

Contents

1	Introduction	1
1.1	What We Can Know	2
1.1.1	Data and Knowledge	2
1.1.2	Generality of Knowledge	3
1.1.3	Comprehensibility	5
1.2	Data Mining	6
1.2.1	Qualitative and Quantitative Modeling	7
1.2.2	Fuzzy Data Analysis	9
1.2.3	Practical Applications	10
1.3	Organization of Work	13
2	Formal Foundations	15
2.1	Fuzzy Logic	16
2.1.1	Fuzzy Sets	16
2.1.2	Logical Operations	19
2.1.3	Special Fuzzy Sets	23
2.1.4	Defuzzification	29
2.2	Metrics	30
2.3	Fuzzy Relation	32

2.3.1	Binary Fuzzy Relations	32
2.3.2	Fuzzy Equality Relations	34
2.3.3	Fuzzy Ordering Relations	37
2.3.4	Fuzzy Relations on Fuzzy Sets	40
2.4	Fuzzy Predicates	41
2.4.1	Introduction	42
2.4.2	Categorical Predicates	43
2.4.3	Numerical Predicates	45
2.4.4	Relational Predicates	46
2.4.5	Compound Predicates	48
2.5	Fuzzy Quantifier	48
2.6	Fuzzy Inference	49
2.6.1	Introduction	49
2.6.2	Inference Models	55
2.7	Fuzzy Decision Trees	58
2.8	A General Purpose First Order Language	61
2.8.1	A General Grammar for Fuzzy Expression	62
2.8.2	A General Purpose First Order Language	65
3	Preprocessing	67
3.1	Data Cleaning	69
3.1.1	Missing Values	69
3.1.2	Outliers	70
3.2	Complexity Reduction	71
3.2.1	Feature Selection	72
3.2.2	Vector Quantization	77
3.2.3	Self-Organizing Maps	78

3.3	Data Interpretation	81
3.3.1	Real Valued Attributes	83
3.3.2	Combined Categorical Predicates	88
3.3.3	Relational Predicates on Different Domains	90
3.4	Data Enhancement	91
3.4.1	FCM Clustering	92
3.4.2	On the Optimal Number of Clusters	96
3.4.3	Creating Cluster Predicates	98
4	Inductive Learning	101
4.1	Introduction	102
4.1.1	Basis Setting	103
4.1.2	General Problem Statement	103
4.1.3	Rule Evaluation Measures	104
4.2	Types of Learning Problems	105
4.2.1	Classification Problems	105
4.2.2	Regression Learning	109
4.2.3	Linguistic Descriptions	113
4.3	Learning Methods	118
4.3.1	Inductive Rule Learning (FS-Miner)	119
4.3.2	Inductive Rule Learning (FS-FOIL)	124
4.3.3	Inductive Learning of Fuzzy Decision Trees (FS-ID3)	127
4.3.4	Inductive Learning of Fuzzy Linear Regression Trees (FS-LiRT)	132
4.4	Reasoning	139
4.4.1	Classification	139
4.4.2	Regression	140
4.4.3	Linguistic Descriptions	142

5	Examples	149
5.1	Classification	149
5.1.1	Wine Data Set	149
5.1.2	Comparison of Results	154
5.2	Regression	159
5.2.1	Two-Dimensional Example A	159
5.2.2	Two-Dimensional Example B	166
5.2.3	High-dimensional Data Set	169
5.2.4	Comparison of Results	175
5.3	Linguistic Descriptions	177
5.3.1	Image Segmentation	177
5.4	On the Optimal Choice of Predicates	181
5.4.1	Number of Fuzzy Sets	181
5.4.2	Ordering-based Predicates	184
5.4.3	Shape of Fuzzy Sets	185
6	Conclusion	187
6.1	Achievements	187
6.2	Future Work	188
A	machine learning framework for <i>Mathematica</i>	191