

Deep Learning with R

FRANCOIS CHOLLET

WITH J.J. ALLAIRE

**MANNING
SHELTER ISLAND**

contents

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

PAR I 1 FUNDAMENTALS OF DEEP LEARNING 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* 12

1.2 Before deep learning: a brief history of machine learning 13

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? 19

Hardware 19 • *Data* 20 • *Algorithms* 21 • *A new wave of investment* 21 • *The democratization of deep learning* 22
Will it last? 22

) *Before we begin: the mathematical building blocks of neural networks* 24

2.1 A first look at a neural network 25

2.2 Data representations for neural networks 29

Scalars (0D tensors) 29 • *Vectors (1D tensors)* 29 • *Matrices (2D tensors)* 30 • *3D tensors and higher-dimensional tensors* 30
Key attributes 30 • *Manipulating tensors in R* 31 • *The notion of data batches* 32 • *Real-world examples of data tensors* 32
Vector data 32 • *Timeseries data or sequence data* 33
Image data 33 • *Video data* 34

2.3 The gears of neural networks: tensor operations 34

Element-wise operations 35 • *Operations involving tensors of different dimensions* 36 • *Tensor dot* 36 • *Tensor reshaping* 38 • *Geometric interpretation of tensor operations* 39
A geometric interpretation of deep learning 40

2.4 The engine of neural networks: gradient-based optimization 41

What's a derivative? 42 • *Derivative of a tensor operation: the gradient* 43 • *Stochastic gradient descent* 44 • *Chaining derivatives: the Backpropagation algorithm* 46

2.5 Looking back at our first example 47

2.6 Summary 49

j *Getting started with neural networks* 50

3.1 Anatomy of a neural network 51

Layers: the building blocks of deep learning 52 • *Models: networks of layers* 52 • *Loss functions and optimizers: keys to configuring the learning process* 53

3.2 Introduction to Keras 54

Keras, TensorFlow, Theano, and CNTK 54 • *Installing Keras* 56 • *Developing with Keras: a quick overview* 56

3.3 Setting up a deep-learning workstation 57

Getting Keras running: two options 58 • *Running deep-learning jobs in the cloud: pros and cons* 58 • *What is the best GPU for deep learning?* 59

- 3.4 Classifying movie reviews:
a binary classification example 59**
The IMDB dataset 59 • Preparing the data 61 • Building your network 62 • Validating your approach 65 • Using a trained network to generate predictions on new data 68 • Further experiments 69 • Wrapping up 69
- 3.5 Classifying newswires: a multiclass classification
example 70**
*The Reuters dataset 70 • Preparing the data 71
Building your network 72 • Validating your approach 73
Generating predictions on new data 74 • A different way to handle the labels and the loss 75 • The importance of having sufficiently large intermediate layers 75 • Further experiments 76 • Wrapping up 76*
- 3.6 Predicting house prices: a regression example 76**
*The Boston Housing Price dataset 77 • Preparing the data 77
Building your network 78 • Validating your approach using K-fold validation 79 • Wrapping up 83*
- 3.7 Summary 83**

Fundamentals of machine learning 84

- 4.1 Four branches of machine learning 85**
*Supervised learning 85 • Unsupervised learning 85
Self-supervised learning 86 • Reinforcement learning 86*
- 4.2 Evaluating machine-learning models 87**
Training, validation, and test sets 88 • Things to keep in mind 91
- 4.3 Data preprocessing, feature engineering,
and feature learning 91**
Data preprocessing for neural networks 91 • Feature engineering 93
- 4.4 Overfitting and underfitting 94**
Reducing the network's size 95 • Adding weight regularization 98 • Adding dropout 100
- 4.5 The universal workflow of machine learning 102**
Defining the problem and assembling a dataset 102 • Choosing a measure of success 103 • Deciding on an evaluation protocol 104 • Preparing your data 104

Developing a model that does better than a baseline 104
Scaling up: developing a model that overfits 105
Regularizing your model and tuning your hyperparameters 106

4.6 Summary 107

PART 2 DEEP LEARNING IN PRACTICE

109

***Deep learning for computer vision* 111**

5.1 Introduction to convnets 111

The convolution operation 114 • *The max-pooling operation* 119

5.2 Training a convnet from scratch on a small dataset 121

The relevance of deep learning for small-data problems 121
Downloading the data 122 • *Building your network* 124
Data preprocessing 126 • *Using data augmentation* 128

5.3 Using a pretrained convnet 132

Feature extraction 133 • *Fine-tuning* 142
Wrapping up 146

5.4 Visualizing what convnets learn 146

Visualizing intermediate activations 146 • *Visualizing convnet filters* 153 • *Visualizing heatmaps of class activation* 159

5.5 Summary 163

***r1 Deep learning for text and sequences* 164**

6.1 Working with text data 165

One-hot encoding of words and characters 166 • *Using word embeddings* 169 • *Putting it all together: from raw text to word embeddings* 174 • *Wrapping up* 180

6.2 Understanding recurrent neural networks 180

A recurrent layer in Keras 182 • *Understanding the LSTM and GRU layers* 186 • *A concrete LSTM example in Keras* 188
Wrapping up 190

6.3 Advanced use of recurrent neural networks 190

A temperature-forecasting problem 191 • *Preparing the data* 193
A common-sense, non-machine-learning baseline 197 • *A basic machine-learning approach* 198 • *A first recurrent baseline* 199
Using recurrent dropout to fight overfitting 201 • *Stacking recurrent layers* 202 • *Using bidirectional RNNs* 204
Going even further 207 • *Wrapping up* 208

6.4 Sequence processing with convnets 209

Understanding 1D convolution for sequence data 209
1D pooling for sequence data 210 • Implementing a 1D
convnet 210 • Combining CNNs and RNNs to process long
sequences 212 • Wrapping up 216

6.5 Summary 216

7 *Advanced deep-learning best practices* 218

7.1 Going beyond the sequential model: the Keras functional API 219

Introduction to the functional API 221 • Multi-input
models 222 • Multi-output models 224 • Directed acyclic
graphs of layers 227 • Layer weight sharing 231 • Models as
layers 232 • Wrapping up 233

7.2 Inspecting and monitoring deep-learning models using Keras callbacks and TensorBoard 233

Using callbacks to act on a model during training 233
Introduction to TensorBoard: the TensorFlow visualization
framework 236 • Wrapping up 241

7.3 Getting the most out of your models 241

Advanced architecture patterns 241 • Hyperparameter
optimization 245 • Model ensembling 246 • Wrapping
up 248

7.4 Summary 249

Generative deep learning 250

8.1 Text generation with LSTM 252

A brief history of generative recurrent networks 252
How do you generate sequence data? 253 • The importance
of the sampling strategy 253 • Implementing character-level
LSTM text generation 255 • Wrapping up 260

8.2 DeepDream 260

Implementing DeepDream in Keras 261 • Wrapping up 261

8.3 Neural style transfer 267

The content loss 268 • The style loss 268 • Neural style transfer
in Keras 269 • Wrapping up 274

8.4 Generating images with variational autoencoders 276

Sampling from latent spaces of images 276 • Concept vectors for
image editing 277 • Variational autoencoders 278
Wrapping up 284

8.5 Introduction to generative adversarial networks 284

A schematic GAN implementation 286 • A bag of tricks 286
The generator 287 • The discriminator 288
The adversarial network 289 • How to train your
DCGAN 290 • Wrapping up 292

8.6 Summary 292

C) Conclusions 293

J

9.1 Key concepts in review 294

Various approaches to AI 294 • What makes deep learning special
within the field of machine learning 294 • How to think about
deep learning 295 • Key enabling technologies 296
The universal machine-learning workflow 297 • Key network
architectures 298 • The space of possibilities 302

9.2 The limitations of deep learning 303

The risk of anthropomorphizing machine-learning models 304
Local generalization vs. extreme generalization 306
Wrapping up 307

9.3 The future of deep learning 307

Models as programs 308 • Beyond backpropagation and
differentiable layers 310 • Automated machine learning 310
Lifelong learning and modular subroutine reuse 311 • The long-
term vision 313

9.4 Staying up to date in a fast-moving field 313

Practice on real-world problems using Kaggle 314 • Read about
the latest developments on arXiv 314 • Explore the Keras
ecosystem 315

9.5 Final words 315

appendix A Installing Keras and its dependencies on Ubuntu 316

appendix B Running RStudio Server on an EC2 GPU instance 320

index 327