
Contents

1	Some Insights into the Geometry and Training of Neural Networks	1
1.1	Introduction	1
1.2	Formation of Decision Regions	4
1.3	Region Properties and Approximation	15
1.4	Gradients	21
1.5	Optimization	28
1.6	Conclusions	40
2	Variable Selection in Gaussian Markov Random Fields	45
2.1	Introduction	46
2.2	Background	47
2.3	Preliminaries	49
2.4	Block Coordinate Descent Method	51
2.5	Experimental Results	54
2.6	Conclusions and Future Work	61
2.7	Appendix: Technical Lemma	63
2.8	Appendix: Additional Experimental Results	64
3	Log-Nonlinear Formulations for Robust High-dimensional Modeling	71
3.1	Introduction	72
3.2	Model Setup	74
3.3	A Regularized Integrated Squared Error Estimator	75
3.4	Simulation Study	82
3.5	Applications	85
3.6	Concluding Remarks	90
3.7	Appendix: Proof of Theorem 1	91
4	Semistochastic Quadratic Bound Methods	95
4.1	Introduction	96
4.2	Quadratic Bound Methods	98

4.3	Stochastic and Semistochastic Extensions	99
4.4	Implementation Details	104
4.5	Experiments	107
4.6	Concluding Remarks	107
5	Use of Deep Learning Features in Log-Linear Models	113
5.1	Introduction	114
5.2	A Framework of Using Deep Learning Features for Classification	117
5.3	DSN and Kernel-DSN Features for Log-Linear Classifiers . . .	121
5.4	DNN Features for Conditional-Random-Field Sequence Classifiers	130
5.5	Log-Linear Stacking to Combine Multiple Deep Learning Systems	132
5.6	Discussion and Conclusions	134
6	Informative Nonstationarity in Paleoclimatological Log-Linear Models	145
6.1	Introduction	146
6.2	Principles of Flicker-Noise Spectroscopy	147
6.3	FNS Smoothing Procedure	148
6.4	FNS Nonstationarity Factor	150
6.5	Experimental Data	151
6.6	Procedure for Analysis	151
6.7	Results and Discussion	152
6.8	Concluding Remarks	158
7	Log-Linear Quandloids and Information Fusion	161
7.1	Introduction	161
7.2	An Illustrative Example	162
7.3	Algebra of Information Fusion	165
7.4	What Is an Information Fusion Tangle?	172
7.5	Information Fusion Tangle Equivalence	173
7.6	A More Sophisticated Example	175
8	Optimization Techniques for Large Speech Tasks	179
8.1	Introduction	180
8.2	Related Work	182
8.3	Parallel Stochastic Gradient Descent	184
8.4	Hessian-Free Optimization	188
8.5	Low-Rank Matrices	191
8.6	Analysis of Optimization Techniques	191
8.7	Performance on Larger Tasks	198
8.8	Conclusions and Future Work	201