Algorithms for Fuzzy Clustering

Methods in c-Means Clustering with Applications
Recently many researchers are working on cluster analysis as a main tool for exploratory data analysis and data mining. A notable feature is that specialists in different fields of sciences are considering the tool of data clustering to be useful. A major reason is that clustering algorithms and software are flexible in the sense that different mathematical frameworks are employed in the algorithms and a user can select a suitable method according to his application. Moreover clustering algorithms have different outputs ranging from the old dendrograms of agglomerative clustering to more recent self-organizing maps. Thus, a researcher or user can choose an appropriate output suited to his purpose, which is another flexibility of the methods of clustering.

An old and still most popular method is the $K$-means which use $K$ cluster centers. A group of data is gathered around a cluster center and thus forms a cluster. The main subject of this book is the fuzzy $c$-means proposed by Dunn and Bezdek and their variations including recent studies. A main reason why we concentrate on fuzzy $c$-means is that most methodology and application studies in fuzzy clustering use fuzzy $c$-means, and fuzzy $c$-means should be considered to be a major technique of clustering in general, regardless whether one is interested in fuzzy methods or not. Moreover recent advances in clustering techniques are rapid and we require a new textbook that includes recent algorithms. We should also note that several books have recently been published but the contents do not include some methods studied herein.

Unlike most studies in fuzzy $c$-means, what we emphasize in this book is a family of algorithms using entropy or entropy-regularized methods which are less known, but we consider the entropy-based method to be another useful method of fuzzy $c$-means. For this reason we call the method of fuzzy $c$-means by Dunn and Bezdek as the standard method to distinguish it from the entropy-based method. Throughout this book one of our intentions is to uncover theoretical and methodological differences between the standard method and the entropy-based method. We do not claim that the entropy-based method is better than the standard method, but we believe that the methods of fuzzy $c$-means become complete by adding the entropy-based method to the standard one by Dunn.
and Bezdek, since we can observe natures of the both methods more deeply by contrasting these two methods.

Readers will observe that the entropy-based method is similar to the statistical model of Gaussian mixture distribution since both of them are using the error functions, while the standard method is very different from a statistical model. For this reason the standard method is purely fuzzy while the entropy-based method connects a statistical model and a fuzzy model.

The whole text is divided into two parts: The first part that consists of Chapters 1∼5 is theoretical and discusses basic algorithms and variations. This part has been written by Sadaaki Miyamoto.

The second part is application-oriented. Chapter 6 which has been written by Hidetomo Ichihashi studies classifier design; Katsuhiro Honda has written Chapters 7∼9 where clustering algorithms are applied to a variety of methods in multivariate analysis.

The authors are grateful to Prof. Janusz Kacprzyk, the editor, for his encouragement to contribute this volume to this series and helpful suggestions throughout the publication process. We also thank Dr. Mika Sato-Ilic and Dr. Yasumori Endo for their valuable comments to our works.

We finally note that studies related to this book have partly been supported by the Grant-in-Aid for Scientific Research, Japan Society for the Promotion of Science, No.16300065.

January 2008

Sadaaki Miyamoto
Hidetomo Ichihashi
Katsuhiro Honda
# Contents

1 Introduction ................................................ 1  
   1.1 Fuzziness and Neural Networks in Clustering .......... 3  
   1.2 An Illustrative Example ............................... 4  

2 Basic Methods for c-Means Clustering .................. 9  
   2.1 A Note on Terminology .................................. 9  
   2.2 A Basic Algorithm of c-Means .......................... 11  
   2.3 Optimization Formulation of Crisp c-Means Clustering .. 12  
   2.4 Fuzzy c-Means ........................................ 16  
   2.5 Entropy-Based Fuzzy c-Means ........................... 20  
   2.6 Addition of a Quadratic Term .......................... 23  
       2.6.1 Derivation of Algorithm in the Method of the  
            Quadratic Term .................................. 24  
   2.7 Fuzzy Classification Rules ................................ 25  
   2.8 Clustering by Competitive Learning .................... 29  
   2.9 Fixed Point Iterations – General Consideration .......... 30  
   2.10 Heuristic Algorithms of Fixed Point Iterations ....... 31  
   2.11 Direct Derivation of Classification Functions ......... 33  
   2.12 Mixture Density Model and the EM Algorithm ........... 36  
       2.12.1 The EM Algorithm .............................. 37  
       2.12.2 Parameter Estimation in the Mixture Densities ...... 39  

3 Variations and Generalizations - I ..................... 43  
   3.1 Possibilistic Clustering ................................. 43  
       3.1.1 Entropy-Based Possibilistic Clustering ............ 44  
       3.1.2 Possibilistic Clustering Using a Quadratic Term .... 46  
       3.1.3 Objective Function for Fuzzy c-Means and Possibilistic  
            Clustering ....................................... 46  
   3.2 Variables for Controlling Cluster Sizes ................ 47  
       3.2.1 Solutions for $J_{efca}(U, V, A)$ .................. 50  
       3.2.2 Solutions for $J_{fema}(U, V, A)$ .................. 50
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3</td>
<td>Covariance Matrices within Clusters</td>
<td>51</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Solutions for <strong>FCMAS</strong> by the GK(Gustafson-Kessel) Method</td>
<td>53</td>
</tr>
<tr>
<td>3.4</td>
<td>The KL (Kullback-Leibler) Information Based Method</td>
<td>55</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Solutions for <strong>FCMAS</strong> by the Method of KL Information Method</td>
<td>55</td>
</tr>
<tr>
<td>3.5</td>
<td>Defuzzified Methods of $c$-Means Clustering</td>
<td>56</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Defuzzified $c$-Means with Cluster Size Variable</td>
<td>57</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Defuzzification of the KL-Information Based Method</td>
<td>58</td>
</tr>
<tr>
<td>3.5.3</td>
<td>Sequential Algorithm</td>
<td>58</td>
</tr>
<tr>
<td>3.5.4</td>
<td>Efficient Calculation of Variables</td>
<td>59</td>
</tr>
<tr>
<td>3.6</td>
<td>Fuzzy $c$-Varieties</td>
<td>60</td>
</tr>
<tr>
<td>3.6.1</td>
<td>Multidimensional Linear Varieties</td>
<td>62</td>
</tr>
<tr>
<td>3.7</td>
<td>Fuzzy $c$-Regression Models</td>
<td>62</td>
</tr>
<tr>
<td>3.8</td>
<td>Noise Clustering</td>
<td>65</td>
</tr>
<tr>
<td>4</td>
<td>Variations and Generalizations - II</td>
<td>67</td>
</tr>
<tr>
<td>4.1</td>
<td>Kernelized Fuzzy $c$-Means Clustering and Related Methods</td>
<td>67</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Transformation into High-Dimensional Feature Space</td>
<td>68</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Kernelized Crisp $c$-Means Algorithm</td>
<td>71</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Kernelized Learning Vector Quantization Algorithm</td>
<td>73</td>
</tr>
<tr>
<td>4.1.4</td>
<td>An Illustrative Example</td>
<td>74</td>
</tr>
<tr>
<td>4.2</td>
<td>Similarity Measure in Fuzzy $c$-Means</td>
<td>77</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Variable for Controlling Cluster Sizes</td>
<td>80</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Kernelization Using Cosine Correlation</td>
<td>81</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Clustering by Kernelized Competitive Learning Using Cosine Correlation</td>
<td>84</td>
</tr>
<tr>
<td>4.3</td>
<td>Fuzzy $c$-Means Based on $L_1$ Metric</td>
<td>86</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Finite Termination Property of the $L_1$ Algorithm</td>
<td>88</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Classification Functions in the $L_1$ Case</td>
<td>89</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Boundary between Two Clusters in the $L_1$ Case</td>
<td>90</td>
</tr>
<tr>
<td>4.4</td>
<td>Fuzzy $c$-Regression Models Based on Absolute Deviation</td>
<td>91</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Termination of Algorithm Based on Least Absolute Deviation</td>
<td>93</td>
</tr>
<tr>
<td>4.4.2</td>
<td>An Illustrative Example</td>
<td>96</td>
</tr>
<tr>
<td>5</td>
<td>Miscellanea</td>
<td>99</td>
</tr>
<tr>
<td>5.1</td>
<td>More on Similarity and Dissimilarity Measures</td>
<td>99</td>
</tr>
<tr>
<td>5.2</td>
<td>Other Methods of Fuzzy Clustering</td>
<td>100</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Ruspini’s Method</td>
<td>100</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Relational Clustering</td>
<td>101</td>
</tr>
<tr>
<td>5.3</td>
<td>Agglomerative Hierarchical Clustering</td>
<td>102</td>
</tr>
<tr>
<td>5.3.1</td>
<td>The Transitive Closure of a Fuzzy Relation and the Single Link</td>
<td>106</td>
</tr>
<tr>
<td>5.4</td>
<td>A Recent Study on Cluster Validity Functions</td>
<td>108</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Two Types of Cluster Validity Measures</td>
<td>108</td>
</tr>
</tbody>
</table>
5.4.2 Kernelized Measures of Cluster Validity .................. 110
5.4.3 Traces of Covariance Matrices .......................... 110
5.4.4 Kernelized Xie-Beni Index ................................. 111
5.4.5 Evaluation of Algorithms ................................. 111
5.5 Numerical Examples ........................................ 112
5.5.1 The Number of Clusters ................................ 112
5.5.2 Robustness of Algorithms ................................ 117

6 Application to Classifier Design ............................... 119
6.1 Unsupervised Clustering Phase ................................ 119
  6.1.1 A Generalized Objective Function ......................... 120
  6.1.2 Connections with $k$-Harmonic Means .................... 123
  6.1.3 Graphical Comparisons .................................. 125
6.2 Clustering with Iteratively Reweighted Least Square Technique .................................................. 130
6.3 FCM Classifier ................................................ 133
  6.3.1 Parameter Optimization with CV Protocol and Deterministic Initialization .................. 134
  6.3.2 Imputation of Missing Values ............................ 136
  6.3.3 Numerical Experiments .................................. 139
6.4 Receiver Operating Characteristics .................................................. 144
6.5 Fuzzy Classifier with Crisp $c$-Means Clustering ............. 150
  6.5.1 Crisp Clustering and Post-supervising .................... 150
  6.5.2 Numerical Experiments .................................. 153

7 Fuzzy Clustering and Probabilistic PCA Model .................. 157
7.1 Gaussian Mixture Models and FCM-Type Fuzzy Clustering ... 157
  7.1.1 Gaussian Mixture Models ................................ 157
  7.1.2 Another Interpretation of Mixture Models ................ 159
  7.1.3 FCM-Type Counterpart of Gaussian Mixture Models ...... 160
7.2 Probabilistic PCA Mixture Models and Regularized Fuzzy Clustering .............................................. 162
  7.2.1 Probabilistic Models for Principal Component Analysis .................................................. 162
  7.2.2 Linear Fuzzy Clustering with Regularized Objective Function .............................................. 164
  7.2.3 An Illustrative Example .................................. 167

8 Local Multivariate Analysis Based on Fuzzy Clustering ....... 171
8.1 Switching Regression and Fuzzy $c$-Regression Models ....... 171
  8.1.1 Linear Regression Model ................................ 171
  8.1.2 Switching Linear Regression by Standard Fuzzy $c$-Regression Models .......................... 174
  8.1.3 Local Regression Analysis with Centered Data Model ... 175
  8.1.4 Connection of the Two Formulations ...................... 177
  8.1.5 An Illustrative Example .................................. 177
8.2 Local Principal Component Analysis and Fuzzy c-Varieties .... 179
8.2.1 Several Formulations for Principal Component Analysis ................... 179
8.2.2 Local PCA Based on Fitting Low-Dimensional Subspace ...................... 182
8.2.3 Linear Clustering with Variance Measure of Latent Variables ............... 183
8.2.4 Local PCA Based on Lower Rank Approximation of Data Matrix ................ 184
8.2.5 Local PCA Based on Regression Model ............................................. 186
8.3 Fuzzy Clustering-Based Local Quantification of Categorical Variables ............ 188
8.3.1 Homogeneity Analysis ................................................................. 188
8.3.2 Local Quantification Method and FCV Clustering of Categorical Data ........ 190
8.3.3 Application to Classification of Variables ......................................... 192
8.3.4 An Illustrative Example ................................................................. 193

9 Extended Algorithms for Local Multivariate Analysis ............ 195
9.1 Clustering of Incomplete Data .............................................................. 195
9.1.1 FCM Clustering of Incomplete Data Including Missing Values .................. 195
9.1.2 Linear Fuzzy Clustering with Partial Distance Strategy ....................... 197
9.1.3 Linear Fuzzy Clustering with Optimal Completion Strategy ................... 199
9.1.4 Linear Fuzzy Clustering with Nearest Prototype Strategy ...................... 201
9.1.5 A Comparative Experiment .............................................................. 202
9.2 Component-Wise Robust Clustering ..................................................... 202
9.2.1 Robust Principal Component Analysis ............................................. 203
9.2.2 Robust Local Principal Component Analysis .................................... 203
9.2.3 Handling Missing Values and Application to Missing Value Estimation .......... 207
9.2.4 An Illustrative Example ................................................................. 207
9.2.5 A Potential Application: Collaborative Filtering ............................. 208
9.3 Local Minor Component Analysis Based on Least Absolute Deviations ............. 211
9.3.1 Calculation of Optimal Local Minor Component Vectors ...................... 211
9.3.2 Calculation of Optimal Cluster Centers .......................................... 214
9.3.3 An Illustrative Example ................................................................. 215
9.4 Local PCA with External Criteria ....................................................... 216
9.4.1 Principal Components Uncorrelated with External Criteria .................... 216
9.4.2 Local PCA with External Criteria .................................................... 219
9.5 Fuzzy Local Independent Component Analysis .................. 220
   9.5.1 ICA Formulation and Fast ICA Algorithm ............... 221
   9.5.2 Fuzzy Local ICA with FCV Clustering ................. 222
   9.5.3 An Illustrative Example .................................. 224
9.6 Fuzzy Local ICA with External Criteria ..................... 226
   9.6.1 Extraction of Independent Components Uncorrelated
to External Criteria ............................................. 226
   9.6.2 Extraction of Local Independent Components
Uncorrelated to External Criteria ............................. 227
9.7 Fuzzy Clustering-Based Variable Selection in Local PCA ...... 228
   9.7.1 Linear Fuzzy Clustering with Variable Selection .. 228
   9.7.2 Graded Possibilistic Variable Selection .................. 231
   9.7.3 An Illustrative Example .................................. 232
References ........................................................................ 235
Index .............................................................................. 245