

Contents

Preface	ix
Notation	xiii
1 Introduction	1
1.1 Natural Language Processing and Its Neighbors	1
1.2 Three Themes in Natural Language Processing	5
I LEARNING	11
2 Linear Text Classification	13
2.1 The Bag of Words	13
2.2 Naïve Bayes	17
2.3 Discriminative Learning	24
2.4 Loss Functions and Large-Margin Classification	28
2.5 Logistic Regression	34
2.6 Optimization	37
2.7 *Additional Topics in Classification	40
2.8 Summary of Learning Algorithms	42
3 Nonlinear Classification	47
3.1 Feedforward Neural Networks	48
3.2 Designing Neural Networks	50
3.3 Learning Neural Networks	53
3.4 Convolutional Neural Networks	61
4 Linguistic Applications of Classification	67
4.1 Sentiment and Opinion Analysis	67
4.2 Word Sense Disambiguation	71
4.3 Design Decisions for Text Classification	74
4.4 Evaluating Classifiers	78
4.5 Building Datasets	85

5	Learning Without Supervision	91
5.1	Unsupervised Learning	91
5.2	Applications of Expectation-Maximization	99
5.3	Semi-Supervised Learning	102
5.4	Domain Adaptation	105
5.5	*Other Approaches to Learning with Latent Variables	109
II	SEQUENCES AND TREES	117
6	Language Models	119
6.1	<i>N</i> -Gram Language Models	120
6.2	Smoothing and Discounting	122
6.3	Recurrent Neural Network Language Models	127
6.4	Evaluating Language Models	132
6.5	Out-of-Vocabulary Words	134
7	Sequence Labeling	137
7.1	Sequence Labeling as Classification	137
7.2	Sequence Labeling as Structure Prediction	139
7.3	The Viterbi Algorithm	140
7.4	Hidden Markov Models	145
7.5	Discriminative Sequence Labeling with Features	149
7.6	Neural Sequence Labeling	158
7.7	*Unsupervised Sequence Labeling	161
8	Applications of Sequence Labeling	167
8.1	Part-of-Speech Tagging	167
8.2	Morphosyntactic Attributes	173
8.3	Named Entity Recognition	175
8.4	Tokenization	176
8.5	Code Switching	177
8.6	Dialogue Acts	178
9	Formal Language Theory	183
9.1	Regular Languages	184
9.2	Context-Free Languages	198
9.3	*Mildly Context-Sensitive Languages	209
10	Context-Free Parsing	215
10.1	Deterministic Bottom-Up Parsing	216
10.2	Ambiguity	219
10.3	Weighted Context-Free Grammars	222
10.4	Learning Weighted Context-Free Grammars	227
10.5	Grammar Refinement	231
10.6	Beyond Context-Free Parsing	238

-1__
0__
+1__

11	Dependency Parsing	243
11.1	Dependency Grammar	243
11.2	Graph-Based Dependency Parsing	248
11.3	Transition-Based Dependency Parsing	253
11.4	Applications	261
III	MEANING	267
12	Logical Semantics	269
12.1	Meaning and Denotation	270
12.2	Logical Representations of Meaning	270
12.3	Semantic Parsing and the Lambda Calculus	274
12.4	Learning Semantic Parsers	280
13	Predicate-Argument Semantics	289
13.1	Semantic Roles	291
13.2	Semantic Role Labeling	295
13.3	Abstract Meaning Representation	302
14	Distributional and Distributed Semantics	309
14.1	The Distributional Hypothesis	309
14.2	Design Decisions for Word Representations	311
14.3	Latent Semantic Analysis	313
14.4	Brown Clusters	315
14.5	Neural Word Embeddings	317
14.6	Evaluating Word Embeddings	322
14.7	Distributed Representations beyond Distributional Statistics	324
14.8	Distributed Representations of Multiword Units	327
15	Reference Resolution	333
15.1	Forms of Referring Expressions	334
15.2	Algorithms for Coreference Resolution	339
15.3	Representations for Coreference Resolution	348
15.4	Evaluating Coreference Resolution	353
16	Discourse	357
16.1	Segments	357
16.2	Entities and Reference	359
16.3	Relations	362
IV	APPLICATIONS	377
17	Information Extraction	379
17.1	Entities	381
17.2	Relations	387

17.3	Events	395
17.4	Hedges, Denials, and Hypotheticals	397
17.5	Question Answering and Machine Reading	399
18	Machine Translation	405
18.1	Machine Translation as a Task	405
18.2	Statistical Machine Translation	410
18.3	Neural Machine Translation	415
18.4	Decoding	423
18.5	Training toward the Evaluation Metric	424
19	Text Generation	431
19.1	Data-to-Text Generation	431
19.2	Text-to-Text Generation	437
19.3	Dialogue	440
	Appendix A: Probability	447
A.1	Probabilities of Event Combinations	447
A.2	Conditional Probability and Bayes' Rule	449
A.3	Independence	451
A.4	Random Variables	451
A.5	Expectations	452
A.6	Modeling and Estimation	453
	Appendix B: Numerical Optimization	455
B.1	Gradient Descent	456
B.2	Constrained Optimization	456
B.3	Example: Passive-Aggressive Online Learning	457
	Bibliography	459
	Index	509