2 Introduction to Keras 61

Keras, TensorFlow, Theano, and CNTK 62 ▪ *Developing with Keras: a quick overview* 62

3.3 Setting up a deep-learning workstation 65

Jupyter notebooks: the preferred way to run deep-learning experiments 65 ▪ *Getting Keras running: two options* 66

	<i>Running deep-learning jobs in the cloud: pros and cons</i>	66
	<i>What is the best GPU for deep learning?</i>	66
3.4	Classifying movie reviews: a binary classification example	68
	<i>The IMDB dataset</i>	68
	<i>Preparing the data</i>	69
	<i>Building your network</i>	70
	<i>Validating your approach</i>	73
	<i>Using a trained network to generate predictions on new data</i>	76
	<i>Further experiments</i>	77
	<i>Wrapping up</i>	77
3.5	Classifying newswires: a multiclass classification example	78
	<i>The Reuters dataset</i>	78
	<i>Preparing the data</i>	79
	<i>Building your network</i>	79
	<i>Validating your approach</i>	80
	<i>Generating predictions on new data</i>	83
	<i>A different way to handle the labels and the loss</i>	83
	<i>The importance of having sufficiently large intermediate layers</i>	83
	<i>Further experiments</i>	84
	<i>Wrapping up</i>	84
3.6	Predicting house prices: a regression example	85
	<i>The Boston Housing Price dataset</i>	85
	<i>Preparing the data</i>	86
	<i>Building your network</i>	86
	<i>Validating your approach using K-fold validation</i>	87
	<i>Wrapping up</i>	91
3.7	Chapter summary	92
4	<i>Fundamentals of machine learning</i>	93
4.1	Four branches of machine learning	94
	<i>Supervised learning</i>	94
	<i>Unsupervised learning</i>	94
	<i>Self-supervised learning</i>	94
	<i>Reinforcement learning</i>	95
4.2	Evaluating machine-learning models	97
	<i>Training, validation, and test sets</i>	97
	<i>Things to keep in mind</i>	100
4.3	Data preprocessing, feature engineering, and feature learning	101
	<i>Data preprocessing for neural networks</i>	101
	<i>Feature engineering</i>	102
4.4	Overfitting and underfitting	104
	<i>Reducing the network's size</i>	104
	<i>Adding weight regularization</i>	107
	<i>Adding dropout</i>	109
4.5	The universal workflow of machine learning	111
	<i>Defining the problem and assembling a dataset</i>	111
	<i>Choosing a measure of success</i>	112
	<i>Deciding on an</i>	

- evaluation protocol* 112 ▪ *Preparing your data* 112
- Developing a model that does better than a baseline* 113
- Scaling up: developing a model that overfits* 114
- Regularizing your model and tuning your hyperparameters* 114

4.6 Chapter summary 116

PART 2 DEEP LEARNING IN PRACTICE117

5 *Deep learning for computer vision* 119

5.1 Introduction to convnets 120

- The convolution operation* 122 ▪ *The max-pooling operation* 127

5.2 Training a convnet from scratch on a small dataset 130

- The relevance of deep learning for small-data problems* 130
- Downloading the data* 131 ▪ *Building your network* 133
- Data preprocessing* 135 ▪ *Using data augmentation* 138

5.3 Using a pretrained convnet 143

- Feature extraction* 143 ▪ *Fine-tuning* 152 ▪ *Wrapping up* 159

5.4 Visualizing what convnets learn 160

- Visualizing intermediate activations* 160 ▪ *Visualizing convnet filters* 167 ▪ *Visualizing heatmaps of class activation* 172

5.5 Chapter summary 177

6 *Deep learning for text and sequences* 178

6.1 Working with text data 180

- One-hot encoding of words and characters* 181 ▪ *Using word embeddings* 184 ▪ *Putting it all together: from raw text to word embeddings* 188 ▪ *Wrapping up* 195

6.2 Understanding recurrent neural networks 196

- A recurrent layer in Keras* 198 ▪ *Understanding the LSTM and GRU layers* 202 ▪ *A concrete LSTM example in Keras* 204 ▪ *Wrapping up* 206

6.3 Advanced use of recurrent neural networks 207

- A temperature-forecasting problem* 207 ▪ *Preparing the data* 210 ▪ *A common-sense, non-machine-learning baseline* 212 ▪ *A basic machine-learning approach* 213
- A first recurrent baseline* 215 ▪ *Using recurrent dropout*

	<i>to fight overfitting</i>	216	▪	<i>Stacking recurrent layers</i>	217
	<i>Using bidirectional RNNs</i>	219	▪	<i>Going even further</i>	222
	<i>Wrapping up</i>	223			
6.4	Sequence processing with convnets	225			
	<i>Understanding 1D convolution for sequence data</i>	225			
	<i>1D pooling for sequence data</i>	226	▪	<i>Implementing a 1D convnet</i>	226
	▪ <i>Combining CNNs and RNNs to process long sequences</i>	228	▪	<i>Wrapping up</i>	231
6.5	Chapter summary	232			
7	<i>Advanced deep-learning best practices</i>	233			
7.1	Going beyond the Sequential model: the Keras functional API	234			
	<i>Introduction to the functional API</i>	236	▪	<i>Multi-input models</i>	238
	▪ <i>Multi-output models</i>	240	▪	<i>Directed acyclic graphs of layers</i>	242
	▪ <i>Layer weight sharing</i>	246	▪	<i>Models as layers</i>	247
	▪ <i>Wrapping up</i>	248			
7.2	Inspecting and monitoring deep-learning models using Keras callbacks and TensorBoard	249			
	<i>Using callbacks to act on a model during training</i>	249			
	<i>Introduction to TensorBoard: the TensorFlow visualization framework</i>	252	▪	<i>Wrapping up</i>	259
7.3	Getting the most out of your models	260			
	<i>Advanced architecture patterns</i>	260	▪	<i>Hyperparameter optimization</i>	263
	▪ <i>Model ensembling</i>	264	▪	<i>Wrapping up</i>	266
7.4	Chapter summary	268			
8	<i>Generative deep learning</i>	269			
8.1	Text generation with LSTM	271			
	<i>A brief history of generative recurrent networks</i>	271	▪	<i>How do you generate sequence data?</i>	272
	▪ <i>The importance of the sampling strategy</i>	272	▪	<i>Implementing character-level LSTM text generation</i>	274
	▪ <i>Wrapping up</i>	279			
8.2	DeepDream	280			
	<i>Implementing DeepDream in Keras</i>	281	▪	<i>Wrapping up</i>	286
8.3	Neural style transfer	287			
	<i>The content loss</i>	288	▪	<i>The style loss</i>	288
	▪ <i>Neural style transfer in Keras</i>	289	▪	<i>Wrapping up</i>	295

8.4	Generating images with variational autoencoders	296
	<i>Sampling from latent spaces of images</i>	296
	<i>Concept vectors for image editing</i>	297
	<i>Variational autoencoders</i>	298
	<i>Wrapping up</i>	304
8.5	Introduction to generative adversarial networks	305
	<i>A schematic GAN implementation</i>	307
	<i>A bag of tricks</i>	307
	<i>The generator</i>	308
	<i>The discriminator</i>	309
	<i>The adversarial network</i>	310
	<i>How to train your DCGAN</i>	310
	<i>Wrapping up</i>	312
8.6	Chapter summary	313
9	Conclusions	314
9.1	Key concepts in review	315
	<i>Various approaches to AI</i>	315
	<i>What makes deep learning special within the field of machine learning</i>	315
	<i>How to think about deep learning</i>	316
	<i>Key enabling technologies</i>	317
	<i>The universal machine-learning workflow</i>	318
	<i>Key network architectures</i>	319
	<i>The space of possibilities</i>	322
9.2	The limitations of deep learning	325
	<i>The risk of anthropomorphizing machine-learning models</i>	325
	<i>Local generalization vs. extreme generalization</i>	327
	<i>Wrapping up</i>	329
9.3	The future of deep learning	330
	<i>Models as programs</i>	330
	<i>Beyond backpropagation and differentiable layers</i>	332
	<i>Automated machine learning</i>	332
	<i>Lifelong learning and modular subroutine reuse</i>	333
	<i>The long-term vision</i>	335
9.4	Staying up to date in a fast-moving field	337
	<i>Practice on real-world problems using Kaggle</i>	337
	<i>Read about the latest developments on arXiv</i>	337
	<i>Explore the Keras ecosystem</i>	338
9.5	Final words	339
appendix A	<i>Installing Keras and its dependencies on Ubuntu</i>	340
appendix B	<i>Running Jupyter notebooks on an EC2 GPU instance</i>	345
	<i>index</i>	353