

Index

- A/B testing, 424
- accuracy, 215, 585
- activation function, 578
- activation map, 578
- adjust for confounders, 134, 278, 511
- alternative hypothesis, 391
- area under the ROC curve, 587, 596
- average treatment effect, 274
- averaging, 241

- backpropagation, 580
- bagging, 570
- Bayes' rule, 18
- Bayesian statistics, 225
- Bernoulli distribution, 47
- beta distribution, 228
- bias, 329
- bias (neural networks), 578
- bilinear model, 465
- binomial distribution, 47
- Bonferroni's correction, 419
- Boole's inequality, 338
- boosting, 573
- bootstrap, 370
- bootstrap aggregating, 570
- bootstrap Gaussian confidence interval, 375
- bootstrap percentile confidence interval, 377
- bootstrap standard error, 371
- Borel set, 68
- box plot, 75
- Brier score, 588

- calibration of a classifier, 588
- categorical distribution, 64
- Cauchy distribution, 347
- causal effect, 127
- causal inference, 127, 274, 315, 423, 510
- centering, 434
- central limit theorem, 349
- chain rule, 15
- chain rule for continuous random variables, 175
- chain rule for discrete and continuous random variables, 208
- chain rule for discrete random variables, 119
- Chebyshev's inequality, 337
- classification, 138, 212, 495
- classification tree, 565
- clustering, 217
- coefficient of determination, 304, 509
- collaborative filtering, 483
- completing the square, 356
- composite hypothesis, 392
- concordance, 587
- conditional continuous distribution, 174, 204
- conditional cumulative distribution function, 204
- conditional discrete distribution, 118, 207
- conditional independence of continuous random variables, 180
- conditional independence of discrete and continuous random variables, 211
- conditional independence of discrete random variables, 125
- conditional independence of events, 26
- conditional mean, 262
- conditional mean function, 263
- conditional probability, 13
- conditional probability density function, 175, 204
- conditional probability mass function, 118, 207
- conditional variance, 282
- confidence intervals, 363
- confounder, 130, 276, 318
- confounding factor, 130, 276, 318
- confusion matrix, 585
- conjugate prior, 228
- consistency, 340, 523
- continuous random variable, 66
- continuous variables, 66
- control for confounders, 134, 278, 511
- convergence in distribution, 52, 357
- convergence in mean square, 335
- convergence in probability, 340
- convergence of Markov chains, 148
- convolution, 350
- correlation, 284
- correlation and causation, 315
- correlation coefficient, 284, 288
- correlation matrix, 442
- counterfactual, 128, 316
- covariance, 288
- covariance matrix, 436
- cross-covariance, 498
- cross-entropy, 583

- cross-validation, 533
- c-statistic, 587
- cumulative distribution function, 69
- curse of dimensionality, 137, 274
- decile, 73
- decomposition of variance, 508
- deep learning, 577
- design matrix, 505
- deterministic variables, 37
- deviation bounds, 337
- dimensionality reduction, 457
- directions of maximum and minimum variance, 446, 451
- discrete and continuous variables, 202
- discrete random variable, 37
- discrete variables, 37
- discriminative model, 549
- domain shift, 215
- double centering, 493
- early stopping, 581, 594
- election, 231
- empirical cdf, 74
- empirical joint probability mass function, 113
- empirical probability, 19, 48
- empirical probability mass function, 43
- ensemble, 570
- entropy, 568
- estimator, 19
- event, 7
- exchangeability, 411
- expectation, 241
- expectation maximization, 218, 223
- explained variance, 300, 507
- exponential distribution, 90
- F1 score, 586, 597
- false negative, 399
- false positive, 399
- false positive rate, 585
- feature, 138, 212
- feature map, 578
- financial crisis, 359
- first moment, 241
- Fisher's transformation, 383
- fitting a model, 52
- frequentist statistics, 329, 369
- Frobenius inner product, 469
- Frobenius norm, 469
- function of a continuous random variable, 81
- function of a discrete random variable, 42
- Gaussian approximation, 357
- Gaussian discriminant analysis, 212
- Gaussian distribution, 95
- Gaussian mixture model, 205, 217
- Gaussian random vector, 186
- generalization, 58, 215, 514, 562
- generalized linear model, 548
- generative model, 549
- geometric distribution, 50
- Gini index, 569
- grand mean, 493
- hard-impute, 486
- held-out data, 58
- hidden variable, 578
- histogram, 44, 84
- hit rate, 585
- hypothesis testing, 390
- independence of continuous random variables, 177
- independence of discrete and continuous random variables, 211
- independence of discrete random variables, 122
- independence of events, 22
- independent identically distributed continuous random variables, 180
- independent identically distributed discrete random variables, 53, 125
- inner product between random variables, 310
- intercept, 496
- interquartile range, 73, 75
- inverse-transform sampling, 101
- iterated expectation, 262, 266
- joint cumulative distribution function, 163
- joint probability density function, 166
- joint probability mass function, 110
- kernel density estimation, 86
- label, 138, 212
- lasso, 543
- latent variable, 217, 578
- law of large numbers, 339
- law of total probability, 16
- learning rate, 580
- least squares, 256, 503
- likelihood, 53
- likelihood (Bayesian), 225
- likelihood (continuous), 97
- likelihood (discrete), 54
- linear causal effect, 316, 510
- linear coefficients, 496
- linear dimensionality reduction, 457
- linear discriminant analysis, 217
- linear minimum mean-squared-error estimation, 293, 496
- linear regression, 495
- linearity of expectation, 248, 436
- link function, 548
- logistic function, 548
- logistic regression, 546
- logit, 549, 553, 583
- log-likelihood (continuous), 97
- log-likelihood (discrete), 54

606

long regression, 511
 low-rank matrix completion, 483
 low-rank model, 464

machine learning, 570, 577
 marginal continuous distribution, 171
 marginal discrete distribution, 115
 marginal probability density function, 171
 marginal probability mass function, 115
 Markov chain, 141
 Markov property, 159
 Markov's inequality, 337
 matrix completion, 483
 maximum a posteriori, 138, 213
 maximum likelihood, 549, 553, 565, 583
 maximum likelihood (continuous), 96
 maximum likelihood (discrete), 52
 maximum likelihood (Gaussian random vector), 194
 mean, 241
 mean of a random vector, 433
 mean square, 255
 mean squared error, 255, 499
 measurable event, 39
 median, 73, 254
 memoryless property, 63, 92
 method of moments, 253
 minimum mean-squared-error estimation, 271
 mixture model, 205
 model evaluation, 57
 Monte Carlo method, 29
 multicollinearity, 519
 multidimensional density estimation, 170
 multidimensional kernel density estimation, 170
 multinoulli distribution, 64
 multiple continuous variables, 161
 multiple discrete variables, 109
 multiple testing, 417
 multivariate continuous distribution, 161
 multivariate discrete and continuous distributions, 202
 multivariate discrete distribution, 109

naïve Bayes, 138
 nearest neighbors, 462
 neural nets, 577
 neural networks, 577
 noise, 59
 noise amplification, 520
 nonlinear classification, 565, 582
 nonlinear regression, 557, 570, 577
 nonnegative matrix factorization, 494
 nonparametric model, 57, 59, 84, 170
 nonparametric testing, 408
 normal distribution, 95
 null hypothesis, 391

observational study, 128, 276
 one-tailed test, 396

Index

ordinary least squares, 503
 orthogonal projection, 313, 497, 525
 outlier, 76
 overfitting, 59, 489, 514, 555, 562, 581
 overparametrization, 581

parametric model, 46, 57, 90, 186, 225
 parametric testing, 392
 percentile, 73
 periodic Markov chain, 151
 permutation test, 408
 p -hacking, 427
 Poisson distribution, 51
 population parameter, 325
 positive predictive value, 585
 posterior distribution, 226
 potential outcome, 127
 power, 402
 power function, 402
 practical significance, 426
 precision, 585
 principal component, 447
 principal component analysis, 445
 principal direction, 446
 prior distribution, 225
 probability, 6, 7, 9
 probability density function, 76
 probability integral transform, 83
 probability mass function, 38
 probability measure, 10
 probability space, 13
 publication bias, 427
 p -value, 393, 409
 p -value function, 393

quadratic discriminant analysis, 215
 quantile, 72, 74
 quartile, 73

R^2 coefficient, 304, 509
 random forest, 571
 random sampling, 325
 random variable, 37, 66
 random vector (continuous), 162
 random vector (discrete), 109
 randomized controlled trial, 133
 randomized experiment, 132, 275, 318
 realization, 38
 recall, 585
 receiver operating characteristic curve, 596
 rectified linear unit, 578
 recursive binary splitting, 561
 regression, 271, 495
 regression tree, 557
 regularization, 531, 544, 556
 rejection of null hypothesis, 393, 399
 rejection region, 402
 reliability diagram, 588
 ridge regression, 531

- sample, 38
- sample conditional mean, 264
- sample correlation coefficient, 290
- sample covariance, 290
- sample covariance matrix, 441
- sample cross-covariance, 503
- sample mean, 247
- sample size, 334
- sample space, 6
- sample standard deviation, 259
- sample variance, 259
- second moment, 255
- selectivity, 585
- sensitivity, 585
- short regression, 511
- σ -algebra, 10
- signal, 59
- significance level, 399
- simple hypothesis, 392
- simple linear regression, 293, 313
- Simpson’s paradox, 130
- simulating continuous random variables, 101
- simulating multiple continuous random variables, 184
- singular-value decomposition, 467
- softmax, 553, 582
- softmax regression, 552
- sparse, 543
- sparse regression, 543
- sparsity, 543
- specificity, 585
- spectral theorem, 445, 454
- sphering, 493
- spurious correlation, 315
- St Petersburg paradox, 344
- standard deviation, 257
- standard error, 332
- standardized data, 291
- standardized variable, 287
- state diagram, 146
- state vector, 145
- stationary distribution, 148
- statistical significance, 399
- stochastic gradient descent, 580
- sum of Gaussian random variables, 354
- sum of independent random variables, 350
- test error, 528
- test set, 58
- test statistic, 392
- time series, 142
- time-homogenous Markov chain, 142
- topic modeling, 494
- training error, 523
- training set, 58
- transition matrix, 145
- treatment, 127, 315
- tree-based model, 556
- true negative rate, 585
- true positive rate, 585
- two-sample test, 395
- two-tailed test, 396
- type 1 error, 399
- type 2 error, 399
- unbiased estimator, 329
- uncertainty quantification, 225, 363
- uncorrelation, 285
- uncorrelation and independence, 306
- underfitting, 60, 489
- underpowered test, 404
- uniform distribution, 71, 81
- union bound, 338
- validation set, 489, 533
- variable selection, 543
- variance, 257
- vector space of matrices, 469
- vector space of random variables, 310
- Venn diagram, 11
- whitening, 493
- z test, 395