

Contents

<i>Preface</i>	<i>page xi</i>
<i>Acknowledgements</i>	xv
<i>Notation</i>	xvi
Part I Basics	1
1 Introduction	3
1.1 What is NLP?	3
1.2 NLP Tasks	5
1.3 NLP from a Machine Learning Perspective	21
Summary	22
Chapter Notes	22
Exercises	23
2 Counting Relative Frequencies	25
2.1 Probabilistic Modelling	25
2.2 n -gram Language Models	32
2.3 A Probabilistic Model for Text Classification	41
Summary	47
Chapter Notes	47
Exercises	48
3 Feature Vectors	50
3.1 Modelling Documents in Vector Spaces	50
3.2 Multi-Class Classification	61
3.3 Discriminative Linear Models	65
3.4 Vector Spaces and Model Training	67
Summary	69
Chapter Notes	69
Exercises	69

4 Discriminative Linear Classifiers	73
4.1 Log-Linear Models	73
4.2 SGD Training of SVMs	81
4.3 A Generalised Linear Model	84
4.4 Choosing and Combining Models	89
Summary	94
Chapter Notes	94
Exercises	94
5 A Perspective from Information Theory	98
5.1 The Maximum Entropy Principle	98
5.2 KL-Divergence and Cross-Entropy	105
5.3 Mutual Information	109
Summary	116
Chapter Notes	116
Exercises	116
6 Hidden Variables	120
6.1 Expectation Maximisation	120
6.2 Using EM for Training Models with Hidden Variables	127
6.3 Theory behind EM	139
Summary	142
Chapter Notes	142
Exercises	142
Part II Structures	145
7 Generative Sequence Labelling	147
7.1 Sequence Labelling	147
7.2 Hidden Markov Models	148
7.3 Finding Marginal Probabilities	154
7.4 EM for Unsupervised HMM Training	159
Summary	166
Chapter Notes	166
Exercises	167
8 Discriminative Sequence Labelling	169
8.1 Locally Trained Models for Discriminative Sequence Labelling	169
8.2 The Label Bias Problem	172
8.3 Conditional Random Fields	174
8.4 Structured Perceptron	182

8.5 Structured SVM	185
Summary	187
Chapter Notes	188
Exercises	188
9 Sequence Segmentation	190
9.1 Segmentation by Sequence Labelling	190
9.2 Discriminative Models for Sequence Segmentation	195
9.3 Structured Perceptron and Beam Search	206
Summary	209
Chapter Notes	210
Exercises	210
10 Predicting Tree Structures	212
10.1 Generative Constituent Parsing	212
10.2 More Features for Constituent Parsing	221
10.3 Reranking	228
10.4 Beyond Sequences and Trees	231
Summary	231
Chapter Notes	231
Exercises	232
11 Transition-Based Methods for Structured Prediction	235
11.1 Transition-Based Structured Prediction	235
11.2 Transition-Based Constituent Parsing	242
11.3 Transition-Based Dependency Parsing	246
11.4 Joint Parsing Models	251
Summary	256
Chapter Notes	256
Exercises	257
12 Bayesian Network	259
12.1 A General Probabilistic Model	259
12.2 Training Bayesian Networks	262
12.3 Inference	272
12.4 Latent Dirichlet Allocation	275
12.5 Bayesian IBM Model 1	283
Summary	284
Chapter Notes	284
Exercises	284

Part III Deep Learning	287
13 Neural Network	289
13.1 From One Layer to Multiple Layers	289
13.2 Building a Text Classifier without Manual Features	300
13.3 Improving Neural Network Training	306
Summary	310
Chapter Notes	310
Exercises	311
14 Representation Learning	314
14.1 Recurrent Neural Network	314
14.2 Neural Attention	320
14.3 Representing Trees	325
14.4 Representing Graphs	329
14.5 Analysing Representation	333
14.6 More on Neural Network Training	334
Summary	338
Chapter Notes	338
Exercises	339
15 Neural Structured Prediction	343
15.1 Local Graph-Based Models	343
15.2 Local Transition-Based Models	351
15.3 Global Structured Models	358
Summary	367
Chapter Notes	367
Exercises	367
16 Working with Two Texts	370
16.1 Sequence-to-Sequence Models	370
16.2 Text Matching Models	382
Summary	392
Chapter Notes	392
Exercises	393
17 Pre-training and Transfer Learning	396
17.1 Neural Language Models and Word Embedding	396
17.2 Contextualised Word Representations	409
17.3 Transfer Learning	414

Summary	419
Chapter Notes	420
Exercises	420
18 Deep Latent Variable Models	423
18.1 Introducing Latent Variables into a Neural Network Model	423
18.2 Working with Categorical Latent Variables	424
18.3 Working with Structured Latent Variables	429
18.4 Variational Inference	434
18.5 Neural Topic Models	443
18.6 VAEs for Language Modelling	444
Summary	450
Chapter Notes	450
Exercises	450
<i>Bibliography</i>	453
<i>Index</i>	468