
Contents

Preface	xi
Part 1. Computational Data Analysis and Methods	1
Chapter 1. Semi-supervised Learning Based on Distributionally Robust Optimization	3
Jose BLANCHET and Yang KANG	
1.1. Introduction	3
1.2. Alternative semi-supervised learning procedures	7
1.3. Semi-supervised learning based on DRO	9
1.3.1. Defining the optimal transport discrepancy	9
1.3.2. Solving the SSL-DRO formulation	10
1.4. Error improvement of our SSL-DRO formulation	13
1.5. Numerical experiments	16
1.6. Discussion on the size of the uncertainty set	17
1.7. Conclusion	20
1.8. Appendix: supplementary material: technical details for theorem 1.1	21
1.8.1. Assumptions of theorem 1.1	21
1.8.2. Revisit theorem 1.1	22
1.8.3. Proof of theorem 1.1	23
1.9. References	31

Chapter 2. Updating of PageRank in Evolving Treegraphs	35
Benard ABOLA, Pitos Seleka BIGANDA, Christopher ENGSTRÖM, John Magero MANGO, Godwin KAKUBA and Sergei SILVESTROV	
2.1. Introduction	35
2.2. Abbreviations and definitions	38
2.3. Finding components	39
2.3.1. Isolation of vertices in the graph	39
2.3.2. Keeping track of every vertex in the components	40
2.4. Maintaining the level of cycles	40
2.5. Calculating PageRank	41
2.6. PageRank of a tree with at least a cycle after addition of an edge	43
2.7. Updating PageRank of evolving treegraph with cyclic components	45
2.8. Aggregation/disaggregation methods of stochastic matrices	46
2.9. Numerical experiment	47
2.10. Procedure to compute PageRank	49
2.11. Conclusion	50
2.12. Acknowledgements	50
2.13. References	50
Chapter 3. Exploring The Relationship Between Ordinary PageRank, Lazy PageRank and Random Walk with Backstep PageRank for Different Graph Structures	53
Pitos Seleka BIGANDA, Benard ABOLA, Christopher ENGSTRÖM, John Magero MANGO, Godwin KAKUBA and Sergei SILVESTROV	
3.1. Introduction	53
3.2. Notations and basic concepts	56
3.3. Mathematical relationships between variants of PageRank	57
3.3.1. Ordinary PageRank $\vec{\pi}^{(t)}$	57
3.3.2. Generalized lazy PageRank $\vec{\pi}^{(g)}$	58
3.3.3. Random walk with backstep PageRank $\vec{\pi}^{(b)}$	60
3.4. Convergence rates of the variants of PageRank	63

3.5. Comparison of ranking behaviors for the variants of PageRank	66
3.5.1. Comparing PageRank of simple networks	66
3.5.2. Numerical experiments for large network	69
3.6. Conclusion	71
3.7. Acknowledgements	71
3.8. References	72
Chapter 4. On the Behavior of Alternative Splitting Criteria for CUB Model-based Trees	75
Carmela CAPPELLI, Rosaria SIMONE and Francesca DI IORIO	
4.1. Introduction	75
4.2. Cubremot	77
4.3. Application and comparison	80
4.4. Further developments	85
4.5. References	88
Chapter 5. Investigation on Life Satisfaction Through (Stratified) Chain Regression Graph Models	89
Federica NICOLUSSI and Manuela CAZZARO	
5.1. Introduction	89
5.2. Methodology	90
5.3. Application	93
5.3.1. Survey on multiple aims analysis	94
5.4. Conclusion	99
5.5. References	99
Part 2. Classification Data Analysis and Methods	101
Chapter 6. Selection of Proximity Measures for a Topological Correspondence Analysis	103
Rafik ABDESSELAM	
6.1. Introduction	103
6.2. Topological correspondence	109
6.2.1. Comparison and selection of proximity measures	110
6.2.2. Statistical comparisons between two proximity measures	112

6.3. Application to real data and empirical results	114
6.4. Conclusion and perspectives	118
6.5. Appendix	119
6.6. References	120

Chapter 7. Support Vector Machines: A Review and Applications in Statistical Process Monitoring 123

Anastasios APSEMIDIS and Stelios PSARAKIS

7.1. Introduction	123
7.2. Review of the literature	126
7.3. Application	134
7.4. Conclusion	138
7.5. Acknowledgement	138
7.6. References	138

Chapter 8. Binary Classification Techniques: An Application on Simulated and Real Bio-medical Data 145

Fragkiskos G. BERSIMIS, Iraklis VARLAMIS, Malvina VAMVAKARI and Demosthenes B. PANAGIOTAKOS

8.1. Introduction	145
8.2. Related work	148
8.3. Materials and methods	150
8.3.1. Data-driven health index construction	150
8.3.2. Classification methods for discrete data	151
8.4. Experimental evaluation	155
8.4.1. Synthetic data generation	155
8.4.2. ATTICA study: dietary data collection	156
8.4.3. Evaluation of classification performance	157
8.5. Results	159
8.5.1. Results on synthetic data	159
8.5.2. Results on real data	165
8.6. Discussion	167
8.7. Conclusion	169
8.8. Acknowledgements	170
8.9. References	170

Chapter 9. Some Properties of the Multivariate Generalized Hyperbolic Models	177
Stergios B. FOTOPOULOS, Venkata K. JANDHYALA and Alex PAPANAS	
9.1. Introduction	177
9.2. The MGH family of distributions and their limiting forms	179
9.3. The conditional MGH distribution and its limits	187
9.4. References	192
Chapter 10. On Determining the Value of Online Customer Satisfaction Ratings – A Case-based Appraisal	195
Jim FREEMAN	
10.1. Introduction	195
10.2. Incomplete, inconsistent and contradictory results	197
10.3. Sample size volatility	201
10.4. Technical inadequacies of the customer score criterion	202
10.5. Non-standard weighting of survey responses	206
10.6. Survey bias	208
10.6.1. Population	209
10.6.2. Web population	209
10.6.3. Web survey panel	209
10.6.4. Web survey sample	210
10.6.5. Web survey non-response	210
10.7. Conclusion	210
10.8. References	211
Chapter 11. Projection Clustering Unfolding: A New Algorithm for Clustering Individuals or Items in a Preference Matrix	215
Mariangela SCIANDRA, Antonio D'AMBROSIO and Antonella PLAIA	
11.1. Introduction	215
11.2. Preference data	216
11.3. Projection pursuit	217
11.3.1. Projection indices	219
11.4. Projection pursuit clustering	220
11.5. Clustering preference data	221

11.5.1. The projection clustering unfolding (PCU)	222
11.5.2. The projection clustering unfolding: a real example	223
11.6. Conclusion	228
11.7. References	228
List of Authors	231
Index	235