
Contents

An Introduction to Feature Extraction

<i>Isabelle Guyon, André Elisseeff</i>	1
1 Feature Extraction Basics	1
2 What is New in Feature Extraction?	7
3 Getting Started	9
4 Advanced Topics and Open Problems	16
5 Conclusion	22
References	23
A Forward Selection with Gram-Schmidt Orthogonalization	24
B Justification of the Computational Complexity Estimates	25

Part I Feature Extraction Fundamentals

1 Learning Machines

<i>Norbert Jankowski, Krzysztof Grabczewski</i>	29
1.1 Introduction	29
1.2 The Learning Problem	29
1.3 Learning Algorithms	35
1.4 Some Remarks on Learning Algorithms	57
References	58

2 Assessment Methods

<i>Gérard Dreyfus, Isabelle Guyon</i>	65
2.1 Introduction	65
2.2 A Statistical View of Feature Selection: Hypothesis Tests and Random Probes	66
2.3 A Machine Learning View of Feature Selection	78
2.4 Conclusion	86
References	86

3 Filter Methods

<i>Włodzisław Duch</i>	89
3.1 Introduction to Filter Methods for Feature Selection	89
3.2 General Issues Related to Filters	91
3.3 Correlation-Based Filters	96
3.4 Relevance Indices Based on Distances Between Distributions	99
3.5 Relevance Measures Based on Information Theory	101
3.6 Decision Trees for Filtering	104
3.7 Reliability and Bias of Relevance Indices	106
3.8 Filters for Feature Selection	108
3.9 Summary and Comparison	110
3.10 Discussion and Conclusions	113
References	114

4 Search Strategies

<i>Juha Reunanen</i>	119
4.1 Introduction	119
4.2 Optimal Results	119
4.3 Sequential Selection	121
4.4 Extensions to Sequential Selection	123
4.5 Stochastic Search	129
4.6 On the Different Levels of Testing	133
4.7 The Best Strategy?	134
References	135

5 Embedded Methods

<i>Thomas Navin Lal, Olivier Chapelle, Jason Weston, André Elisseeff</i>	137
5.1 Introduction	137
5.2 Forward-Backward Methods	139
5.3 Optimization of Scaling Factors	150
5.4 Sparsity Term	156
5.5 Discussions and Conclusions	161
References	162

6 Information-Theoretic Methods

<i>Kari Torkkola</i>	167
6.1 Introduction	167
6.2 What is Relevance?	167
6.3 Information Theory	169
6.4 Information-Theoretic Criteria for Variable Selection	172
6.5 MI for Feature Construction	178
6.6 Information Theory in Learning Distance Metrics	179
6.7 Information Bottleneck and variants	180
6.8 Discussion	181
References	182

7 Ensemble Learning

<i>Eugene Tu</i>	187
7.1 Introduction.....	187
7.2 Overview of Ensemble Methods	188
7.3 Variable Selection and Ranking with Tree Ensembles	191
7.4 Bayesian Voting	200
7.5 Discussions.....	201
References	203

8 Fuzzy Neural Networks

<i>Madan M. Gupta, Noriyasu Homma, Zeng-Guang Hou</i>	205
8.1 Introduction.....	205
8.2 Fuzzy Sets and Systems: An Overview	207
8.3 Building Fuzzy Neurons Using Fuzzy Arithmetic and Fuzzy Logic Operations	215
8.4 Hybrid Fuzzy Neural Networks (HFNNs).....	222
8.5 Concluding Remarks	230
References	231

Part II Feature Selection Challenge

9 Design and Analysis of the NIPS2003 Challenge

<i>Isabelle Guyon, Steve Gunn, Asa Ben Hur, Gideon Dror</i>	237
9.1 Introduction.....	237
9.2 Benchmark Design	239
9.3 Challenge Results	245
9.4 Post-Challenge Verifications	253
9.5 Conclusions and Future Work	259
References	260
A Details About the Fifty Feature Subset Study	261

10 High Dimensional Classification with Bayesian Neural Networks and Dirichlet Diffusion Trees

<i>Radford M. Neal, Jianguo Zhang</i>	265
10.1 Bayesian Models vs. Learning Machines	266
10.2 Selecting Features with Univariate Tests	269
10.3 Reducing Dimensionality Using PCA	271
10.4 Bayesian Logistic Regression	271
10.5 Bayesian Neural Network Models	274
10.6 Dirichlet Diffusion Tree Models	276
10.7 Methods and Results for the Challenge Data Sets	280
10.8 Conclusions	294
References	295

11 Ensembles of Regularized Least Squares Classifiers for High-Dimensional Problems

<i>Kari Torkkola and Eugene Tu</i>	297
11.1 Introduction	297
11.2 Regularized Least-Squares Classification (RLSC)	298
11.3 Model Averaging and Regularization	300
11.4 Variable Filtering with Tree-Based Ensembles	301
11.5 Experiments with Challenge Data Sets	302
11.6 Future Directions	310
11.7 Conclusion	312
References	313

12 Combining SVMs with Various Feature Selection Strategies

<i>Yi-Wei Chen, Chih-Jen Lin</i>	315
12.1 Introduction	315
12.2 Support Vector Classification	316
12.3 Feature Selection Strategies	317
12.4 Experimental Results	320
12.5 Competition Results	321
12.6 Discussion and Conclusions	322
References	323

13 Feature Selection with Transductive Support Vector Machines

<i>Zhili Wu, Chunhung Li</i>	325
13.1 Introduction	325
13.2 SVMs and Transductive SVMs	326
13.3 Feature Selection Methods Related with SVMs	329
13.4 Transductive SVM-Related Feature Selection	331
13.5 Experimentation	332
13.6 Conclusion and Discussion	339
13.7 Acknowledgement	340
References	340

14 Variable Selection using Correlation and Single Variable Classifier Methods: Applications

<i>Amir Reza Saffari Azar Alamdari</i>	343
14.1 Introduction	343
14.2 Introduction to Correlation and Single Variable Classifier Methods	344
14.3 Ensemble Averaging	347
14.4 Applications to NIPS 2003 Feature Selection Challenge	350
14.5 Conclusion	355
References	357

15 Tree-Based Ensembles with Dynamic Soft Feature Selection

<i>Alexander Borisov, Victor Erubimov, Eugene Tuv</i>	359
15.1 Background	359
15.2 Dynamic Feature Selection	361
15.3 Experimental results	367
15.4 Summary	373
References	374

16 Sparse, Flexible and Efficient Modeling using L_1 Regularization

<i>Saharon Rosset, Ji Zhu</i>	375
16.1 Introduction	375
16.2 The L_1 -Norm Penalty	380
16.3 Piecewise Linear Solution Paths	384
16.4 A Robust, Efficient and Adaptable Method for Classification	388
16.5 Results on the NIPS-03 Challenge Datasets	390
16.6 Conclusion	392
References	393

17 Margin Based Feature Selection and Infogain with Standard Classifiers

<i>Ran Gilad-Bachrach, Amir Navot</i>	395
17.1 Methods	395
17.2 Results	398
17.3 Discussion	399
References	400

18 Bayesian Support Vector Machines for Feature Ranking and Selection

<i>Wei Chu, S. Sathiya Keerthi, Chong Jin Ong, Zoubin Ghahramani</i>	403
18.1 Introduction	403
18.2 Bayesian Framework	404
18.3 Post-processing for Feature Selection	410
18.4 Numerical Experiments	414
18.5 Conclusion	416
References	416

19 Nonlinear Feature Selection with the Potential Support Vector Machine

<i>Sepp Hochreiter, Klaus Obermayer</i>	419
19.1 Introduction	419
19.2 The Potential Support Vector Machine	420
19.3 P-SVM Discussion and Redundancy Control	424
19.4 Nonlinear P-SVM Feature Selection	427
19.5 Experiments	429

19.6 Conclusion	436
References	436
20 Combining a Filter Method with SVMs	
<i>Thomas Navin Lal, Olivier Chapelle, Bernhard Schölkopf</i>	439
20.1 The Parameters σ and C of the SVM	439
20.2 Feature Ranking	440
20.3 Number of Features	441
20.4 Summary	443
References	445
21 Feature Selection via Sensitivity Analysis with Direct Kernel PLS	
<i>Mark J. Embrechts, Robert A. Bress, Robert H. Kewley</i>	447
21.1 Introduction	447
21.2 Partial Least Squares Regression (PLS)	448
21.3 Regression Models Based on Direct Kernels	450
21.4 Dealing with the Bias: Centering the Kernel	452
21.5 Metrics for Assessing the Model Quality	453
21.6 Data Conditioning and Preprocessing	454
21.7 Sensitivity Analysis	455
21.8 Heuristic Feature Selection Policies for the NIPS Feature Selection Challenge	456
21.9 Benchmarks	459
21.10 Conclusions	460
References	461
22 Information Gain, Correlation and Support Vector Machines	
<i>Danny Roobaert, Grigoris Karakoulas, Nitesh V. Chawla</i>	463
22.1 Introduction	463
22.2 Description of Approach	464
22.3 Final Results	467
22.4 Alternative Approaches Pursued	468
22.5 Discussion and Conclusion	469
References	470
23 Mining for Complex Models Comprising Feature Selection and Classification	
<i>Krzysztof Grąbczewski, Norbert Jankowski</i>	471
23.1 Introduction	471
23.2 Fundamental Algorithms	472
23.3 Fully Operational Complex Models	481
23.4 Challenge Data Exploration	483
23.5 Conclusions	486
References	487

24 Combining Information-Based Supervised and Unsupervised Feature Selection

<i>Sang-Kyun Lee, Seung-Joon Yi, Byoung-Tak Zhang</i>	489
24.1 Introduction	489
24.2 Methods	490
24.3 Experiments	494
24.4 Conclusions	496
References	498

25 An Enhanced Selective Naïve Bayes Method with Optimal Discretization

<i>Marc Boullé</i>	499
25.1 Introduction	499
25.2 The Enhanced Selective Naïve Bayes Method	500
25.3 The MODL Discretization Method	502
25.4 Results on the NIPS Challenge	504
25.5 Conclusion	506
References	506

26 An Input Variable Importance Definition based on Empirical Data Probability Distribution

<i>V. Lemaire, F. Clérot</i>	509
26.1 Introduction	509
26.2 Analysis of an Input Variable Influence	510
26.3 Application to Feature Subset Selection	512
26.4 Results on the NIPS Feature Selection Challenge	513
26.5 Conclusions	516
References	516

Part III New Perspectives in Feature Extraction

27 Spectral Dimensionality Reduction

<i>Yoshua Bengio, Olivier Delalleau, Nicolas Le Roux, Jean-François Païement, Pascal Vincent, Marie Ouimet</i>	519
27.1 Introduction	519
27.2 Data-Dependent Kernels for Spectral Embedding Algorithms	524
27.3 Kernel Eigenfunctions for Induction	532
27.4 Learning Criterion for the Leading Eigenfunctions	539
27.5 Experiments	541
27.6 Conclusion	544
References	547

28 Constructing Orthogonal Latent Features for Arbitrary Loss

<i>Michinari Momma, Kristin P. Bennett</i>	551
28.1 Introduction	551
28.2 General Framework in BLF	554
28.3 BLF with Linear Functions	557
28.4 Convergence Properties of BLF	561
28.5 PLS and BLF	563
28.6 BLF for Arbitrary Loss	564
28.7 Kernel BLF	571
28.8 Computational Results	572
28.9 Conclusion	581
References	582

29 Large Margin Principles for Feature Selection

<i>Ran Gilad-Bachrach, Amir Navot, Naftali Tishby</i>	585
29.1 Introduction	585
29.2 Margins	586
29.3 Algorithms	589
29.4 Theoretical Analysis	592
29.5 Empirical Assessment	593
29.6 Discussion and Further Research Directions	602
References	604
A Complementary Proofs	604

30 Feature Extraction for Classification of Proteomic Mass Spectra: A Comparative Study

<i>Ilya Levner, Vadim Bulitko, Guohui Lin</i>	607
30.1 Introduction	607
30.2 Existing Feature Extraction and Classification Methods	611
30.3 Experimental Results	615
30.4 Conclusion	622
30.5 Acknowledgements	623
References	623

31 Sequence Motifs: Highly Predictive Features of Protein Function

<i>Asa Ben-Hur, Douglas Brutlag</i>	625
31.1 Introduction	625
31.2 Enzyme Classification	627
31.3 Methods	628
31.4 Results	636
31.5 Discussion	642
31.6 Conclusion	643
References	643

Appendix A Elementary Statistics

Elementary Statistics

<i>Gérard Dreyfus</i>	649
1 Basic Principles	649
2 Estimating and Learning	651
3 Some Additional Useful Probability Distributions	654
4 Confidence Intervals	655
5 Hypothesis Testing	657
6 Probably Approximately Correct (PAC) Learning and Guaranteed Estimators	660
References	662

Appendix B Feature Selection Challenge Datasets

Experimental Design

<i>Isabelle Guyon</i>	665
ARCENE	669
GISETTE	677
DEXTER	683
DOROTHEA	687
MADELON	691
MATLAB Code of the Lambda Method	697
MATLAB Code Used to Generate Madelon	699

Appendix C Feature Selection Challenge Fact Sheets

10 High Dimensional Classification with Bayesian Neural Networks and Dirichlet Diffusion Trees

<i>Radford M. Neal, Jianguo Zhang</i>	707
---	-----

11 Ensembles of Regularized Least Squares Classifiers for High-Dimensional Problems

<i>Kari Torkkola and Eugene Tuv</i>	709
---	-----

12 Combining SVMs with Various Feature Selection Strategies

Yi-Wei Chen, Chih-Jen Lin 711

13 Feature Selection with Transductive Support Vector Machines

Zhili Wu, Chunhung Li 713

14 Variable Selection using Correlation and SVC Methods: Applications

Amir Reza Saffari Azar Alamdari 715

15 Tree-Based Ensembles with Dynamic Soft Feature Selection

Alexander Borisov, Victor Eruhimov, Eugene Tuv 717

16 Sparse, Flexible and Efficient Modeling using L_1 Regularization

Saharon Rosset, Ji Zhu 719

17 Margin Based Feature Selection and Infogain with Standard Classifiers

Ran Gilad-Bachrach, Amir Navot 721

18 Bayesian Support Vector Machines for Feature Ranking and Selection

Wei Chu, S. Sathiya Keerthi, Chong Jin Ong, Zoubin Ghahramani 723

19 Nonlinear Feature Selection with the Potential Support Vector Machine

Sepp Hochreiter, Klaus Obermayer 725

20 Combining a Filter Method with SVMs

Thomas Navin Lal, Olivier Chapelle, Bernhard Schölkopf 729

21 Feature Selection via Sensitivity Analysis with Direct Kernel PLS

Mark J. Embrechts, Robert A. Bress, Robert H. Kewley 731

22 Information Gain, Correlation and Support Vector Machines

Danny Roobaert, Grigoris Karakoulas, Nitesh V. Chawla 733

23 Mining for Complex Models Comprising Feature Selection and Classification

Krzysztof Grabczewski, Norbert Jankowski 735

24 Combining Information-Based Supervised and Unsupervised Feature Selection

Sang-Kyun Lee, Seung-Joon Yi, Byoung-Tak Zhang 737

25 An Enhanced Selective Naïve Bayes Method with Optimal Discretization

Marc Boullé 741

26 An Input Variable Importance Definition based on Empirical Data Probability Distribution

V. Lemaire, F. Clérot 743

Appendix D Feature Selection Challenge Results Tables

Result Tables of the NIPS2003 Challenge

Isabelle Guyon, Steve Gunn 747

ARCENE 749

DEXTER 753

DOROTHEA 757

GISETTE 761

DOROTHEA 765

Overall Results 769

Index 773