# Contents

- **Preface**
  - page xi

- **PART ONE FUNDAMENTALS**
  - 1 Introduction
    - 1.1 Feature Engineering
    - 1.2 Evaluation
    - 1.3 Cycles
    - 1.4 Analysis
    - 1.5 Other Processes
    - 1.6 Discussion
    - 1.7 Learning More
  - 2 Features, Combined: Normalization, Discretization and Outliers
    - 2.1 Normalizing Features
    - 2.2 Discretization and Binning
    - 2.3 Descriptive Features
    - 2.4 Dealing with Outliers
    - 2.5 Advanced Techniques
    - 2.6 Learning More
  - 3 Features, Expanded: Computable Features, Imputation and Kernels
    - 3.1 Computable Features
    - 3.2 Imputation
    - 3.3 Decomposing Complex Features
    - 3.4 Kernel-Induced Feature Expansion
    - 3.5 Learning More
# Table of Contents

4  **Features, Reduced: Feature Selection, Dimensionality Reduction and Embeddings**  
4.1 Feature Selection  
4.2 Regularization and Embedded Feature Selection  
4.3 Dimensionality Reduction  
4.4 Learning More  

5  **Advanced Topics: Variable-Length Data and Automated Feature Engineering**  
5.1 Variable-Length Feature Vectors  
5.2 Instance-Based Engineering  
5.3 Deep Learning and Feature Engineering  
5.4 Automated Feature Engineering  
5.5 Learning More  

PART TWO  **CASE STUDIES**  

6  **Graph Data**  
6.1 WikiCities Dataset  
6.2 Exploratory Data Analysis (EDA)  
6.3 First Feature Set  
6.4 Second Feature Set  
6.5 Final Feature Sets  
6.6 Learning More  

7  **Timestamped Data**  
7.1 WikiCities: Historical Features  
7.2 Time Lagged Features  
7.3 Sliding Windows  
7.4 Third Featurization: EMA  
7.5 Historical Data as Data Expansion  
7.6 Time Series  
7.7 Learning More  

8  **Textual Data**  
8.1 WikiCities: Text  
8.2 Exploratory Data Analysis  
8.3 Numeric Tokens Only  
8.4 Bag-of-Words  
8.5 Stop Words and Morphological Features  
8.6 Features in Context  
8.7 Skip Bigrams and Feature Hashing
## Contents

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

### Bibliography

246

### Index

270