
Index

- Akaike information criterion (AIC), 313
 ascending factorial, 92
 asymptotic mean integrated squared error (AMISE), 262
- bagging, *see* bootstrap aggregation
 Bayes factor, 22, 54 *see also* model selection
 Bayes information criterion (BIC), 313
 Bayes model averaging (BMA), 15, 454
 containment principle, 455
 generalized linear models, 459
 introduction, 454
 kernel methods, 461
 local Bayes factors, 462
 neural networks, 459
 Occam's window, 461
 predictions, 455
 predictive intervals, 456
 recursive partitioning, 460
 splines, 459
 wavelets, 458
 Bayes' rule, 51, 111
 Bayesian paradigm, 35, 51–56
 credible set, 52, 53
 de Finetti's theorem, 51
 highest posterior density predictive region
 normal case, 45
 hypothesis testing, 53
 parameter estimation, 51
 posterior risk, 55
 prediction, 55
 predictive density, 54
 big data, 525
 boosting, 27, 480–489
 Adaboost, 31
 and logistic regression, 485
 classification, 481
 comparisons, 484
 considerations, 486
 regression, 486
 bootstrap aggregation (bagging), 17, 31, 71, 462–471
 average expected prediction error, 464
 generalization error, 467
 kernel methods, 470
 neural networks, 469
 out-of-bag (OOB) error, 466
 random forests, 465, 468
 variable importance, 466
 bootstrapping, 83
 Box–Cox transformation, *see* longitudinal data analysis
- central limit theorem, 49, 53
 clustering, partitioning-around-medoids, 31
 complexity, 520, 600
 credibility interval, 36
 cross-validation (CV), 71, 83
 cumulative predictive error (CPE), 14, 39, 47, 64, 83, 104, 106, 117, 237, 295
- data generator (DG), 3
 data quality, 79
 decision theory
 axioms, 36
 Bayesian, Savage's expected utility theorem, 54
 classical, 25
 decision rule, 6
 frequentist
 Neyman–Pearson, 49
 von Neumann–Morgenstern expected utility theorem, 49, 333
 regret, 79
 Savage's axioms, 333
 sequential quantile maximization, 333
 survey sampling, 57
 deep learning, 349, 482 *and see* neural networks
 dilution, 327, 453
 dimension reduction, 600
 distribution
 asymmetric Laplace, 102
 Bernoulli, 35, 36, 100, 108, 193, 457
 beta, 110, 287–289, 349
 beta-binomial, 349
 binomial, 35, 36, 92, 214, 349, 391
 chi-squared, 43, 95, 163, 176, 198

- Dirichlet, 91, 287, 385
 - Dirichlet process, 91, 219, 222
 - Dirichlet process mixture, 94
 - double exponential, 235
 - Erlang, 155
 - exponential, 151, 207, 227, 240, 245, 578
 - extreme value, 241, 245
 - F, 366
 - gamma, 235, 242, 415
 - gamma process, 235
 - inverse-chi-squared, 42, 46, 198
 - inverse-gamma, 46, 95, 100, 384
 - inverse-Wishart, 195
 - log-gamma, 198
 - log-logistic, 245
 - logistic, 245
 - matrix-variate normal, 180
 - multinomial, 91, 287
 - multivariate normal, 139, 186, 189
 - multivariate Student's t -, 96, 98, 100, 140, 163
 - normal, 42, 93, 94, 100, 110, 162, 165, 177, 203, 242, 292, 384, 410, 415
 - normal-inverse-gamma, 95, 97, 100
 - normal-inverse-Wishart, 163
 - normal-Wishart, 93
 - Pareto, 578
 - Poisson, 151, 181, 193, 578
 - predictive, 37
 - probit, 386
 - scale-free, 578
 - Student's t -, 43, 46, 176, 366
 - noncentral, 110
 - uniform, 36, 88, 283, 415
 - Weibull, 208, 240, 245
 - Wishart, 163–165
 - Zellner's g -prior, 99, 330, 349, 458
- empirical distribution function (EDF), 87
 - ensemble methods, 449 *and see* Bayes model
 - averaging bootstrap aggregation (bagging) stacking
 - comparisons, 471
 - conceptual methodology, 451
 - introduction to, 449
 - median and other averages
 - introduction to, 489
 - median intrinsic Bayes factor, 494
 - median model, 490
 - Occam's window predictor, 492
 - posterior weighted median predictor, 493
 - model list, 452
 - performance, 495
 - variance-bias, 453
 - weak learners vs. greedy learners, 496
 - wisdom of crowds, 450
 - examples of methods
 - classification, 26
 - cosmic ray soil moisture, 142
 - Doppler function, 429
 - food production index, 105
 - heart data, 348
 - hypothesis testing, 21
 - model selection in linear regression, 312
 - Natural Resource Monitoring, Modeling and Management (NRM3) Project, 194
 - NOAA rainfall data, 153
 - prediction, 11
 - simulation study of ensembles, 497
 - soil water content, 295
 - time to relapse, 222, 236, 243
 - Tour de France, 12
 - tumor micro-satellite status, 119
 - Vegout, 433, 507, 524
 - experimental design, 599
 - fallacy of Greek letters, 232
 - final predictive error (FPE), 15, 39, 47, 61, 64, 83
 - frequentist paradigm, 35, 48–51
 - parametric, 41
 - permutation testing, 23
 - plug-in predictor, 35
 - prediction interval
 - general distribution, known sigma, 42
 - normal, known sigma, 42
 - normal, sigma = 1, 43
 - normal, unknown sigma, 42
 - normal, unknown sigma, known mean, 43
 - Gaussian process prior (GPP), *see* nonparametric methods
 - generalized estimating equation (GEE)
 - introduction, 183
 - working correlation matrix, 183
 - generalized linear mixed model
 - estimation, 193
 - logistic regression, 193
 - Poisson regression, 193
 - prediction of new observation, 193
 - generalized linear model (GLM)
 - estimating equation, 181
 - exponential families, 181
 - link function, 181
 - point predictors, 182
 - prediction intervals, 182
 - quasi-likelihood estimation, 183
 - Gibbs sampling, 94
 - graph models, *see* network (graph) models
 - graphical model structure, 491
 - hidden Markov models
 - Baum-Welch algorithm, 447
 - Bayes, 447
 - first-order finite-state discrete-time Markov chain, 446
 - highest-posterior-density predictive region,
 - definition, 44

- hinge loss, 414
- hypothesis testing of expectations, 332
- internet of things (IoT), 547
- kernal density estimation, *see* nonparametric method
- kernel methods
 - and Bayes' rule, 412
 - Bayes, 409
 - Bayes support vector machine (SVM), 413
 - frequentist predictors, 416
 - Gaussian process regression, 412
 - Hilbert space, 407
 - instance-based learners, 409
 - kernel trick, 419
 - Lagrangian, 417
 - margin, 417
 - Mercer's theorem, 413
 - radial basis function, 414
 - Reisz representation theorem, 407
 - relevance vector machines (RVMs), 8, 17, 409, 421
 - representer theorem, 17, 408–409
 - reproducing kernel, 17
 - reproducing kernel Hilbert space (RKHS), 17, 71, 407
 - slack variables, 418
 - support vector machine (SVM), 27
 - for regression, 419
 - support vectors, 418
- kernel regression, *see* nonparametric methods 281
- L_1 distance, 199
- Laplace approximation, 45
- linear mixed models
 - Bayesian, 185
 - best linear unbiased predictor (BLUP), 186
 - conditional expectations, 185
 - empirical BLUP, 188
 - Henderson's equations, 187
 - hierarchical, 185
 - log-likelihood, 186
 - maximum likelihood estimates, 187
 - prediction, 189–190
 - restricted maximum likelihood (REML), 188
 - Shi–Weiss–Taylor, 191–192
- logistic regression, 108–111
 - Bayes' rule, 108
 - Bayesian, 110
 - classification, 109
 - frequentist, 109
- longitudinal data analysis
 - balanced data, 167
 - Bayesian
 - posterior, 163
 - predictive, 163
 - frequentist, 163
 - generalization of time series, 161
 - generalized linear mixed models, 193
 - growth curve models, 172, 202
 - Cole and Green, 204
 - fixed effect, 203
 - introduction, 161
 - linear models
 - compound symmetry, 173
 - introduction, 172
 - maximum likelihood estimates, 173
 - prediction, 175–180
 - restricted maximum likelihood (REML), 175
 - sampling distributions, 175
 - measurement and random error, 167
 - no covariates, 166
 - random effects, 184
 - repeated measures ANOVA, 167
 - split-plot
 - Bayesian, 171
 - design, 167
 - homogenous compound symmetry, 169
 - prediction, 169–171
- Markov chain Monte Carlo (MCMC), 344–353
 - candidate generating distribution, 344
 - Gibbs sampler, 346
 - Metropolis–Hastings (MH), 344–347
 - vs. SA, 344
 - reversibility condition, 344
- Metropolis–Hastings, 344 *see* Markov chain Monte Carlo
- model accessibility classes
 - \mathcal{M} -closed, 68, 86, 123
 - \mathcal{M} -complete, 69, 123
 - \mathcal{M} -complete vs. \mathcal{M} -open, 300
 - \mathcal{M} -mixed, 72
 - \mathcal{M} -open, 70
 - complexity, 72
 - validation, 72
- model averaging, 71, 449 *see also* Bayes model averaging
- model identification, overview, 5
- model selection, 16
 - as the standard statistical problem, 307
 - assumptions, 308
 - bail-out effect, 314
 - Bayes
 - modal model, 328
 - testing, 330
 - within and across model priors, 328
 - Bayes factors, 332
 - consistency, 325
 - cross-validation (CV)
 - J -fold CV, 336
 - choice of number of folds, 337
 - consistency, 339
 - expected prediction error, 335
 - introduction, 334

- leave-one-out (LOO) CV, 336
- one-standard-error rule, 338
- robustness, 338
- variability of CV error, 337
- variance–bias tradeoff, 339
- effective dimension, 354
- examination of procedures, 313
- expected risk, 333
- information criteria, 320–327
 - Akaike information criterion (AIC), 313, 321
 - Bayes information criterion (BIC), 16, 313, 329
 - Bethel–Shumway, 321–325
 - deviance information criterion (DIC), 354
 - focused information criterion (FIC), 320
 - information complexity, 321
 - posterior predictive information criterion (PPIC), 355
 - Takeuchi information criterion, 321
- information-theoretic
 - description length, 355
 - Shannon code length, 355
- linear models, 312
- log scoring rule, 356
- median model, 327
- sampling distribution visualization, 317
- Shtarkov solution, 309
- stepwise, 313
- sure independence screening, 319
- vs. predictor selection, 309
- model uncertainty
 - sparsity, 10
 - variance–bias tradeoff, 11
 - vs. modeling information, 119
- modeling
 - as predictor generation, 120
 - Bayes
 - bias, 329
 - calibration, 329
 - scoring rules, 329
 - bootstrapping, 326
 - dangers, 75
 - dilution, 327, 338
 - justification, 73
 - model assessment, 310
 - generalization error, 311
 - model identification, 311
 - predictuals, 308
- model-selection procedures (MSPs), 308
- multidimensional scaling (MDS), 316
- multitype data, 585–599
 - analytical goal, 586
 - informationally disjoint, 595
 - interpretability, 594–596
 - double sparsity, 594
 - shrinkage, 595
 - meta-dimensional analysis, 593
 - multistaged analysis, 593
- omics data, 586–592
 - next-generation sequencing (NGS), 588
 - proteomics, 589
 - transcriptomics, 590
- supervised, 592
- unsupervised
 - Bayes cluster ensembles, 598
 - Bayes consensus clustering, 598
 - consensus clustering, 596
 - integrated clustering, 597
 - integrating phenotype framework, 599
- nearest neighbors, 30, 115–119, 275–286
 - classification, 304
 - density estimation, 275–281
 - AMISE, 279
 - asymptotics, 276
 - confidence interval, 280
 - distances, 276
 - prediction, 280
 - selection of k , 276, 280
 - smoothed, 276
 - regression, 282–285
 - AMISE, 282
 - asymptotics, 282
 - prediction, 284
 - selection of k , 282
 - sequential, 285–286
- network (graph) models, 570–585
 - adjacency matrix, 570
 - affinity matrix, 574
 - bipartite graph, 571
 - block model, 574
 - Boolean, 571
 - Clifford–Hammersley theorem, 574
 - degree, 570
 - directed graphs, p_1 and p_2 models, 576
 - distance between vertices, 570
 - dynamic networks, 581–585
 - dynamic preferential attachment, 581
 - edge-oriented dynamics, 583
 - generalized block model, 582
 - graph Laplacian, 584
 - intensity matrix, 582
 - linear dynamical system, 584
 - potential function, 583
 - prediction, 584
 - random graph model, 582
 - vertex-oriented dynamics, 583
 - graph prediction via CV, 579
 - isomorphism, 570
 - motifs, 571
 - power law, 577
 - static
 - Horvitz–Thompson estimator, 572
 - random graphs, 572
 - simple random sampling, 572
 - static vs. dynamic, 571

- walk and path, 570
- neural networks, 386–399
 - architecture selection, 393
 - automatic relevance determination, 395
 - backwards propagation, 389
 - Bayes, 395–397
 - Boltzmann machine, 404
 - curse of dimensionality, 397
 - deep learning
 - convolution, 402
 - convolutional neural network (CNN), 401
 - deep belief networks (DBN), 404
 - feature map, 402
 - filter or kernel, 402
 - generative adversarial neural network (GAN), 404
 - local receptive fields, 401
 - NN vs. CNN, 403
 - pooling layer, 402
 - dilution, 397
 - feedforward, 388
 - fitting, 388
 - implementation, 399
 - interpretation, 398
 - minimum description length (MDL), 393
 - multitype data, 398
 - Occam's window, 396
 - prediction intervals, 392, 393
 - recursive neural network (RNN), 404
 - regularization, 388
 - self-organizing neural networks, 393
 - simulated annealing (SA), 392
 - stability, 398
 - transfer function, 387
 - vanishing gradient problem, 400
- no free lunch theorems, 522
- nonpredictive statistics, 62
- nonlinear mixed models, 194
- nonlinear regression, *see* recursive partitioning
 - (trees), neural networks, kernel methods, penalized methods
 - classical, 365–368
 - errors-in-variables, 447
 - projection pursuit, 443
- nonparametric methods, *see* orthonormal basis
 - expansions, nearest neighbors
 - Bayes, 286–294
 - Dirichlet process prior (DPP), 286
 - Gaussian process prior (GPP), 287, 291–294
 - Polya tree, 287
 - Polya tree process priors, 288–291
 - predictive distribution, 286
 - regression, 287
 - global vs. local, 251
 - Hilbert space, 252
 - kernel density estimation, 260–266
 - AMISE, 262
 - bias, 261
 - boxcar kernel, 261
 - confidence interval, 264
 - dominated convergence theorem (DCT), 265
 - Epanechnikov kernel, 261
 - multivariate, 263
 - pointwise convergence, 263
 - prediction, 265
 - specification, 261
 - variance, 261
 - kernel regression, 266–274
 - Nadaraya–Watson (NW), 270–274
 - Priestley–Chao, 266–270
 - model averaging, 298
 - model classes, 249
 - point predictors, 90
 - predictive interval, 250
 - Bayes, 92
 - Bayes as mixture of DPPs, 93
 - frequentist density estimation, 88
 - frequentist EDF, 87
 - from Bayes posterior mean, 92
 - predictor reselection, 298
 - smoothing splines, 303–304
 - cubic, 303
 - point predictor, 304
 - testing as prediction
 - Bayes, 304
 - frequentist, 305
 - tolerance intervals, 250
- online data, *see* streaming data
- orthonormal basis expansions, 252–260
 - bases as explanatory variables, 255
 - bias, 256
 - confidence interval, 255
 - curse of dimensionality, 256
 - estimation and prediction, 253
 - estimation with modulator, 255
 - global vs. local, 259
 - hard thresholding, 256
 - partitioning estimate, 259
 - random x , 257
 - Runge function, 260
 - Sobolev ellipsoids, 256
 - variance–bias tradeoff, 254
- paradigms, *see* frequentist, Bayesian, predictive, survey sampling
 - fiducial, 41
 - information-theoretic, 41
 - sunrise problem, 34–40
- penalized methods, 422–429
 - adaptive LASSO (ALASSO), 424
 - Bayes, 423
 - least absolute shrinkage and selection operator (LASSO), 425
 - oracle property, 424
 - ridge regression, 405, 422

- shrinkage, 422
- smoothly clipped absolute deviation (SCAD), 70, 424, 426
- super-efficiency, 428
- prediction
 - adaptive predictor reselection, 601
 - along a string
 - regret, 521
 - Shtarkov solution, 520
 - classification
 - area under the ROC curve (AUC), 113
 - Bayes, 111
 - Bayes linear discriminant analysis (LDA), 113
 - Bayes linear discriminant function (LDF), 114
 - comparisons, 116
 - nearest neighbor, 115
 - receiver operator characteristic (ROC) curve, 112
 - sensitivity, 112
 - specificity, 112
 - decision rule, 6
 - limits to, 601
 - linear regression, 94–101
 - Bayesian, 95
 - Bayesian noninformative, 97
 - Bayesian with Jeffrey's prior, 99
 - Bayesian with spike and slab, 100
 - Bayesian with Zellner's g -prior, 99
 - Bayesian with Zellner's informative g -prior, 100
 - frequentist, 94
 - minimax prediction, 310
 - of extreme values, 495
 - predictive density, 44
 - predictor evaluation, 601
 - predictor uncertainty, 47
 - quantile regression
 - Bayesian, 102
 - frequentist, 101
 - uncertainty, 8
 - validation, 525
 - versus interpretability, 7
 - versus modeling, 32
- predictive paradigm, 39, 58–63, 87
 - estimation, 62
 - expected loss, 63
 - hypothesis testing, 62
 - model selection, 59, 61
 - motivation, 4
 - predictor assessment, 59
 - unified framework, 63
 - versus modeling, 9
- predictual, 137, 309, 513
- prequentialism, 80–85
 - adaptive, 84
 - error, 82
 - estimation, 80
 - generalization error, 81
 - model-based criteria, 83
 - Prequential Principle, 64, 82, 299
 - sequential prediction, 80
 - training error, 81
 - variance–bias tradeoff, 84
 - probability theory, new, 603
- R statistical language, 104
- random forests, *see* bootstrap aggregation (bagging)
- recommender systems, 526–537
 - collaborative filtering, 526–530
 - association rules, 528
 - cosine similarity, 528
 - singular value decomposition, 528
 - content-based, 530–533
 - Dice coefficient, 532
 - naive Bayes, 533
 - Ruche's algorithm, 532
 - term-frequency inverse-document frequency (TF-IDF), 531
 - evaluation, 536
 - hybrid, 535
 - knowledge-based, 534–535
- recursive partitioning (trees), 27, 368–386
 - Bayes, 383
 - Bayesian additive regression tree (BART), 385
 - bootstrapping, 374
 - classification error, 377
 - clustering, 372
 - cost-complexity, 379
 - entropy, 377
 - Gini index, 376
 - logic trees, 445–446
 - model selection, 379
 - optimality, 373
 - out-of-bag (OOB) error, 381
 - prediction intervals, 374
 - pruning, 379
 - random forests, 27, 348
 - Shannon mutual information, 378
 - simulation extrapolation (simex), 382
 - splitting and splitting rules, 371, 375, 380
 - variable importance factor (VIF), 381
 - vs. linear models, 369
 - vs. nonparametric models, 370
- regression
 - alternatives to least squares, 495
- regret, 71
- relevance vector machines (RVMs), *see* kernel methods
- remote sensing, 562–565
 - area to point prediction, 564
 - electromagnetic radiation (EMR), 563
 - radiance, 563
 - regularization, 563
 - rescaling, 563
 - super-resolution mapping, 564

- representer theorem, 17
- risk, 465
- sensor data, 547–550
 - data snapshot framework, 548
 - data types, 548
 - events, 550
 - inference, 548
 - missing data, 549
 - trend cluster, 549
 - windowing, 548
- silhouette distance, 31
- simulated annealing (SA)
 - and Markov chain, 340
 - and Metropolis–Hastings (MH), 340
 - convergence, 340
 - cooling schedule, 340
 - prediction, 343
 - simulation–extrapolation (simex), 343
 - stationary distribution, 340
 - temperature, 340
- sources of information, 602
- spatio-temporal data, 556–562
 - areal data, 568
 - Bayes, 569
 - conditionally autoregressive, 568
 - proper Gaussian Markov random field, 569
 - classes, 556
 - dynamic spatio-temporal model (DSTM), 562
 - GLMs, 561
 - innovations, 561
 - Kriging, 557–558, 560–561
 - measurement equations, 562
 - orthogonal functions, 559
 - point data, 559
 - point process, 565–568
 - K -function, 567
 - inhomogeneous Poisson, 567
 - intensity function, 566
 - kernel estimators, 566
 - marked, 565
 - second-order intensity, 566
 - process model, 561
 - spectral measures, 561
 - separability, 561
 - transformation equations, 562
 - variogram, 561
- stacking, 471–479
 - model list, 473
 - regularization, 475
 - super learner, 477
 - weights, 475
- streaming data, 537–547
 - decisions, 551–556
 - anomaly detection, 551, 554
 - automated feature extraction, 555
 - theory, 554
 - estimation
 - Partition procedure, 545
 - Quickselect procedure, 545
 - count-min sketch, 541
 - frequency moments, 542
 - frequency of each element, 541
 - median of medians, 545
 - Munro–Paterson, 545
 - number of distinct elements, 540
 - tug-of-war procedure, 543
 - turnstile model, 541
 - hash function, 538
 - multiset, 540
 - performance assessment, 539
 - representative sample, 546
 - doubling algorithm, 546
 - sketch, 541
- support vector machines (SVMs), 27, 28
- survey sampling paradigm, 37, 56–58
 - confidence interval, 56
 - estimation, 56
 - hypothesis testing, 57
 - prediction, 57
- survival analysis
 - accelerated failure time, 245–246
 - censoring, 206
 - competing risks, 246–248
 - Cox proportional hazards, 208
 - hazard function, 207, 226
 - lifetime, 206
 - nonparametric
 - Aalen estimator, 216
 - Bayes, 219–221
 - calibration, 219, 221
 - conditional survival function, 215
 - confidence interval, 217, 218
 - consistency, 210
 - delta method, 213, 215, 216
 - discrimination, 221
 - Greenwood’s formula, 213, 215, 216
 - Kaplan–Meier (KM) predictor, 208–210, 217
 - L^1 loss, 224
 - martingale, 214
 - median as predictor, 216
 - model uncertainty, 218
 - Nelson–Aalen estimator, 209
 - predictive distribution, 218
 - predictive interval, 211, 213, 215
 - probability forecasting, 218
 - prognostic models, 221
 - receiver operating characteristic (ROC) curve, 221
 - parametric, 239–245
 - Bayes, 241–243
 - exponential, 240
 - Weibull, 241
 - proportional hazards, 226–236
 - baseline hazard function, 227
 - Bayes, 233–236

- Bayes posterior distribution, 235
- caveats, 233
- conditional likelihood, 229
- confidence interval, 232
- constant of proportionality, 227
- Cox models, 227
- factorization of the hazard function, 227
- Fisher information, 230
- frequentist estimation, 229
- likelihood, 228
- marginal likelihood, 229
- partial likelihood, 229
- plug-in predictor, 231
- predictive interval, 231
- profile likelihood, 229
- risk set, 229
- survival function, 206
- symmetric positive semidefinite matrix, 292
- testing error, 81
- time series analysis
 - ARMAX and (S)ARFIMA, 156
 - Bayes general ARMA, 139
 - Bayes SARIMA, 139–142
 - Box–Jenkins, 128–138
 - ARMA, 128
 - autocorrelation function (ACF), 131
 - autoregressive process, 130
 - backshift, 128
 - forecasting, 135–138
 - moving average process, 130
 - nonseasonal and seasonal polynomials, 129
 - parameter estimation, 132
 - partial autocorrelation function (PACF), 131
 - prediction, 129
 - predictor class identification, 129
 - SARIMA, 128
 - SARMA, 129
 - stationarity, 130
 - validation, 133
 - classical decomposition model, 125–128
 - backshift operator, 127
 - detrending, 126
 - differencing, 127
 - exponential smoothing, 126
 - moving average, 126
 - seasonality, 125, 127
 - second-order stationary, 125
 - trend, 125, 127
 - dynamic linear models, 159
 - forecasting, 159
 - observation equation, 159
 - system equation, 159
 - linear regression, 157–159
 - Bayesian, 158
 - frequentist, 157
 - stochastic modeling, 150
 - compound Poisson process, 151
 - Poisson process, 151
 - prediction, 152
 - probabilistic forecasts, 152
 - proper scoring rule, 153
- ultrametric space, 525
- ultrametric topology, 519
- validation, 47
- Vapnik’s ϵ -insensitive loss, 439
- Vapnik–Chervonenkis dimension, 65, 84
- variance–bias decomposition, 64
- variance–bias tradeoff, 77, 80
- weak learner, 386