

# Contents

<b>1</b>	<b>Introduction</b>	1
1.1	A Natural Representation of Data Similarity	3
1.2	Goals	8
1.3	Outline	10
<b>2</b>	<b>Basic Structure of High-Dimensional Spaces</b>	13
2.1	Comparing Attributes	13
2.2	Comparing Records	14
2.3	Similarity	14
2.4	High-Dimensional Spaces	16
2.5	Summary	18
<b>3</b>	<b>Algorithms</b>	19
3.1	Improving the Natural Geometry	19
3.1.1	Projection	20
3.1.2	Singular Value Decompositions	20
3.1.3	Random Projections	22
3.2	Algorithms that Find Standalone Clusters	23
3.2.1	Clusters Based on Density	23
3.2.2	Parallel Coordinates	24
3.2.3	Independent Component Analysis	24
3.2.4	Latent Dirichlet Allocation	25
3.3	Algorithms that Find Clusters and Their Relationships	25
3.3.1	Clusters Based on Distance	25
3.3.2	Clusters Based on Distribution	26
3.3.3	Semidiscrete Decomposition	27
3.3.4	Hierarchical Clustering	29
3.3.5	Minimum Spanning Tree with Collapsing	29
3.4	Overall Process for Constructing a Skeleton	30

3.5	Algorithms that Wrap Clusters . . . . .	31
3.5.1	Distance-Based . . . . .	32
3.5.2	Distribution-Based . . . . .	32
3.5.3	1-Class Support Vector Machines . . . . .	32
3.5.4	Autoassociative Neural Networks . . . . .	33
3.5.5	Covers . . . . .	34
3.6	Algorithms to Place Boundaries Between Clusters . . . . .	34
3.6.1	Support Vector Machines . . . . .	35
3.6.2	Random Forests . . . . .	35
3.7	Overall Process for Constructing Empty Space . . . . .	36
3.8	Summary . . . . .	37
<b>4</b>	<b>Spaces with a Single Center . . . . .</b>	<b>39</b>
4.1	Using Distance . . . . .	39
4.2	Using Density . . . . .	40
4.3	Understanding the Skeleton . . . . .	42
4.4	Understanding Empty Space . . . . .	43
4.5	Summary . . . . .	45
<b>5</b>	<b>Spaces with Multiple Centers . . . . .</b>	<b>47</b>
5.1	What is a Cluster? . . . . .	48
5.2	Identifying Clusters . . . . .	50
5.2.1	Clusters Known Already . . . . .	50
5.3	Finding Clusters . . . . .	50
5.4	Finding the Skeleton . . . . .	55
5.5	Empty Space . . . . .	58
5.5.1	An Outer Boundary and Novel Data . . . . .	58
5.5.2	Interesting Data . . . . .	60
5.5.3	One-Cluster Boundaries . . . . .	63
5.5.4	One-Cluster-Against-the-Rest Boundaries . . . . .	63
5.6	Summary . . . . .	64
<b>6</b>	<b>Representation by Graphs . . . . .</b>	<b>67</b>
6.1	Building a Graph from Records . . . . .	68
6.2	Local Similarities . . . . .	68
6.3	Embedding Choices . . . . .	69
6.4	Using the Embedding for Clustering . . . . .	70
6.5	Summary . . . . .	71
<b>7</b>	<b>Using Models of High-Dimensional Spaces . . . . .</b>	<b>73</b>
7.1	Understanding Clusters . . . . .	73
7.2	Structure in the Set of Clusters . . . . .	76
7.2.1	Semantic Stratified Sampling . . . . .	77
7.3	Ranking Using the Skeleton . . . . .	78

7.4	Ranking Using Empty Space. . . . .	87
7.4.1	Applications to Streaming Data . . . . .	89
7.4.2	Concealment. . . . .	90
7.5	Summary . . . . .	91
<b>8</b>	<b>Including Contextual Information. . . . .</b>	<b>93</b>
8.1	What is Context?. . . . .	93
8.1.1	Changing Data . . . . .	93
8.1.2	Changing Analyst and Organizational Properties. . . . .	94
8.1.3	Changing Algorithmic Properties . . . . .	95
8.2	Letting Context Change the Models. . . . .	95
8.2.1	Recomputing the View . . . . .	95
8.2.2	Recomputing Derived Structures. . . . .	96
8.2.3	Recomputing the Clustering . . . . .	97
8.3	Summary . . . . .	98
<b>9</b>	<b>Conclusions . . . . .</b>	<b>99</b>
	<b>References . . . . .</b>	<b>103</b>
	<b>Index . . . . .</b>	<b>107</b>



<http://www.springer.com/978-3-642-33397-2>

Understanding High-Dimensional Spaces

Skillicorn, D.B.

2012, IX, 108 p. 29 illus., Softcover

ISBN: 978-3-642-33397-2