

contents

preface xiii
acknowledgments xv
about this book xvi
about the author xx
about the cover xxi

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

Running deep-learning jobs in the cloud: pros and cons 66
What is the best GPU for deep learning? 66

3.4 Classifying movie reviews: a binary classification example 68

The IMDB dataset 68 ▪ *Preparing the data* 69
Building your network 70 ▪ *Validating your approach* 73
Using a trained network to generate predictions on new data 76 ▪ *Further experiments* 77 ▪ *Wrapping up* 77

3.5 Classifying newswires: a multiclass classification example 78

The Reuters dataset 78 ▪ *Preparing the data* 79
Building your network 79 ▪ *Validating your approach* 80
Generating predictions on new data 83 ▪ *A different way to handle the labels and the loss* 83 ▪ *The importance of having sufficiently large intermediate layers* 83 ▪ *Further experiments* 84 ▪ *Wrapping up* 84

3.6 Predicting house prices: a regression example 85

The Boston Housing Price dataset 85 ▪ *Preparing the data* 86 ▪ *Building your network* 86 ▪ *Validating your approach using K-fold validation* 87 ▪ *Wrapping up* 91

3.7 Chapter summary 92

4 Fundamentals of machine learning 93

4.1 Four branches of machine learning 94

Supervised learning 94 ▪ *Unsupervised learning* 94
Self-supervised learning 94 ▪ *Reinforcement learning* 95

4.2 Evaluating machine-learning models 97

Training, validation, and test sets 97 ▪ *Things to keep in mind* 100

4.3 Data preprocessing, feature engineering, and feature learning 101

Data preprocessing for neural networks 101 ▪ *Feature engineering* 102

4.4 Overfitting and underfitting 104

Reducing the network's size 104 ▪ *Adding weight regularization* 107 ▪ *Adding dropout* 109

4.5 The universal workflow of machine learning 111

Defining the problem and assembling a dataset 111
Choosing a measure of success 112 ▪ *Deciding on an*

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*

to fight overfitting 216 ▪ *Stacking recurrent layers* 217
Using bidirectional RNNs 219 ▪ *Going even further* 222
Wrapping up 223

6.4 Sequence processing with convnets 225

Understanding 1D convolution for sequence data 225
1D pooling for sequence data 226 ▪ *Implementing a 1D convnet* 226 ▪ *Combining CNNs and RNNs to process long sequences* 228 ▪ *Wrapping up* 231

6.5 Chapter summary 232

7 Advanced deep-learning best practices 233

7.1 Going beyond the Sequential model: the Keras functional API 234

Introduction to the functional API 236 ▪ *Multi-input models* 238 ▪ *Multi-output models* 240 ▪ *Directed acyclic graphs of layers* 242 ▪ *Layer weight sharing* 246 ▪ *Models as layers* 247 ▪ *Wrapping up* 248

7.2 Inspecting and monitoring deep-learning models using Keras callbacks and TensorBoard 249

Using callbacks to act on a model during training 249
Introduction to TensorBoard: the TensorFlow visualization framework 252 ▪ *Wrapping up* 259

7.3 Getting the most out of your models 260

Advanced architecture patterns 260 ▪ *Hyperparameter optimization* 263 ▪ *Model ensembling* 264 ▪ *Wrapping up* 266

7.4 Chapter summary 268

8 Generative deep learning 269

8.1 Text generation with LSTM 271

A brief history of generative recurrent networks 271 ▪ *How do you generate sequence data?* 272 ▪ *The importance of the sampling strategy* 272 ▪ *Implementing character-level LSTM text generation* 274 ▪ *Wrapping up* 279

8.2 DeepDream 280

Implementing DeepDream in Keras 281 ▪ *Wrapping up* 286

8.3 Neural style transfer 287

The content loss 288 ▪ *The style loss* 288 ▪ *Neural style transfer in Keras* 289 ▪ *Wrapping up* 295

8.4 Generating images with variational autoencoders 296

Sampling from latent spaces of images 296 ▪ *Concept vectors for image editing* 297 ▪ *Variational autoencoders* 298
Wrapping up 304

8.5 Introduction to generative adversarial networks 305

A schematic GAN implementation 307 ▪ *A bag of tricks* 307
The generator 308 ▪ *The discriminator* 309 ▪ *The adversarial network* 310 ▪ *How to train your DCGAN* 310 ▪ *Wrapping up* 312

8.6 Chapter summary 313

9 Conclusions 314

9.1 Key concepts in review 315

Various approaches to AI 315 ▪ *What makes deep learning special within the field of machine learning* 315 ▪ *How to think about deep learning* 316 ▪ *Key enabling technologies* 317
The universal machine-learning workflow 318 ▪ *Key network architectures* 319 ▪ *The space of possibilities* 322

9.2 The limitations of deep learning 325

The risk of anthropomorphizing machine-learning models 325
Local generalization vs. extreme generalization 327
Wrapping up 329

9.3 The future of deep learning 330

Models as programs 330 ▪ *Beyond backpropagation and differentiable layers* 332 ▪ *Automated machine learning* 332
Lifelong learning and modular subroutine reuse 333
The long-term vision 335

9.4 Staying up to date in a fast-moving field 337

Practice on real-world problems using Kaggle 337
Read about the latest developments on arXiv 337
Explore the Keras ecosystem 338

9.5 Final words 339

appendix A Installing Keras and its dependencies on Ubuntu 340

appendix B Running Jupyter notebooks on an EC2 GPU instance 345

index 353