

Contents

1	Introduction	1
1.1	Problem Definition	2
1.2	General Assumptions	3
1.3	Outline of This Book	4
1.4	General Notation	7
2	Survey and Overview	9
2.1	Active Learning	9
2.2	Imbalanced Classification	11
2.3	Anomaly Detection (Outlier Detection)	12
2.4	Clustering	13
2.5	Co-clustering	13
2.6	Unsupervised Feature Selection	15
2.7	Rare Category Detection	15
3	Rare Category Detection	17
3.1	Rare Category Detection with Priors for Data with Features .	17
3.1.1	Rare Category Detection for the Binary Case	18
3.1.2	Rare Category Detection for Multiple Classes	23
3.1.3	Experimental Results	29
3.2	Prior-free Rare Category Detection for Data with Features .	38
3.2.1	Semiparametric Density Estimation for Rare Category Detection	39
3.2.2	Algorithm	47
3.2.3	Experimental Results	49
3.3	Rare Category Detection for Graph Data	54
3.3.1	GRADE Algorithm	55
3.3.2	GRADE-LI Algorithm	61
3.3.3	Experimental Results	61

3.3.4	Discussion	67
3.4	Summary of Rare Category Detection	72
3.4.1	Limitations	73
4	Rare Category Characterization	75
4.1	Optimization Framework	76
4.1.1	Pre-processing: Filtering	77
4.1.2	Problem Formulations	77
4.2	Optimization Algorithm: <i>RACH</i>	80
4.2.1	Initialization Step	81
4.2.2	Projected Subgradient Method for Problem 4.3 . . .	82
4.2.3	<i>RACH</i> for Problem 4.1	84
4.3	Kernelized <i>RACH</i> Algorithm	87
4.4	Experimental Results	87
4.5	Summary of Rare Category Characterization	96
4.5.1	Limitations	97
5	Unsupervised Rare Category Analysis	99
5.1	Optimization Framework	100
5.1.1	Additional Notation	100
5.1.2	Objective Function	101
5.1.3	Justification	102
5.2	Partial Augmented Lagrangian Method	104
5.3	Experimental Results	107
5.3.1	Synthetic Data Sets	107
5.3.2	Real Data Sets	112
5.4	Summary of Unsupervised Rare Category Analysis	121
5.4.1	Limitations	123
6	Conclusion and Future Directions	125
	Bibliography	129