
Index

- ϵ 0 Bootstrap, 179–182
- .632 Bootstrap, 179–183, 186, 198–199
- 2-fold matched sample t-test, 217–226, 264–265
- 2-fold matched pair t-test, 225
- 5×2 CV F-Test, 261–263
- 5×2 CV t-test, 261
- 10×10 CV, 284–289

- Accuracy, 86–88, 315
- Active learning, 25, 311
- Agreement statistic, 89–94, 110
- Aggregation, 10, 12, 14, 322, 36, 380
- ANOVA (Analysis of Variance), 214–216, 239–240, 242, 245–248, 250–255, 257–258, 268–269, 272, 275–276, 289–290, 347, 381–383, 385
- Artificial data approach, 294, 301–304
- Asymmetric misclassification cost, 12, 105
- Area Under the Curve (AUC), 128–131, 136, 138, 153, 156, 196, 309–310, 314, 370, 376, 382, 385–386, 388, 391
- AUC of the multi-class ROC
- Artificial data, 18, 294, 301–304

- Balanced bootstrap sampling, 182
- Balanced F-measure, 70, 103–104
- Bayesian Information Reward (BIR), 81, 143
- Bernoulli trial, 55–56
- Between-group variation, 240
- Bias, 39, 167–168, 170, 178, 181, 225
- Bias-variance analysis, 34, 40, 162, 167–168, 329
- Bias-variance decomposition, 35–39, 167–168, 170, 395
- Bias-variance tradeoff, 39–40, 169–171
- Binary classification, 2, 25, 31–32, 35, 37, 75, 77, 79–81, 85–86, 90, 94–96, 99, 101, 104, 113, 137, 174, 186, 230, 314, 317, 330, 349
- Binomial distribution, 53–57, 59, 165, 228–229, 232, 325–326, 333

- Bonferroni adjustment, 11
- Bonferroni Test, 254–255
- Bonferroni-Dunn Test, 214, 255
- Bootstrap, 15, 168, 178–183, 196–198, 304, 373
- Bootstrapping, 15, 162, 179–183, 185–187, 196–199, 203, 346–348, 373

- Calibration, 130–131, 140, 142–143, 146, 160, 315
- Candidate evaluation function, 318, 321
- Capacity, 30
- Central limit theorem, 42, 53, 56–59, 325
- Chance agreement, 89–93
- Chi-squared table, 265–266
- Classification, 2, 20–21, 25–27, 31–32, 79–81
- Classifier space, 24, 28–32, 34, 41, 323–324, 327–329
- Class imbalance problem, 102, 104, 317
- Class ratio, 105, 118, 120
- Class skew, 118
- Cohen's d , 72, 221–222, 265, 373
- Cohen's κ (kappa), 109, 304
- Community experiments effect, 15, 21
- Confidence coefficient, 59
- Confidence interval, 42, 47, 59–65, 68, 159, 165–166, 176, 211–212, 265, 325–326, 329, 332–334, 351
- Confidence level, 59–61, 63, 165, 170, 191, 215, 264, 285–287, 388
- Confidence parameter, 59, 61, 70, 166, 324
- Confirmatory data analysis, 65
- Confusion matrix, 77–79
- Corrected random subsampling, 196, 259–260, 278–279
- Cost, 104–106
- Cost curves, 82, 112, 132–135, 155–159
- Cost curve space, 133–134
- Cost ratio, 106, 110, 117, 119, 124, 128, 136, 309
- Credible interval, 60

- Critical value, 67, 232, 234–235, 237–238, 242, 249, 252–256, 351, 355, 362–363, 390
- Cross-entropy, 146
- Cross-validated t-test, 188
- Cross-Validation (CV), 172
- 10-fold Cross Validation, 10, 43, 48, 64, 84, 128, 147–148, 159, 199–200, 203–204, 219, 223, 228, 232, 242, 263, 266, 300, 341, 364
- Cumulative distribution function (CDF), 45, 61
- Data mining, 2, 6, 19–20, 206, 246, 294–297, 307, 350, 370
- Dataset, 15
- Dataset selection, 11, 21, 42, 292, 301, 306, 336–337, 340, 342, 345, 347
- De-Facto approach/culture, 7–8
- DET curve, 136, 156, 160
- Deterministic algorithm, 75
- Domain specific metric, 145
- Dunnnett Test, 253–255, 351, 363, 382, 384
- Effect size, 71–72, 211, 213, 221–222, 258, 264–265, 373
- Efficiency method, 318–320
- Empirical risk, 27–28
- Empirical risk minimization, 29–30, 54, 159
- Error estimation, 35, 74, 84, 161
- Error rate, 86–87, 89, 94, 96, 104, 109, 133–134, 159, 162, 167, 175, 182, 187
- Evaluation framework, 162, 292, 306, 308, 317, 329–330, 335–338, 340–349
- Evaluation framework template, 336–338
- Evaluation measure (performance measure), 75–77
- Evaluation metric (performance metric), 11–12, 42, 82, 84, 145, 160, 207–209, 316, 322–323, 379–381
- Expected cost, 120, 122, 128, 135
- Expected risk, 27–28, 163, 176, 202
- Expected value of a random variable, 45–46
- Exploratory data analysis, 66, 211
- False negative, 69, 79, 89, 91, 95, 116–117, 135
- False negative rate, 95
- False positive, 69
- False positive rate, 84, 94–95, 113–116, 121–128, 135, 149, 151–152, 370–371
- F-measure, 104, 108, 157, 372
- F-Ratio table, 351, 357–360
- Friedman table, 351, 361
- Friedman Test, 11, 239, 248–251, 255, 257–258, 268, 273–275, 290, 385, 388
- Gaussian distribution, 53–54, 60
- Generalization, 16, 32, 34, 39, 41, 69, 87, 90, 93, 110, 131, 144, 164–165, 170, 176, 240, 297, 303, 323, 344
- Generalization error, 16, 28–29, 32, 41, 54, 170, 323–324
- Generic algorithm, 293–294, 301, 336–337, 340
- Geometric distribution, 56
- Geometric mean, 96, 100–101, 103
- Gini coefficient, 122, 139, 159
- Gold standard, 145
- Graphical performance measure, 112
- H Measure, 81, 159, 310
- Holdout method, 70, 161–162, 167, 169, 176, 202, 226
- Holdout risk bound, 325–326, 345
- Honestly significant difference (HSD), 252
- Hypothesis testing, 42, 59–62, 64–68, 162, 184, 188, 211, 216–217, 230, 239, 290, 324
- Indicator function, 32, 85–86, 141, 230, 291, 325
- Inductive inference, 23–24, 48
- Information theoretic measure, 137–138, 143, 159
- Information Reward, 140, 143
- Information Score, 84, 140–142, 157, 317, 376, 384–385, 388, 391
- Iso-curves, 122
- Isometrics, 119–124, 133, 314
- Iso-precision lines, 120
- jackknife, 175, 186, 203
- Kappa statistic, 93
- Kendall coefficient, 144
- k-fold Cross-Validation, 7, 14, 127–128, 162, 171–173, 189–191, 193–194, 202–204, 258, 260, 262, 280, 290, 344
- Kononenko and Bratko's Information Score, 84, 140–142
- Kullback-Leibler divergence, 139
- Learning bias, 31, 33, 177, 323–324, 329, 339
- Leave One Out, 162, 171, 173, 175–176, 194, 202, 327, 341
- Lift chart, 132
- Lift curve, 153–154
- Likelihood ratio, 97–99
- Loss function, 27, 32, 34–38, 40, 48, 52, 75, 106, 137, 144, 168, 204
- Mann-Whitney U test, 129
- McNemar's Test, 68, 217, 226–231, 236, 264–265, 289, 373
- McNemar's contingency matrix, 227–229
- Machine learning, 23–42, 279–280
- Matched samples design, 216
- Matched pairs design, 216, 320–321
- Measure-based method, 318,
- Meta-learning, 294
- Model selection, 11–12, 29, 31–32, 39–40, 177–178
- Monotonic performance measure, 65, 77, 216
- Multiclass, 25, 38, 97, 110, 131, 144, 230, 315
- Multi-class classification, 85, 97, 110, 174–175
- Multiclass focus, 85–86, 101
- Multiclass ROC curve, 131

- Multiple resampling, 161–163, 178–179, 183, 185–186, 188, 202–203
- Multiplicity effect, 15, 299–300, 345
- Negative predictive value (NPV), 99, 371
- Nemenyi Test, 256–257, 275, 382, 388, 390–391
- Nested k-fold cross validation, 178
- “No Free Lunch” theorems, 292
- noise, 36
- Non-parametric test, 110, 131, 144, 164–165, 170, 176, 197, 240, 303, 323, 344
- Normal distribution, 47, 53–57, 60, 62–63, 65, 68, 71, 112, 218, 222, 224, 230, 235, 247, 347–348
- Normalized expected cost (NEC), 135
- Null hypothesis, 65, 216
- Null hypothesis statistical testing (NHST), 207, 289
- Omnibus statistical test, 251
- One-tailed test/One-sided test, 70, 220
- One-Way ANOVA, 214, 245–246
- One-Way repeated measure ANOVA, 240–245
- Online learning, 25, 311
- Ontology of performance measures, 81–82
- Ontology of error-estimation methods, 163
- Operating point, 112, 114, 116, 119, 122–123
- Overview of the statistical tests, 214
- Overfitting, 30, 32, 39, 167, 229, 299, 302, 325
- Panacea approach to evaluation, 4, 7, 349
- Parameter estimation, 34
- Parameter selection, 4, 15, 76, 178
- Parametric test, 15, 69, 215, 217, 226, 230, 239, 247, 263, 289, 347–348
- Parametric hypothesis testing, 68
- Passive learning, 25
- Perfect calibration, 140, 143
- Performance metric, 4, 21, 83, 85–89, 94, 308–309, 311
- Permutation test, 186, 199–203, 329, 373
- Poisson distribution, 55–56
- Positive predictive values (PPV), 13
- Power, 69–70, 145, 203, 226, 230, 256, 260, 275
- Precision, 13, 70–71, 84, 99–100, 102, 104, 109
- Precision-Recall curve (PR Curve), 132–133
- Prior class probability, 13
- Probabilistic algorithm, 75
- Probability density function (PDF), 45, 54
- Probability distribution, 44–46, 55, 58, 67, 139, 331
- Probability space, 44
- Qualitative metric, 311–312, 319
- Quantitative metric, 309–311
- R, 42
- Random variable, 42, 44–48, 52, 54, 58, 60
- Randomization, 15, 162, 183–185, 188, 202, 345, 348
- Random subsampling, 162, 179, 194–196, 258–259, 278–279, 290–291
- Ranking classifier, 75, 118, 129, 248
- Recall, 137–138
- Receiver Operating Characteristic (ROC), 111, 112–113
- regression, 32–33
- Regularization, 29–31, 34
- Relative Superiority Graph, 135–136, 156, 160
- Reliability metric, 76, 111, 137, 145, 157–159, 316–317
- Repeated-measures design, 216
- Replicability, 185, 203, 262–263, 291, 294
- Repository approach, 294
- Resampled matched pair t-test, 225
- Resampling, 166–167, 175–176
- Resampling framework, 171–172, 180
- Resampling statistics, 14–15
- Resubstitution error, 161, 164, 181, 202
- risk, 27–28
- $r \times k$ CV, 261–263
- ROC Analysis, 112–113
- ROC Convex Hull (ROCCH), 122
- ROC curve, 124–128, 134–135, 148–153
- ROC Curve generation, 124–126
- ROCR, 112, 146–148, 152–153, 370–372
- ROC Space, 113–120, 123, 128, 130, 133–134, 136, 314
- Root Mean Squared Error (RMSE), 137–143, 157
- Sample mean, 45–46, 48, 51–53, 57–58, 218, 240
- Sample standard deviation, 47, 60–61, 176, 218
- Sampling distribution, 57–59, 63, 216, 218, 247, 332, 348
- S Coefficient, 90–91, 110
- Scoring classifier, 117–118
- Scott’s π (pi) coefficient, 91
- Semi-supervised learning, 24–25
- Sensitivity, 95–99
- Sign Test, 228–229, 231–233, 236–239, 267, 289, 355
- Silver standard, 89
- Simple and intuitive measure (SIM), 319–320
- Simple resampling, 161–163, 171–173, 176–179, 185–186, 215, 258–260, 290–291
- Single class focus, 87, 94, 102
- Skew, 104–106
- Skew ratio, 110, 112, 119–120, 123, 130, 310
- Specialization, 34, 39
- Specificity, 13, 82, 95
- Standard deviation, 47, 52
- Standard error, 60–61, 251
- Statistical distribution, 46
- Statistical hypothesis testing, 61–62, 65–66, 68, 162, 230, 290
- Statistical significance testing, 6, 14–15, 21, 57, 74, 162, 185, 188, 203, 206

- Statistical learning theory, 16, 202, 308,
 323–325, 329–330
 Statistical table, 22, 351
 Statistical test, 206–215
 Stratified k-fold Cross-Validation, 189, 191, 193
 Structural risk minimization, 29, 30
 Summary statistic, 128–130, 310
 Supervised learning, 1–3, 21, 24–25, 310
- Task specific algorithm, 293–294
 Total variation, 246
 Training set bound, 126–128, 176, 324
 True negative, 19, 96, 102, 104, 113
 True negative rate, 95–96, 113
 True positive, 79, 84, 94–95, 116, 128, 132
 True positive rate, 94–95, 128, 132
 True risk, 27–29, 61, 68, 162, 164–165, 204, 324
 t-table, 261
 t-test, 224
 Tukey Test, 251–253, 257, 275, 362
 Two-tailed test/Two-sided test, 67–68
- Type I error, 14, 69–71, 203, 208, 225–226, 231,
 239, 255, 261, 263, 291, 329
 Type II error, 42, 69–70, 347
- UCI Machine Learning Repository, 7
- variance, 11, 37, 40, 46
 visualization approach, 321–323
 visualization-based combination metric,
 321–323
- WEKA, 7, 13, 22, 73, 83–84, 92, 107–108, 110,
 138, 146–147, 157, 163, 174, 187–188, 273,
 278, 364, 373
 Wilcoxon’s Signed Rank Test, 233–235, 236
 Wilcoxon’s Rank Sum Test, 129
 Wilcoxon Table, 235, 238, 267–268, 356
 Within-group variation, 240, 242
- Zero-one loss, 27, 32, 36, 39, 163, 169
 Z-table, 351