

Contents

Contents	5
Preface	11
1. Assessment of Numeric Predictions	15
Notation	16
Overview of Performance Measures	17
Consistency and Evolutionary Stability	20
Selection Bias and the Need for Three Datasets	24
Cross Validation and Walk Forward Testing	29
Bias in Cross Validation	30
Overlap Considerations	31
Assessing Nonstationarity Using Walk Forward Testing	33
Nested Cross Validation Revisited	34
Common Performance Measures	36
Mean Squared Error	36
Mean Absolute Error	39
R-Squared	39
RMS Error	40
Nonparametric Correlation	40
Success Ratios	42
Alternatives to Common Performance Measures	44
Stratification For Consistency	45
Confidence Intervals	47
The Confidence Set	47
Serial Correlation	49
Multiplicative Data	50
Normally Distributed Errors	51
Empirical Quantiles as Confidence Intervals	53
Confidence Bounds for Quantiles	55
Tolerance Intervals	59
2. Assessment of Class Predictions	63
The Confusion Matrix	64
Expected Gain/Loss	65
ROC (Receiver Operating Characteristic) Curves	67
Hits, False Alarms, and Related Measures	67

Computing the ROC Curve	69
Area Under the ROC Curve	75
Cost and the ROC Curve	77
Optimizing ROC-Based statistics	79
Optimizing the Threshold: Now or Later?	80
Maximizing Precision	84
Generalized Targets	84
Maximizing Total Gain	85
Maximizing Mean Gain	86
Maximizing the Standardized Mean Gain	87
Confidence in Classification Decisions	89
Hypothesis Testing	89
Confidence in the Confidence	96
Bayesian Methods	102
Multiple Classes	107
Hypothesis Testing Versus Bayes' Method	108
Final Thoughts on Hypothesis Testing	114
Confidence Intervals for Future Performance	121
3. Resampling for Assessing Parameter Estimates	123
Bias and Variance of Statistical Estimators	124
Plug-In Estimators and Empirical Distributions	126
Bias of an Estimator	127
Variance of an Estimator	128
Bootstrap Estimation of Bias and Variance	130
Code For Bias and Variance Estimation	133
Plug-In Estimators Can Provide Better Bootstraps	137
A Model Parameter Example	141
Confidence Intervals	145
Is the Interval Backwards?	150
Improving the Percentile Method	154
Hypothesis Tests For Parameter Values	161
Bootstrapping Ratio Statistics	163
Jackknife Estimates of Bias and Variance	175
Bootstrapping Dependent Data	178
Estimating the Extent of Autocorrelation	179
The Stationary Bootstrap	182
Choosing a Block Size for the Stationary Bootstrap	185

The Tapered Block Bootstrap	190
Choosing a Block Size for the Tapered Block Bootstrap	198
What If the Block Size Is Wrong?	200
4. Resampling for Assessing Prediction and Classification	213
Partitioning the Error	214
Cross Validation	217
Bootstrap Estimation of Population Error	220
Efron's E0 Estimate of Population Error	223
Efron's E632 Estimate of Population Error	227
Comparing the Error Estimators For Prediction	229
Comparing the Error Estimators For Classification	231
Summary	233
5. Miscellaneous Resampling Techniques	235
Bagging	236
A Quasi-Theoretical Justification	237
The Component Models	240
Code for Bagging	241
AdaBoost	246
Binary AdaBoost for Pure Classification Models	247
Probabilistic Sampling for Inflexible Models	255
Binary AdaBoost When the Model Provides Confidence	258
AdaBoost.MH for More Than Two Classes	266
AdaBoost.OC for More Than Two Classes	274
Comparing the Boosting Algorithms	292
A Binary Classification Problem	292
A Multiple-Class Problem	295
Final Thoughts on Boosting	296
Permutation Training and Testing	298
The Permutation Training Algorithm	300
Partitioning the Training Performance	301
A Demonstration of Permutation Training	304
6. Combining Numeric Predictions	313
Simple Average	314
Code for Averaging Predictions	316
Unconstrained Linear Combinations	318

Constrained Linear Combinations	321
Constrained Combination of Unbiased Models	326
Variance-Weighted Interpolation	328
Combination by Kernel Regression Smoothing	331
Code for the GRNN	336
Comparing the Combination Methods	341
7. Combining Classification Models	345
Introduction and Notation	346
Reduction Versus Ordering	348
The Majority Rule	349
Code for the Majority Rule	350
The Borda Count	353
The Average Rule	355
Code for the Average Rule	355
The Median Alternative	357
The Product Rule	357
The MaxMax and MaxMin Rules	358
The Intersection Method	359
The Union Rule	365
Logistic Regression	368
Code For the Combined Weight Method	372
The Logit Transform and Maximum Likelihood Estimation	376
Code for Logistic Regression	381
Separate Weight Sets	385
Model Selection By Local Accuracy	388
Code For Local Accuracy Selection	390
Maximizing the Fuzzy Integral	399
What Does This Have To Do With Classifier Combination?	402
Code for the Fuzzy Integral	404
Pairwise Coupling	412
Pairwise Threshold Optimization	421
A Cautionary Note	422
Comparing the Combination Methods	423
Small Training Set, Three Models	424
Large Training Set, Three Models	425
Small Training Set, Three Good Models, One Worthless	426
Large Training Set, Three Good Models, One Worthless	427

Small Training Set, Worthless and Noisy Models Included	428
Large Training Set, Worthless and Noisy Models Included	429
Five Classes	430
8. Gating Methods	431
Preordained Specialization	432
Learned Specialization	433
After-the-Fact Specialization	434
Code for After-the-Fact Specialization	434
Some Experimental Results	442
General Regression Gating	444
Code for GRNN Gating	447
Experiments with GRNN Gating	454
9. Information and Entropy	457
Entropy	458
Entropy of a Continuous Random Variable	461
Partitioning a Continuous Variable for Entropy	462
An Example of Improving Entropy	467
Joint and Conditional Entropy	470
Code for Conditional Entropy	474
Mutual Information	475
Fano's Bound and Selection of Predictor Variables	477
Confusion Matrices and Mutual Information	480
Extending Fano's Bound for Upper Limits	482
Simple Algorithms for Mutual Information	486
The TEST_DIS Program	491
Continuous Mutual Information	494
The Parzen Window Method	495
Adaptive Partitioning	504
The TEST_CON Program	518
Predictor Selection Using Mutual Information	519
Maximizing Relevance While Minimizing Redundancy	519
The MI_DISC and MI_CONT Programs	522
A Contrived Example of Information Minus Redundancy	523
A Superior Selection Algorithm for Binary Variables	525
Screening Without Redundancy	530
Asymmetric Information Measures	540

Uncertainty Reduction	540
Transfer Entropy: Schreiber's Information Transfer	544
References	553
Index	557