CONTENTS

1 INTRODUCTION 1
- 1.1 Cluster analysis 2
- 1.2 Parameter estimation 5
- 1.3 Outline of the thesis 9

I ON CLUSTER ANALYSIS 11

2 CLUSTERING BASICS 13
- 2.1 The notion of similarity 14
- 2.2 Clustering quality 14
- 2.3 Problem Definitions 16

3 ANALYSIS OF AGGLOMERATIVE CLUSTERING 17
- 3.1 Preliminaries 18
- 3.2 Discrete k-center clustering 20
- 3.3 k-center clustering 24
- 3.4 Diameter k-clustering 34
- 3.5 The one-dimensional case 41
- 3.6 Lower bounds 45

II ON PARAMETER ESTIMATION 55

4 STATISTICAL MODELS 57
- 4.1 The parameter estimation problem 58
- 4.2 General mixture models 60
- 4.3 Gaussian mixture models 61
- 4.4 Handy notions from probability theory 67

5 THE CLASSICAL EM ALGORITHM 71
- 5.1 Convergence of the EM algorithm 73
- 5.2 The EM algorithm for mixture distributions 77
- 5.3 The EM algorithm for Gaussian mixtures 78

6 THE STOCHASTIC EM ALGORITHM 81
- 6.1 The generic SEM algorithm 82
- 6.2 The SEM algorithm for mixture distributions 83
- 6.3 The SEM algorithm for Gaussian mixtures 86

7 EXPERIMENTAL ANALYSIS 99
- 7.1 Implementation 99
- 7.2 Data sets 100
- 7.3 Experiments 102
- 7.4 Results 105

Bibliografische Informationen
http://d-nb.info/1048369935

digitalisiert durch DEUTSCHE NATIONAL BIBLIOTHEK