
Subject Index

- abc method, 194, 204
- Accelerated gradient descent, 359
- Acceleration, 192, 205
- Accuracy, 14
 - after model selection, 402–408
- Accurate but not correct, 402
- Activation function, 355, 361
 - leaky rectified linear, 362
 - rectified linear, 362
 - ReLU, 362
 - tanh, 362
- Active set, 302, 308
- adaboost** algorithm, 341–345, 447
- Adaboost.M1, 342
- Adaptation, 404
- Adaptive estimator, 404
- Adaptive rate control, 359
- Additive model, 324
 - adaptive, 346
- Adjusted compliance, 404
- Admixture modeling, 256–260
- AIC, *see* Akaike information criterion
- Akaike information criterion, 208, 218, 226, 231, 246, 267
- Allele frequency, 257
- American Statistical Association, 449
- Ancillary, 44, 46, 139
- Apparent error, 211, 213, 219
- arcsin transformation, 95
- Arthur Eddington, 447
- Asymptotics, xvi, 119, 120
- Autoencoder, 362–364
- Backfitting, 346
- Backpropagation, 356–358
- Bagged estimate, 404, 406
- Bagging, 226, 327, 406, 408, 419
- Balance equations, 256
- Barycentric plot, 259
- Basis expansion, 375
- Bayes
 - deconvolution, 421–424
 - factor, 244, 285
 - false-discovery rate, 279
 - posterior distribution, 254
 - posterior probability, 280
 - shrinkage, 212
 - t*-statistic, 255
 - theorem, 22
- Bayes–frequentist estimation, 412–417
- Bayesian
 - inference, 22–37
 - information criterion, 246
 - lasso, 420
 - lasso prior, 415
 - model selection, 244
 - trees, 349
- Bayesian information criterion, 267
- Bayesianism, 3
- BCa
 - accuracy and correctness, 206
 - confidence density, 202, 207, 237, 242, 243
 - interval, 202
 - method, 192
- Benjamini and Hochberg, 276
- Benjamini–Yekutieli, 400
- Bernoulli, 338
- Best-approximating linear subspace, 363
- Best-subset selection, 299
- Beta
 - distribution, 54, 239
- BH_q, 276
- Bias, 14, 352
- Bias-corrected, 330
 - and accelerated, *see* BCa method
 - confidence intervals, 190–191
 - percentile method, 190

- Bias-correction value, 191
- Biased estimation, 321
- BIC, *see* Bayesian information criterion
- Big-data era, xv, 446
- Binomial, 109, 117
 - distribution, 54, 117, 239
 - log-likelihood, 380
 - standard deviation, 111
- Bioassay, 109
- Biometrika*, 449
- Bivariate normal, 182
- Bonferroni bound, 273
- Boole's inequality, 274
- Boosting, 320, 324, 333–350
- Bootstrap, 7, 155–180, 266, 327
 - Baron Munchausen, 177
 - Bayesian, 168, 179
 - cdf, 187
 - confidence intervals, 181–207
 - ideal estimate, 160, 179
 - jackknife after, 179
 - moving blocks, 168
 - multisample, 167
 - nonparametric, 159–162, 217
 - out of bootstrap, 232
 - packages, 178
 - parametric, 169–173, 223, 312, 429
 - probabilities, 164
 - replication, 159
 - sample, 159
 - sample size, 179, 205
 - smoothing, 226, 404, 406
 - t , 196
 - t intervals, 195–198
- Bound form, 305
- Bounding hyperplane, 398
- Burn-in, 260
- BY_q algorithm, 400
- Causal inference, xvi
- Censored
 - data, 134–139
 - not truncated, 150
- Centering, 107
- Central limit theorem, 119
- Chain rule for differentiation, 356
- Classic statistical inference, 3–73
- Classification, 124, 209
- Classification accuracy, 375
- Classification error, 209
- Classification tree, 348
- Cochran–Mantel–Haenszel test, 131
- Coherent behavior, 261
- Common task framework, 447
- Compliance, 394
- Computational bottleneck, 128
- Computer age, xv
- Computer-intensive, 127
 - inference, 189, 267
 - statistics, 159
- Conditional, 58
- Conditional distribution
 - full, 253
- Conditional inference, 45–48, 139, 142
 - lasso, 318
- Conditionality, 44
- Confidence
 - density, 200, 201, 235
 - distribution, 198–203
 - interval, 17
 - region, 397
- Conjugate, 253, 259
 - prior, 238
 - priors, 237
- Convex optimization, 304, 308, 321, 323, 377
- Convolution, 422, 445
 - filters, 368
 - layer, 367
- Corrected differences, 411
- Correlation effects, 295
- Covariance
 - formula, 312
 - penalty, 218–226
- Coverage, 181
- Coverage level, 274
- Coverage matching prior, 236–237
- Cox model, *see* proportional hazards model
- C_p , 217, 218, 221, 231, 267, 300, 394, 395, 403
- Cramér–Rao lower bound, 44
- Credible interval, 198, 417
- Cross-validation, 208–232, 267, 335
 - 10-fold, 326
 - estimate, 214
 - K -fold, 300
 - leave one out, 214, 231
- Cumulant generating function, 67
- Curse of dimensionality, 387
- Dark energy, 210, 231
- Data analysis, 450
- Data science, xvii, 450, 451

- Data sets
 - ALS**, 334
 - AML**, *see* leukemia
 - baseball**, 94
 - butterfly**, 78
 - cell infusion**, 112
 - cholesterol**, 395, 402, 403
 - CIFAR-100**, 365
 - diabetes**, 98, 209, 396, 414, 416
 - dose-response**, 109
 - galaxy**, 120
 - handwritten digits**
(**MNIST**), 353
 - head/neck cancer**, 135
 - human ancestry**, 257
 - insurance**, 131
 - kidney function**, 157, 222
 - leukemia**, 176, 196, 377
 - NCOG**, 134
 - nodes**, 424, 427, 430, 438, 439, 442
 - pediatric cancer**, 143
 - police**, 287
 - prostate**, 249, 272, 289, 408, 410, 423, 434–436
 - protein classification**, 385
 - shakespeare**, 81
 - spam**, 113, 127, 209, 215, 300–302, 325
 - student score**, 173, 181, 186, 202, 203
 - supernova**, 210, 212, 217, 221, 224
 - vasoconstriction**, 240, 241, 246, 252
- Data snooping, 398
- De Finetti, B., 35, 36, 251, 450
- De Finetti–Savage school, 251
- Debias, 318
- Decision rule, 275
- Decision theory, xvi
- Deconvolution, 422
- Deep learning, 351–374
- Definitional bias, 431
- Degrees of freedom, 221, 231, 312–313
- Delta method, 15, 414, 420
- Deviance, 112, 118, 119, 301
- Deviance residual, 123
- Diffusion tensor imaging, 291
- Direct evidence, 105, 109, 421
- Directional derivatives, 158
- Distribution
 - beta, 54, 239
 - binomial, 54, 117, 239
 - gamma, 54, 117, 239
 - Gaussian, 54
 - normal, 54, 117, 239
 - Poisson, 54, 117, 239
- Divide-and-conquer algorithm, 325
- Document retrieval, 298
- Dose-response, 109
- Dropout learning, 368, 372
- DTI, *see* diffusion tensor imaging
- Early computer-age, xvi, 75–268
- Early stopping, 362
- Effect size, 272, 288, 399, 408
- Efficiency, 44, 120
- Eigenratio, 162, 173, 194
- Elastic net, 316, 356
- Ellipsoid, 398
- EM algorithm, 146–150
 - missing data, 266
- Empirical Bayes, 75–90, 93, 264
 - estimation strategies, 421–445
 - information, 443
 - large-scale testing, 278–282
- Empirical null, 286
 - estimation, 289–290
 - maximum-likelihood estimation, 296
- Empirical probability distribution, 160
- Ensemble, 324, 334
- Ephemeral predictors, 227
- Epoch, 359
- Equilibrium distribution, 256
- Equivariant, 106
- Exact inferences, 119
- Expectation parameter, 118
- Experimental design, xvi
- Exponential family, 53–72, 225
 - p -parameter, 117, 413, 424
 - curved, 69
 - one-parameter, 116
- F distribution, 397
- F tests, 394
- f -modeling, 424, 434, 440–444
- Fake-data principle, 148, 154, 266
- False coverage
 - control, 399
- False discovery, 275
 - control, 399
 - control theorem, 294
 - proportion, 275
 - rate, 271–297
- False-discovery

- rate, 9
- Family of probability densities, 64
- Family-wise error rate, 274
- FDR, *see* false-discovery rate
- Feed-forward, 351
- Fiducial, 267
 - constructions, 199
 - density, 200
 - inference, 51
- Fisher, 79
- Fisher information, 29, 41, 59
 - bound, 41
 - matrix, 236, 427
- Fisherian correctness, 205
- Fisherian inference, 38–52, 235
- Fixed-knot regression splines, 345
- Flat prior, 235
- Forward pass, 357
- Forward-stagewise, 346
 - fitting, 320
- Forward-stepwise, 298–303
 - computations, 322
 - logistic regression, 322
 - regression, 300
- Fourier
 - method, 440
 - transform, 440
- Frailties, 439
- Frequentism, 3, 12–22, 30, 35, 51, 146, 267
- Frequentist, 413
 - inference, 12–21
 - strongly, 218
- Fully connected layer, 368
- Functional gradient descent, 340
- FWER, *see* family-wise error rate
- g*-modeling, 423
- Gamma, 117
 - distribution, 54, 117, 239
- General estimating equations, xvi
- General information criterion, 248
- Generalized
 - linear mixed model, 437–440
 - linear model, 108–123, 266
 - ridge problem, 384
- Genome, 257
- Genome-wide association studies, 451
- Gibbs sampling, 251–260, 267, 414
- GLM, *see* generalized linear model
- GLMM, *see* generalized linear mixed model
- Google flu trends, 230, 232
- Gradient boosting, 338–341
- Gradient descent, 354, 356
- Gram matrix, 381
- Gram-Schmidt orthogonalization, 322
- Graphical lasso, 321
- Graphical models, xvi
- Greenwood's formula, 137, 151
- Group lasso, 321
- Hadamard product, 358
- Handwritten digits, 353
- Haplotype estimation, 261
- Hazard rate, 131–134
 - parametric estimate, 138
- Hidden layer, 351, 352, 354
- High-order interaction, 325
- Hinge loss, 380
- Hints
 - learning with, 369
- Hoeffding's lemma, 118
- Holm's procedure, 274, 294
- Homotopy path, 306
- Hypergeometric distribution, 141, 152
- Imputation, 149
- Inadmissible, 93
- Indirect evidence, 102, 109, 266, 290, 421, 440, 443
- Inductive inference, 120
- Inference, 3
- Inference after model selection, 394–420
- Inferential triangle, 446
- Infinitesimal forward stagewise, 320
- Infinitesimal jackknife, 167
 - estimate, 406
 - standard deviations, 407
- Influence function, 174–177
 - empirical, 175
- Influenza outbreaks, 230
- Input distortion, 369, 373
- Input layer, 355
- Insample error, 219
- Inverse chi-squared, 262
- Inverse gamma, 239, 262
- IRLS, *see* iteratively reweighted least squares
- Iteratively reweighted least squares, 301, 322
- Jackknife, 155–180, 266, 330
 - estimate of standard error, 156
 - standard error, 178

- James–Stein
 - estimation, 91–107, 282, 305, 410
 - ridge regression, 265
- Jeffreys
 - prior, 237
- Jeffreys’
 - prior, 28–30, 36, 198, 203, 236
 - prior, multiparameter, 242
 - scale, 285
- Jumpiness of estimator, 405
- Kaplan–Meier, 131, 134, 136, 137
 - estimate, 134–139, 266
- Karush–Kuhn–Tucker optimality
 - conditions, 308
- Kernel
 - function, 382
 - logistic regression, 386
 - method, 375–393
 - SVM, 386
 - trick, 375, 381–383, 392
- Kernel smoothing, 375, 387–390
- Knots, 309
- Kullback–Leibler distance, 112
- ℓ_1 regularization, 321
- Lagrange
 - dual, 381
 - form, 305, 308
 - multiplier, 391
- Large-scale
 - hypothesis testing, 271–297
 - testing, 272–275
- Large-scale prediction algorithms, 446
- Lasso, 101, 210, 217, 222, 231, 298–323
 - modification, 312
 - path, 312
 - penalty, 356
- Learning from the experience of others,
 - 104, 280, 290, 421, 443
- Learning rate, 358
- Least squares, 98, 112, 299
- Least-angle regression, 309–313, 321
- Least-favorable family, 262
- Left-truncated, 150
- Lehmann alternative, 294
- Life table, 131–134
- Likelihood function, 38
 - concavity, 118
- Limited-translation rule, 293
- Lindsey’s method, 68
- Linearly separable, 375
- Link function, 237, 340
- Local false-discovery rate, 280, 282–286
- Local regression, 387–390, 393
- Local translation invariance, 368
- Log polynomial regression, 410
- Log-rank statistic, 152
- Log-rank test, 131, 139–142, 152, 266
- Logic of inductive inference, 185, 205
- Logistic regression, 109–115, 139, 214,
 - 299, 375
 - multiclass, 355
- Logit, 109
- Loss plus penalty, 385
- Machine learning, 208, 267, 375
- Mallows’ C_p , *see* C_p
- Mantel–Haenszel test, 131
- MAP, 101
- MAP estimate, 420
- Margin, 376
- Marginal density, 409, 422
- Markov chain Monte Carlo, *see* MCMC
- Markov chain theory, 256
- Martingale theory, 294
- Matching prior, 198, 200
- Matlab, 271
- Matrix completion, 321
- Max pool layer, 366
- Maximized a-posteriori probability, *see* MAP
- Maximum likelihood, 299
- Maximum likelihood estimation, 38–52
- MCMC, 234, 251–260, 267, 414
- McNemar test, 341
- Mean absolute deviation, 447
- Median unbiased, 190
- Memory-based methods, 390
- Meter reader, 30
- Meter-reader, 37
- Microarrays, 227, 271
- Minitab, 271
- Misclassification error, 302
- Missing data, 146–150, 325
 - EM algorithm, 266
- Missing-species problem, 78–84
- Mixed features, 325
- Mixture density, 279
- Model averaging, 408
- Model selection, 243–250, 398
 - criteria, 250
- Monotone lasso, 320
- Monotonic increasing function, 184
- Multinomial

- distribution, 61–64, 425
- from Poisson, 63
- Multiple testing, 272
- Multivariate
 - analysis, 119
 - normal, 55–59
- n*-gram, 385
- N-P complete, 299
- Nadaraya–Watson estimator, 388
- Natural parameter, 116
- Natural spline model, 430
- NCOG, *see* Northern California Oncology Group
- Nested models, 299
- Neural Information Processing Systems, 372
- Neural network, 351–374
 - adaptive tuning, 360
 - number of hidden layers, 361
- Neurons, 351
- Neyman’s construction, 181, 183, 193, 204
- Neyman–Pearson, 18, 19, 293
- Non-null, 272
- Noncentral chi-square variable, 207
- Nonlinear transformations, 375
- Nonlinearity, 361
- Nonparameteric
 - regression, 375
- Nonparametric, 53, 127
 - MLE, 150, 160
 - percentile interval, 187
- Normal
 - correlation coefficient, 182
 - distribution, 54, 117, 239
 - multivariate, 55–59
 - regression model, 414
 - theory, 119
- Northern California Oncology Group, 134
- Nuclear norm, 321
- Nuisance parameters, 142, 199
- Objective Bayes, 36, 267
 - inference, 233–263
 - intervals, 198–203
 - prior distribution, 234–237
- OCR, *see* optical character recognition
- Offset, 349
- OLS
 - algorithm, 403
 - estimation, 395
 - predictor, 221
- One-sample nonparametric bootstrap, 161
- One-sample problems, 156
- OOB, *see* out-of-bag error
- Optical character recognition, 353
- Optimal separating hyperplane, 375–377
- Optimal-margin classifier, 376
- Optimality, 18
- Oracle, 275
- Orthogonal parameters, 262
- Out-of-bag error, 232, 327, 329–330
- Out-the-box learning algorithm, 324
- Output layer, 352
- Outsample error, 219
- Over parametrized, 298
- Overfitting, 304
- Overshrinks, 97
- p*-value, 9, 282
- Package/program
 - gbm**, 335, 348
 - glmnet**, 214, 315, 322, 348
 - h2o**, 372
 - lars**, 312, 320
 - liblineaR**, 381
 - locfdr**, 289–291, 296, 437
 - lowess**, 6, 222, 388
 - nlm**, 428
 - randomForest**, 327, 348
 - selectiveInference**, 323
- Pairwise inner products, 381
- Parameter space, 22, 29, 54, 62, 66
- Parametric bootstrap, 242
- Parametric family, 169
- Parametric models, 53–72
- Partial likelihood, 142, 145, 151, 153, 266, 341
- Partial logistic regression, 152
- Partial residual, 346
- Path-wise coordinate descent, 314
- Penalized
 - least squares, 101
 - likelihood, 101, 428
 - logistic regression, 356
 - maximum likelihood, 226, 307
- Percentile method, 185–190
 - central interval, 187
- Permutation null, 289, 296
- Permutation test, 49–51
- Phylogenetic tree, 261
- Piecewise

- linear, 313
- nonlinear, 314
- Pivotal
 - argument, 183
 - quantity, 196, 198
 - statistic, 16
- .632 rule, 232
- Poisson, 117, 193
 - distribution, 54, 117, 239
 - regression, 120–123, 249, 284, 295, 435
- Poisson regression, 171
- Polynomial kernel, 382, 392
- Positive-definite function, 382
- Post-selection inference, 317, 394–420
- Posterior density, 235, 238
- Posterior distribution, 416
- Postwar era, 264
- Prediction
 - errors, 216
 - rule, 208–213
- Predictors, 124, 208
- Principal components, 362
- Prior distribution, 234–243
 - beta, 239
 - conjugate, 237–243
 - coverage matching, 236–237
 - gamma, 239
 - normal, 239
 - objective Bayes, 234
 - proper, 239
- Probit analysis, 112, 120, 128
- Propagation of errors, 420
- Proper prior, 239
- Proportional hazards model, 131, 142–146, 266
- Proximal-Newton, 315
- q -value, 280
- QQ plot, 287
- QR decomposition, 311, 322
- Quadratic program, 377
- Quasilikelihood, 266
- Quetelet, Adolphe, 449
- \mathbf{R} , 178, 271
- Random forest, 209, 229, 324–332, 347–350
 - adaptive nearest-neighbor estimator, 328
 - leave-one-out cross-validated error, 329
 - Monte Carlo variance, 330
 - sampling variance, 330
 - standard error, 330–331
- Randomization, 49–51
- Rao–Blackwell, 227, 231
- Rate annealing, 360
- Rectified linear, 359
- Regression, 109
- Regression rule, 219
- Regression to the mean, 33
- Regression tree, 124–128, 266, 348
- Regularization, 101, 173, 298, 379, 428
 - path, 306
- Relevance, 290–293
- Relevance function, 293
- Relevance theory, 297
- Reproducing kernel Hilbert space, 375, 384, 392
- Resampling, 162
 - plans, 162–169
 - simplex, 164, 169
 - vector, 163
- Residual deviance, 283
- Response, 124, 208
- Ridge regression, 97–102, 209, 304, 327, 332, 372, 381
 - James–Stein, 265
- Ridge regularization, 368
 - logistic regression, 392
- Right-censored, 150
- Risk set, 144
- RKHS, *see* reproducing-kernel Hilbert space
- Robbins’ formula, 75, 77, 422, 440
- Robust estimation, 174–177
- Royal Statistical Society, 449
- S language, 271
- Sample correlation coefficient, 182
- Sample size coherency, 248
- Sampling distribution, 312
- SAS, 271
- Savage, L. J., 35, 36, 51, 199, 233, 251, 450
- Scale of evidence
 - Fisher, 245
 - Jeffreys, 245
- Scheffé
 - interval, 396, 397, 417
 - theorem, 398
- Score function, 42
- Score tests, 301
- Second-order accuracy, 192–195

- Selection bias, 33, 408–411
- Self-consistent, 149
- Separating hyperplane, 375
 - geometry, 390
- Seven-league boots, 448
- Shrinkage, 115, 316, 338
 - estimator, 59, 91, 94, 96, 410
- Sigmoid function, 352
- Significance level, 274
- Simulation, 155–207
- Simultaneous confidence intervals, 395–399
- Simultaneous inference, 294, 418
- Sinc kernel, 440, 445
- Single-nucleotide polymorphism, *see* SNP
- Smoothing operator, 346
- SNP, 257
- Soft margin classifier, 378–379
- Soft-threshold, 315
- Softmax, 355
- Spam filter, 115
- Sparse
 - models, 298–323
 - principal components, 321
- Sparse matrix, 316
- Sparsity, 321
- Split-variable randomization, 327, 332
- SPSS, 271
- Squared error, 209
- Standard candles, 210, 231
- Standard error, 155
 - external, 408
 - internal, 408
- Standard interval, 181
- Stein's
 - paradox, 105
 - unbiased risk estimate, 218, 231
- Stepwise selection, 299
- Stochastic gradient descent, 358
- Stopping rule, 32, 413
- Stopping rules, 243
- String kernel, 385, 386
- Strong rules, 316, 322
- Structure, 261
- Structure matrix, 97, 424
- Student t
 - confidence interval, 396
 - distribution, 196, 272
 - statistic, 449
 - two-sample, 8, 272
- Studentized range, 418
- Subgradient
 - condition, 308
 - equation, 312, 315
- Subjective prior distribution, 233
- Subjective probability, 233
- Subjectivism, 35, 233, 243, 261
- Sufficiency, 44
- Sufficient
 - statistic, 66, 112, 116
 - vector, 66
- Supervised learning, 352
- Support
 - set, 377, 378
 - vector, 377
 - vector classifiers, 381
 - vector machine, 319, 375–393
- SURE, *see* Stein's unbiased risk estimate
- Survival analysis, 131–154, 266
- Survival curve, 137, 279
- SVM
 - Lagrange dual, 391
 - Lagrange primal, 391
 - loss function, 391
- Taylor series, 157, 420
- Theoretical null, 286
- Tied weights, 368
- Time series, xvi
- Training set, 208
- Transformation invariance, 183–185, 236
- Transient episodes, 228
- Trees
 - averaging, 348
 - best-first, 333
 - depth, 335
 - terminal node, 126
- Tricube kernel, 388, 389
- Trimmed mean, 175
- Triple-point, xv
- True error rate, 210
- True-discovery rates, 286
- Tukey, J. W., 418, 450
- Tukey, J. W., 418
- Tweedie's formula, 409, 419, 440
- Twenty-first-century methods, xvi, 271–446
- Two-groups model, 278
- Uncorrected differences, 411
- Uninformative prior, 28, 169, 233, 261
- Universal approximator, 351
- Unlabeled images, 365

- Unobserved covariates, 288
- Validation set, 213
- Vapnik, V., 390
- Variable-importance plot, 331–332, 336
- Variance, 14
- Variance reduction, 324
- Velocity vector, 360
- Voting, 333
- Warm starts, 314, 363
- Weak learner, 333, 342
- Weight
 - decay, 356
 - regularization, 361, 362
 - sharing, 352, 367
- Weighted exponential loss, 345
- Weighted least squares, 315
- Weighted majority vote, 341
- Weights, 352
- Wide data, 298, 321
- Wilks' likelihood ratio statistic, 246
- Winner's curse, 33, 408
- Winsorized mean, 175
- Working response, 315, 322
- $z^{(\alpha)}$, 188
- Zero set, 296