

Contents

<i>Preface</i>	<i>page xi</i>
PART ONE FUNDAMENTALS	1
1 Introduction	3
1.1 Feature Engineering	6
1.2 Evaluation	10
1.3 Cycles	15
1.4 Analysis	20
1.5 Other Processes	25
1.6 Discussion	30
1.7 Learning More	32
2 Features, Combined: Normalization, Discretization and Outliers	34
2.1 Normalizing Features	35
2.2 Discretization and Binning	43
2.3 Descriptive Features	50
2.4 Dealing with Outliers	54
2.5 Advanced Techniques	56
2.6 Learning More	57
3 Features, Expanded: Computable Features, Imputation and Kernels	59
3.1 Computable Features	60
3.2 Imputation	67
3.3 Decomposing Complex Features	70
3.4 Kernel-Induced Feature Expansion	73
3.5 Learning More	78

4	Features, Reduced: Feature Selection, Dimensionality Reduction and Embeddings	79
4.1	Feature Selection	80
4.2	Regularization and Embedded Feature Selection	94
4.3	Dimensionality Reduction	99
4.4	Learning More	111
5	Advanced Topics: Variable-Length Data and Automated Feature Engineering	112
5.1	Variable-Length Feature Vectors	112
5.2	Instance-Based Engineering	124
5.3	Deep Learning and Feature Engineering	127
5.4	Automated Feature Engineering	130
5.5	Learning More	135
	PART TWO CASE STUDIES	137
6	Graph Data	139
6.1	WikiCities Dataset	142
6.2	Exploratory Data Analysis (EDA)	144
6.3	First Feature Set	150
6.4	Second Feature Set	158
6.5	Final Feature Sets	160
6.6	Learning More	162
7	Timestamped Data	163
7.1	WikiCities: Historical Features	166
7.2	Time Lagged Features	169
7.3	Sliding Windows	172
7.4	Third Featurization: EMA	173
7.5	Historical Data as Data Expansion	174
7.6	Time Series	176
7.7	Learning More	183
8	Textual Data	186
8.1	WikiCities: Text	189
8.2	Exploratory Data Analysis	190
8.3	Numeric Tokens Only	194
8.4	Bag-of-Words	196
8.5	Stop Words and Morphological Features	200
8.6	Features in Context	203
8.7	Skip Bigrams and Feature Hashing	204

Contents	ix
8.8 Dimensionality Reduction and Embeddings	205
8.9 Closing Remarks	208
8.10 Learning More	211
9 Image Data	212
9.1 WikiCities: Satellite Images	215
9.2 Exploratory Data Analysis	216
9.3 Pixels as Features	217
9.4 Automatic Dataset Expansion	222
9.5 Descriptive Features: Histograms	223
9.6 Local Feature Detectors: Corners	225
9.7 Dimensionality Reduction: HOGs	227
9.8 Closing Remarks	228
9.9 Learning More	231
10 Other Domains: Video, GIS and Preferences	233
10.1 Video	234
10.2 Geographical Features	239
10.3 Preferences	242
<i>Bibliography</i>	246
<i>Index</i>	270