

Contents

<i>List of Figures</i>	<i>page</i> x
<i>List of Tables</i>	xv
<i>Preface</i>	xvii
1 Introduction	1
1.1 What This Book Covers	3
1.2 What This Book Does Not Cover	5
1.3 Deep Learning Is Not Perfect	5
1.4 Mathematical Notations	6
2 The Perceptron	8
2.1 Machine Learning Is Easy	8
2.2 Use Case: Text Classification	11
2.3 Evaluation Measures for Text Classification	12
2.4 The Perceptron	14
2.5 Voting Perceptron	22
2.6 Average Perceptron	24
2.7 Drawbacks of the Perceptron	26
2.8 Historical Background	28
2.9 References and Further Readings	29
2.10 Summary	29
3 Logistic Regression	30
3.1 The Logistic Regression Decision Function and Learning Algorithm	30
3.2 The Logistic Regression Cost Function	32
3.3 Gradient Descent	34
3.4 Deriving the Logistic Regression Update Rule	38
3.5 From Binary to Multiclass Classification	40
3.6 Evaluation Measures for Multiclass Text Classification	43
	v

vi	Contents	
3.7	Drawbacks of Logistic Regression	46
3.8	Historical Background	46
3.9	References and Further Readings	47
3.10	Summary	48
4	Implementing Text Classification Using Perceptron and Logistic Regression	49
4.1	Binary Classification	49
4.2	Multiclass Classification	62
4.3	Summary	72
5	Feed-Forward Neural Networks	73
5.1	Architecture of Feed-Forward Neural Networks	73
5.2	Learning Algorithm for Neural Networks	77
5.3	The Equations of Backpropagation	79
5.4	Drawbacks of Neural Networks (So Far)	85
5.5	Historical Background	85
5.6	References and Further Readings	86
5.7	Summary	86
6	Best Practices in Deep Learning	87
6.1	Minibatching	87
6.2	Other Optimization Algorithms	91
6.3	Other Activation Functions	94
6.4	Cost Functions	97
6.5	Regularization	99
6.6	Dropout	101
6.7	Temporal Averaging	102
6.8	Parameter Initialization and Normalization	103
6.9	References and Further Readings	105
6.10	Summary	106
7	Implementing Text Classification with Feed-Forward Networks	107
7.1	Data	108
7.2	Fully Connected Neural Network	109
7.3	Training	111
7.4	Summary	115
8	Distributional Hypothesis and Representation Learning	117
8.1	Traditional Distributional Representations	117

Contents	vii
8.2 Matrix Decompositions and Low-Rank Approximations	120
8.3 Drawbacks of Representation Learning Using Low-Rank Approximation	123
8.4 The Word2vec Algorithm	123
8.5 Drawbacks of the Word2vec Algorithm	128
8.6 Historical Background	129
8.7 References and Further Readings	130
8.8 Summary	131
9 Implementing Text Classification Using Word Embeddings	132
9.1 Pretrained Word Embeddings	132
9.2 Text Classification with Pretrained Word Embeddings	140
9.3 Summary	146
10 Recurrent Neural Networks	147
10.1 Vanilla Recurrent Neural Networks	148
10.2 Deep Recurrent Neural Networks	150
10.3 The Problem with Simple Recurrent Neural Networks: Vanishing Gradient	151
10.4 Long Short-Term Memory Networks	152
10.5 Conditional Random Fields	155
10.6 Drawbacks of Recurrent Neural Networks	163
10.7 Historical Background	163
10.8 References and Further Readings	164
10.9 Summary	164
11 Implementing Part-of-Speech Tagging Using Recurrent Neural Networks	165
11.1 Part-of-Speech Tagging	165
11.2 Summary	177
12 Contextualized Embeddings and Transformer Networks	178
12.1 Architecture of a Transformer Layer	179
12.2 Subword Tokenization	186
12.3 Training a Transformer Network	188
12.4 Drawbacks of Transformer Networks	190
12.5 Historical Background	191
12.6 References and Further Readings	192
12.7 Summary	193

viii Contents

13	Using Transformers with the Hugging Face Library	194
13.1	Tokenization	194
13.2	Text Classification	196
13.3	Part-of-Speech Tagging	204
13.4	Summary	215
14	Encoder-Decoder Methods	216
14.1	BLEU: An Evaluation Measure for Machine Translation	217
14.2	A First Sequence-to-Sequence Architecture	219
14.3	Sequence-to-Sequence with Attention	221
14.4	Transformer-Based Encoder-Decoder Architectures	224
14.5	Drawbacks of Encoder-Decoder Methods	227
14.6	Historical Background	227
14.7	References and Further Readings	227
14.8	Summary	228
15	Implementing Encoder-Decoder Methods	229
15.1	Translating English to Romanian	229
15.2	Implementation of Greedy Generation	235
15.3	Fine-Tuning Romanian to English Translation	237
15.4	Using a Previously Saved Model	244
15.5	Summary	244
16	Neural Architectures for Natural Language Processing Applications	246
16.1	Text Classification	246
16.2	Part-of-Speech Tagging	248
16.3	Named Entity Recognition	252
16.4	Dependency Parsing	255
16.5	Relation Extraction	260
16.6	Question Answering	264
16.7	Machine Translation	269
16.8	Summary	271
<i>Appendix A Overview of the Python Language and Key Libraries</i>		272
A.1	Python	272
A.2	NumPy	286
A.3	PyTorch	296

Contents	ix
<i>Appendix B</i> Character Encodings: ASCII and Unicode	301
B.1 How Do Computers Represent Text?	301
B.2 How to Encode/Decode Characters in Python	304
B.3 Text Normalization	305
<i>References</i>	308
<i>Index</i>	321