

Index

- accuracy, 547
- adjusted coefficient of determination, 739
- Apriori algorithm, 225
- artificial neural networks, *see* neural networks
- association rule, 222, 303
 - antecedent, 303
 - assessment measures, 303
 - Bonferroni correction, 322
 - confidence, 222, 305
 - consequent, 303
 - conviction, 308
 - Fisher exact test, 318
 - general, 317
 - improvement, 317
 - Jaccard coefficient, 307
 - leverage, 307
 - lift, 305
 - multiple hypothesis testing, 322
 - non-redundant, 317
 - odds ratio, 309
 - permutation test, 322
 - swap randomization, 323
 - productive, 317
 - randomization test, 322
 - redundant, 317
 - significance, 322
 - specific, 317
 - support, 222, 304
 - relative, 304
 - swap randomization, 323
 - unproductive, 317
 - bootstrap sampling, 327
 - confidence interval, 327
 - mining algorithm, 237
 - relative support, 222
- association rule mining, 237
- attribute
 - binary, 5
 - categorical, 5
 - nominal, 5
 - ordinal, 5
 - continuous, 5
 - discrete, 5
 - numeric, 5
 - interval-scaled, 5
 - ratio-scaled, 5
- backpropagation
 - CNN, 707
 - convolutional neural networks, 707
 - deep learning, 662
 - deep multilayer perceptron, 662
 - LSTM, 690
 - neural networks, 650
 - recurrent neural networks, 675
 - RNN, 675
- bagging, 574
- Bayes classifier, 469
 - categorical attributes, 473
 - numeric attributes, 470
- Bayes theorem, 469, 495
- Bernoulli distribution
 - mean, 62
 - sample mean, 62
 - sample variance, 62
 - variance, 62
- Bernoulli variable, 61
- BetaCV measure, 444
- bias-variance decomposition, 570
- bidirectional recurrent neural networks, 681
- binary database, 220
 - vertical representation, 220
- Binomial distribution, 63
- bivariate analysis

- categorical, 70
 - numeric, 40
- Bonferroni correction, 322
- boosting, 578
 - AdaBoost, 579
 - combined classifier, 581
- bootstrap
 - resampling, 561
 - sampling, 327
- bootstrap aggregation, *see* bagging
- C-index, 444
- Calinski–Harabasz index, 452
- categorical attributes
 - angle, 86
 - cosine similarity, 86
 - covariance matrix, 67, 81
 - distance, 86
 - Euclidean distance, 86
 - Hamming distance, 86
 - Jaccard coefficient, 86
 - mean, 65, 81
 - bivariate, 72
 - norm, 86
 - sample covariance matrix, 67
 - bivariate, 74
 - sample mean, 66
 - bivariate, 72
- Cauchy–Schwartz inequality, 9
- central limit theorem, 563
- centroid, 334
- Charm algorithm, 250
 - properties, 250
- χ^2 distribution, 78
- chi-squared statistic, 78
- χ^2 statistic, 78, 83
- classification, 467
 - accuracy, 547, 548, 552
 - area under ROC curve, 555
 - assessment measures, 546
 - contingency table based, 548
 - AUC, 555
 - bagging, 574
 - Bayes classifier, 469
 - bias, 571
 - bias-variance decomposition, 570
 - binary classes, 551
 - boosting, 578
 - AdaBoost, 579
 - classifier evaluation, 560
 - combined classifier, 574
 - confidence interval, 563
 - confusion matrix, 548
 - coverage, 549
 - cross-entropy error, 644
 - cross-validation, 560
 - decision trees, 483
 - ensemble classifiers, 574
 - error rate, 547, 551
 - out of bag, 577
 - expected loss, 570
 - F-measure, 549
 - false negative, 551
 - false negative rate, 552
 - false positive, 551
 - false positive rate, 552
 - K nearest neighbors classifier, 479
 - KNN classifier, 479
 - loss function, 570
 - naive Bayes classifier, 475
 - overfitting, 572
 - paired t -test, 568
 - precision, 548, 552
 - random forest, 575
 - recall, 549
 - sensitivity, 552
 - specificity, 552
 - stacking, 582
 - true negative, 551
 - true negative rate, 552
 - true positive, 551
 - true positive rate, 552
 - unstable, 574
 - variance, 571
- classifier evaluation, 560
 - bootstrap resampling, 561
 - confidence interval, 563
 - sample mean, 563
 - cross-validation, 560
 - paired t -test, 568
 - Student's t distribution, 566
- closed itemsets, 245
 - Charm algorithm, 250
 - equivalence class, 246
- cluster stability, 456
- clusterability, 459
- clustering, 332
 - centroid, 334
 - curse of dimensionality, 388
 - DBSCAN, 375
 - border point, 375
 - core point, 375
 - density connected, 376
 - density-based cluster, 376
 - directly density reachable, 375
 - ϵ -neighborhood, 375
 - noise point, 375
 - DENCLUE

- density attractor, 385
- dendrogram, 364
- density-based
 - DBSCAN, 375
 - DENCLUE, 385
- EM, *see* expectation maximization
- EM algorithm, *see* expectation maximization algorithm
- evaluation, 426
- expectation maximization, 343, 344
 - expectation step, 345, 349
 - Initialization, 349
 - maximization step, 346, 349
 - multivariate data, 347
 - univariate data, 345
- expectation maximization algorithm, 350
- external validation, 426
- Gaussian mixture model, 343
- graph cuts, 401
- internal validation, 426
- K-means, 335
 - specialization of EM, 354
- kernel density estimation, 379
- kernel K-means, 339
- Markov chain, 417
- Markov clustering, 417
- Markov matrix, 417
- relative validation, 426
- spectral clustering
 - computational complexity, 407
- stability, 426
- sum of squared errors, 334
- tendency, 426
- validation
 - external, 426
 - internal, 426
 - relative, 426
- clustering evaluation, 426
- clustering stability, 426
- clustering tendency, 426, 459
 - distance distribution, 460
 - Hopkins statistic, 462
 - spatial histogram, 459
- clustering validation
 - BetaCV measure, 444
 - C-index, 444
 - Calinski-Harabasz index, 452
 - clustering tendency, 459
 - conditional entropy, 431
 - contingency table, 427
 - correlation measures, 438
 - Davies–Bouldin index, 446
 - distance distribution, 460
 - Dunn index, 445
 - entropy-based measures, 431
 - external measures, 426
 - F-measure, 428
 - Fowlkes–Mallows measure, 437
 - gap statistic, 454
 - Hopkins statistic, 462
 - Hubert statistic, 439, 447
 - discretized, 439, 440
 - internal measures, 441
 - Jaccard coefficient, 436
 - matching based measures, 427
 - maximum matching, 428
 - modularity, 445
 - mutual information, 432
 - normalized, 432
 - normalized cut, 444
 - pair-wise measures, 435
 - purity, 427
 - Rand statistic, 437
 - relative measures, 450
 - silhouette coefficient, 446, 450
 - spatial histogram, 459
 - stability, 456
 - variation of information, 433
- CNN, *see* convolutional neural networks
- coefficient of determination, 723
- coefficient of multiple determination, 739
- conditional entropy, 431
- confidence interval, 327, 563, 728
 - small sample, 566
 - unknown variance, 565
- confusion matrix, 548
- contingency table, 77
 - χ^2 test, 83
 - clustering validation, 427
 - multiway, 83
- convolution, 694
 - 1D, 694
 - 2D, 695
 - 3D, 696
- convolutional neural networks, 694
 - 3D convolutions, 698
 - avg-pooling, 704
 - filter bias, 700
 - max-pooling, 704
 - padding, 702
 - parameter sharing, 694
 - pooling, 704
 - striding, 703
- correlation, 42
- cosine similarity, 9
- covariance, 41
- covariance matrix, 44, 46

- bivariate, 73
- determinant, 44
- eigen-decomposition, 55
- eigenvalues, 47
- inner product, 48
- outer product, 48
- positive semi-definite, 47
- trace, 44
- cross-entropy error, 626, 632
- cross-validation, 560
 - leave-one-out, 561
- cumulative distribution
 - binomial, 20
- cumulative distribution function, 20
 - empirical CDF, 29
 - empirical inverse CDF, 30
 - inverse CDF, 30
 - joint CDF, 24, 25
 - quantile function, 30
- curse of dimensionality
 - clustering, 388
- data dimensionality, 3
 - extrinsic, 15
 - intrinsic, 15
- data matrix, 3
 - centering, 11
 - column space, 14
 - mean, 11
 - rank, 15
 - row space, 14
 - symbolic, 61
 - total variance, 11
- data normalization
 - range normalization, 50
 - standard score normalization, 50
- Davies–Bouldin index, 446
- DBSCAN algorithm, 375
- decision tree algorithm, 487
- decision trees, 483
 - axis-parallel hyperplane, 485
 - categorical attributes, 487
 - data partition, 485
 - decision rules, 487
 - entropy, 489
 - Gini-index, 489
 - information gain, 489
 - purity, 486
 - split-point, 485
 - split-point evaluation, 490
 - categorical attributes, 494
 - measures, 488
 - numeric attributes, 490
- deep learning, 660
 - backpropagation phase, 662
 - bias gradient, 662
 - binary cross-entropy error, 664
 - CNN, 694
 - convolutional neural networks, 694
 - cross-entropy error, 665
 - dropout regularization, 716
 - exploding gradient, 668, 682
 - feed-forward phase, 662
 - gated RNN, 682
 - inverted dropout, 716
 - L_2 regularization, 713
 - long short-term memory networks, 682, 688
 - LSTM, 682, 688
 - minibatch learning, 668
 - net gradient, 663, 666
 - overfitting, 712
 - recurrent neural networks, 672
 - regularization, 712
 - RNN, 672
 - squared error, 664
 - training, 667
 - vanishing gradient, 668, 682
 - weight gradient, 662
- deep MLP, *see* deep multilayer perceptron
- deep multilayer perceptron, 660
 - backpropagation phase, 662
 - bias gradient, 662
 - binary cross-entropy error, 664
 - cross-entropy error, 665
 - exploding gradient, 668
 - feed-forward phase, 662
 - minibatch learning, 668
 - net gradient, 663, 666
 - squared error, 664
 - training, 667
 - vanishing gradient, 668
 - weight gradient, 662
- degrees of freedom, 39
- DENCLUE
 - center-defined cluster, 387
 - density attractor, 385, 386
 - density reachable, 387
 - density-based cluster, 387
- DENCLUE algorithm, 385
- dendrogram, 364
- density attractor, 386
- density estimation, 379
 - nearest neighbor based, 384
- density-based cluster, 387
- density-based clustering
 - DBSCAN, 375
 - DENCLUE, 385

INDEX

759

- dependent variable, 589
- dimensionality reduction, 184
- discrete random variable, 16
- discretization, 87
 - equal-frequency intervals, 88
 - equal-width intervals, 88
- dominant eigenvector, 104
 - power iteration method, 104
- dropout regularization, 716
- Dunn index, 445
- Eclat algorithm, 227
 - computational complexity, 230
 - dEclat, 231
 - diffsets, 230
 - equivalence class, 228
- empirical joint probability mass function, 459
- ensemble classifiers, 574
 - bagging, 574
 - boosting, 578
 - random forest, 575
 - stacking, 582
- entropy, 489
 - split, 489
- EPMF, *see* empirical joint probability mass function
- error rate, 547
 - out of bag, 577
- error term
 - linear regression, 589
- Euclidean distance, 9
- expectation maximization, 343, 344, 358
 - expectation step, 359
 - maximization step, 359
- expected value, 30
- explanatory variable, *see* independent variable
- exploding gradient, 668, 682
- F-measure, 428
- false negative, 551
- false positive, 551
- Fisher exact test, 318, 320
- Fowlkes–Mallows measure, 437
- FPGrowth algorithm, 233
- frequent itemset, 221
- frequent itemset mining, 223
- frequent pattern mining, 217
- Gamma function, 78, 167
- gap statistic, 454
- gate neurons, 683
- gated RNN, 682
- Gauss error function, 54
- Gaussian mixture model, 343
- generalized itemset, 253
- GenMax algorithm, 247
 - maximality checks, 248
- Gini index, 489
- graph, 282
 - adjacency matrix, 95
 - weighted, 95
 - authority score, 109
 - average degree, 97
 - average path length, 97
 - Barabási–Albert model, 123
 - clustering coefficient, 130
 - diameter, 130
 - Barabási–Albert model
 - degree distribution, 125
 - centrality
 - authority score, 109
 - betweenness, 102
 - closeness, 102
 - degree, 101
 - eccentricity, 101
 - eigenvector centrality, 103
 - hub score, 109
 - pagerank, 106
 - prestige, 103
 - clustering coefficient, 99
 - clustering effect, 114
 - degree, 97
 - degree distribution, 93
 - degree sequence, 93
 - diameter, 98
 - eccentricity, 97
 - effective diameter, 98
 - efficiency, 100
 - Erdős–Rényi model, 114
 - HITS, 109
 - hub score, 109
 - labeled, 282
 - pagerank, 106
 - preferential attachment, 123
 - radius, 98
 - random graphs, 114
 - scale-free property, 112
 - shortest path, 94
 - small-world property, 111
 - transitivity, 100
 - Watts–Strogatz model, 118
 - clustering coefficient, 118
 - diameter, 118, 122
 - Watts–Strogatz model
 - degree distribution, 120
- graph clustering

- average weight, 410
- degree matrix, 395
- graph cut, 402
- k -way cut, 401
- Laplacian matrix, 398
- Markov chain, 417
- Markov clustering, 417
- MCL algorithm, 419
- modularity, 411
- normalized adjacency matrix, 395
- normalized asymmetric Laplacian, 400
- normalized cut, 404
- normalized modularity, 416
- normalized symmetric Laplacian, 399
- objective functions, 403, 410
- ratio cut, 403
- weighted adjacency matrix, 394
- graph cut, 402
- graph isomorphism, 283
- graph kernel, 156
 - exponential, 158
 - power kernel, 157
 - von Neumann, 159
- graph mining
 - canonical DFS code, 289
 - canonical graph, 288
 - canonical representative, 287
 - DFS code, 288
 - edge growth, 285
 - extended edge, 282
 - graph isomorphism, 283
 - gSpan algorithm, 290
 - rightmost path extension, 286
 - rightmost vertex, 287
 - search space, 285
 - subgraph isomorphism, 284
- graph models, 111
 - Barabási–Albert model, 123
 - Erdős–Rényi model, 114
 - Watts–Strogatz model, 118
- graphs
 - degree matrix, 395
 - Laplacian matrix, 398
 - normalized adjacency matrix, 395
 - normalized asymmetric Laplacian, 400
 - normalized symmetric Laplacian, 399
 - weighted adjacency matrix, 394
- GSP algorithm, 263
- gSpan algorithm, 290
 - candidate extension, 293
 - canonicity checking, 297
 - subgraph isomorphisms, 295
 - support computation, 293
- hierarchical clustering, 364
 - agglomerative, 364
 - complete link, 368
 - dendrogram, 364, 365
 - distance measures, 367
 - divisive, 364
 - group average, 368
 - Lance–Williams formula, 370
 - mean distance, 368
 - minimum variance, 368
 - single link, 367
 - update distance matrix, 370
 - Ward’s method, 368
- Hopkins statistic, 462
- Hubert statistic, 439, 447
- hyper-rectangle, 163
- hyperball, 164
 - volume, 167
- hyperbolic tangent, *see* tanh function
- hypercube, 164
 - volume, 167
- hyperplane, 165
 - bias, 165
 - weight vector, 165
- hyperspace, 163
 - density of multivariate normal, 173
 - diagonals, 172
 - angle, 172
- hypersphere, 164
 - asymptotic volume, 169
 - closed, 164
 - inscribed within hypercube, 170
 - surface area, 169
 - volume of thin shell, 171
- hypersphere volume, 177
 - Jacobian, 178–180
 - Jacobian matrix, 177, 179, 180
- hypothesis testing, 744
 - regression, 730
- IID, *see* independent and identically distributed
- inclusion-exclusion principle, 254
- independent and identically distributed, 26
- independent variable, 589
- information gain, 489
- inter-quartile range, 35
- itemset, 219
- itemset mining, 219, 223
 - Apriori algorithm, 225
 - level-wise approach, 225
 - candidate generation, 223
 - Charm algorithm, 250
 - computational complexity, 224

- Eclat algorithm, 227
 - tidset intersection, 227
- FPGrowth algorithm, 233
 - frequent pattern tree, 233
- frequent pattern tree, 233
- GenMax algorithm, 247
- level-wise approach, 225
- negative border, 242
- partition algorithm, 240
- prefix search tree, 223, 226
- support computation, 223
- tidset intersection, 227
- itemsets
 - assessment measures, 311
 - closed, 315
 - maximal, 314
 - minimal generator, 315
 - minimum support threshold, 221
 - productive, 316
 - support, 311
 - relative, 311
 - closed, 245, 250
 - closure operator, 245
 - properties, 245
 - generalized, 253
 - maximal, 244, 247
 - minimal generators, 246
 - non-derivable, 252, 256
 - relative support, 221
 - rule-based assessment measures, 312
 - support, 221
- Jaccard coefficient, 436
- Jacobian matrix, 177, 179, 180
- K nearest neighbors classifier, 479
- K-means
 - algorithm, 335
 - kernel method, 339
- k -way cut, 401
- kernel density estimation, 379
 - discrete kernel, 380, 383
 - Gaussian kernel, 381, 383
 - multivariate, 382
 - univariate, 379
- kernel discriminant analysis, 508
- kernel K-means, 339
- kernel matrix, 135
 - centered, 151
 - normalized, 153
- kernel methods
 - data-specific kernel map, 142
 - diffusion kernel, 156
 - exponential, 158
 - power kernel, 157
 - von Neumann, 159
- empirical kernel map, 140
- Gaussian kernel, 147
- graph kernel, 156
- Hilbert space, 140
- kernel matrix, 135
- kernel operations
 - centering, 151
 - distance, 149
 - mean, 150
 - norm, 149
 - normalization, 153
 - total variance, 150
- kernel trick, 137
- Mercer kernel map, 143
- polynomial kernel
 - homogeneous, 144
 - inhomogeneous, 144
- positive semi-definite kernel, 138
- pre-Hilbert space, 140
- reproducing kernel Hilbert space, 140
- reproducing kernel map, 139
- reproducing property, 140
- spectrum kernel, 155
- string kernel, 155
- vector kernel, 144
- kernel PCA, *see* kernel principal component analysis
- kernel principal component analysis, 203
- kernel regression, 611
 - kernel hat matrix, 612
- kernel ridge regression, 611
- kernel trick, 339
- KL divergence, *see* Kullback–Leibler divergence
- KNN classifier, 479
- Kullback–Leibler divergence, 459
- L_1 regularization, 615
 - linear regression, 615
- L_1 regularized regression, *see* Lasso
- L_2 regularization, 606, 713
 - deep learning, 713
 - linear regression, 606
- L_2 regularized regression, *see* ridge regression
- Lasso, 615
 - cyclical coordinate descent, 618
 - soft-threshold function, 618
 - subdifferential, 616
 - subgradient, 616
- learning rate, 633

- least absolute selection and shrinkage operator, *see* Lasso
- linear discriminant analysis, 501
 - between-class scatter matrix, 504
 - Fisher objective, 504
 - optimal linear discriminant, 504
 - within-class scatter matrix, 505
- linear function
 - derivative, 641
- linear regression
 - bias, 589
 - dependent variable, 589
 - error term, 589
 - independent variable, 589
 - Lasso, 615
 - multiple regression, 596
 - regression coefficient, 589
 - residual error, 590
 - response variable, 589
 - ridge regression, 606
 - squared error, 590
 - SSE, 590
- log-likelihood, 626
 - cross-entropy error, 626, 632
- log-odds ratio, 625
- logistic function, 624
- logistic regression, 623
 - cross-entropy error, 626, 632
 - learning rate, 633
 - log-likelihood, 626
 - log-odds ratio, 625
 - logistic function, 624
 - logit function, 625
 - maximum likelihood estimation, 626
 - multiclass, 630
 - one-hot encoding, 630
 - sigmoid function, 624
 - softmax function, 631
 - stochastic gradient ascent, 628, 633
- logit, *see* log-odds ratio
- long short-term memory networks, 682, 688
 - forget gate, 688
 - gate neurons, 683
 - input gate, 688
 - internal memory, 688
 - output gate, 688
- loss function, 570
 - squared loss, 570
 - zero-one loss, 570
- LSTM, *see* long short-term memory networks
- Mahalanobis distance, 54
- Markov chain, 417
- Markov clustering, 417
- maximal itemsets, 244
 - GenMax algorithm, 247
- maximum likelihood estimation, 344, 354, 626, 632
 - covariance matrix, 356
 - cross-entropy error, 626, 632
 - gradient ascent, 627
 - log-likelihood function, 632
 - mean, 355
 - mixture parameters, 357
- maximum matching, 428
- mean, 30
- median, 33
- minibatch learning, 668, 679
- minimal generator, 246
- MLE, *see* maximum likelihood estimation
- MLP
 - see* multilayer perceptron, 648
- mode, 33
- modularity, 412, 445
- multiclass logistic regression, 630
- multilayer perceptron, 648
 - backpropagation phase, 650
 - bias gradient, 651
 - epoch, 656
 - feed-forward phase, 648
 - gradient descent, 651
 - net gradient, 652
 - stochastic gradient descent, 655, 667
 - training, 655
 - weight gradient, 651
 - weight update, 651
- multinomial distribution, 70
 - covariance, 70
 - mean, 70
 - sample covariance, 70
 - sample mean, 70
- multiple hypothesis testing, 322
- multiple regression, 596, 735
 - F -test, 744
 - f -statistic, 744
 - estimated variance, 737
 - goodness of fit, 738
 - gradient descent, 605
 - Gram–Schmidt orthogonalization, 600
 - hat matrix, 597
 - hypothesis testing, 744
 - normal equations, 599
 - QR-factorization, 600
 - SGD, 605
 - stochastic gradient descent, 605
- multivariate analysis
 - categorical, 81

INDEX

763

- numeric, 46
- multivariate Bernoulli variable, 64, 81
 - covariance matrix, 67, 81
 - empirical PMF, 68
 - joint PMF, 71
 - mean, 65, 81
 - probability mass function, 65, 71
 - sample covariance matrix, 67
 - sample mean, 66
- multivariate variable
 - Bernoulli, 64
 - one-hot encoding, 64
- mutual information, 432
 - normalized, 432
- naive Bayes classifier, 475
 - categorical attributes, 478
 - numeric attributes, 475
- nearest neighbor density estimation, 384
- neural networks, 637
 - activation function, 637
 - activation functions, 637
 - backpropagation phase, 650
 - bias gradient, 651
 - bias neuron, 637
 - binary cross-entropy error, 647, 664
 - classification, 644
 - cross-entropy error, 647, 665
 - epoch, 656
 - error functions, 646
 - feed-forward phase, 648
 - gradient descent, 651
 - identity function, 638
 - linear function, 638
 - logistic regression, 644
 - multilayer perceptron, 648
 - net gradient, 652
 - regression, 642
 - ReLU function, 640
 - sigmoid function, 640
 - softmax function, 640
 - squared error, 646, 664
 - step function, 638
 - stochastic gradient descent, 655, 667
 - tanh function, 640
 - training, 655, 667
 - weight gradient, 651
 - weight update, 651
- non-derivable itemsets, 252, 256
 - inclusion-exclusion principle, 254
 - support bounds, 254
- normal distribution
 - Gauss error function, 54
- normalized cut, 444
- one-hot encoding, 64, 630
- orthogonal complement, 187
- orthogonal projection matrix, 187
 - error vector, 187
- orthogonal subspaces, 187
- padding, 702
- pagerank, 106
- paired *t*-test, 568
- parameter sharing, 694
 - recurrent neural networks, 673
- pattern assessment, 311
- PCA, *see* principal component analysis
- permutation test, 322
 - swap randomization, 323
- pooling, 704
 - avg-pooling, 704
 - max-pooling, 704
- population, 26
- power iteration method, 104
- predictor variable, *see* independent variable
- PrefixSpan algorithm, 267
- principal component, 188
 - kernel PCA, 203
- principal component analysis, 188
 - choosing the dimensionality, 198
 - connection with SVD, 213
 - eigen-decomposition, 201
 - mean squared error, 194, 197
 - minimum squared error, 190
 - total projected variance, 193, 197
- probability distribution
 - bivariate normal, 23
 - normal, 19
- probability density function, 18
 - joint PDF, 22, 25
- probability distribution
 - Bernoulli, 17, 61
 - binomial, 17
 - Gaussian, 19
 - multivariate normal, 54
 - normal, 52
- probability mass function, 16
 - empirical joint PMF, 40
 - empirical PMF, 30
 - joint PMF, 22, 25
- projection
 - error vector, 12
 - orthogonal projection, 12
 - scalar projection, 13
- purity, 427
- quantile function, 30

- quartile, 35
- R^2 statistic, *see* coefficient of determination
- Rand statistic, 437
- random forest, 575
- random graphs, 114
 - average degree, 115
 - clustering coefficient, 116
 - degree distribution, 116
 - diameter, 117
- random sample, 26
 - multivariate, 26
 - statistic, 27
 - univariate, 26
- random variable, 16
 - Bernoulli, 61
 - bivariate, 21
 - continuous, 16
 - correlation, 42
 - covariance, 41
 - covariance matrix, 44, 46
 - discrete, 16
 - empirical joint PMF, 40
 - expectation, 30
 - expected value, 30
 - generalized variance, 44, 47
 - independent and identically distributed, 26
 - inter-quartile range, 35
 - mean, 30
 - bivariate, 41
 - multivariate, 46
 - median, 33
 - mode, 33
 - moments about the mean, 36
 - multivariate, 25
 - standard deviation, 36
 - standardized covariance, 42
 - total variance, 41, 44, 47
 - value range, 35
 - variance, 36
 - vector, 25
- receiver operating characteristic curve, 554
 - class imbalance, 560
- rectified linear unit, *see* ReLU function
- recurrent neural networks, 672
 - backpropagation phase, 675
 - bidirectional, 681
 - feed-forward in time, 675
 - feed-forward phase, 675
 - forget gate, 684
 - minibatch learning, 679
 - net gradients, 677
 - parameter sharing, 673
 - stochastic gradient descent, 677
 - weight tying, 673
- regression, 587
 - f -statistic, 731
 - adjusted coefficient of determination, 739
 - adjusted R^2 statistic, 739
 - coefficient of determination, 723
 - coefficient of multiple determination, 739
 - confidence interval, 728
 - deterministic component, 720
 - estimated variance, 722
 - hypothesis testing, 730
 - kernel regression, 611
 - kernel trick, 611
 - linear regression, 589
 - multiple regression, 596
 - overfitting, 712
 - probabilistic model, 720
 - R^2 statistic, 723, 739
 - random error component, 720
 - regression sum of squares, 723
 - ridge regression, 606
 - standard error of regression, 722
 - standardized residuals, 733
 - total scatter, 722
 - total sum of squares, 722
- regression effect, 730
- regression sum of squares, 723
- regularization, 712
 - inverted dropout, 716
- ReLU function, 640
 - derivative, 641
- response variable, *see* dependent variable
- ridge regression, 606
 - SGD, 610
 - stochastic gradient descent, 610
- RNN, *see* recurrent neural networks
- ROC curve, *see* receiver operating characteristic curve
- RSS, *see* regression sum of squares
- rule assessment, 303
- sample covariance matrix
 - bivariate, 74
- sample mean, 27
- sample space, 16
- sample variance
 - geometric interpretation, 39
- scatter matrix, 49
- sequence, 261
 - closed, 262
 - maximal, 262

INDEX

765

- sequence mining
 - alphabet, 261
 - GSP algorithm, 263
 - prefix, 261
 - PrefixSpan algorithm, 267
 - relative support, 262
 - search space, 263
 - sequence, 261
 - Spade algorithm, 265
 - subsequence, 261
 - consecutive, 261
 - substring, 261
 - substring mining, 269
 - suffix, 261
 - suffix tree, 269
 - support, 262
 - SGA, *see* stochastic gradient ascent
 - sigmoid function, *see* logistic function, 640
 - derivative, 641
 - silhouette coefficient, 446, 450
 - singular value decomposition, 210
 - connection with PCA, 213
 - Frobenius norm, 212
 - left singular vector, 211
 - reduced SVD, 211
 - right singular vector, 211
 - singular value, 211
 - spectral decomposition, 211
 - softmax function, 631, 640
 - derivative, 641
 - Spade algorithm
 - sequential joins, 265
 - spectral clustering
 - average weight, 410
 - computational complexity, 407
 - degree matrix, 395
 - k -way cut, 401
 - Laplacian matrix, 398
 - modularity, 411
 - normalized adjacency matrix, 395
 - normalized asymmetric Laplacian, 400
 - normalized cut, 404
 - normalized modularity, 416
 - normalized symmetric Laplacian, 399
 - objective functions, 403, 410
 - ratio cut, 403
 - weighted adjacency matrix, 394
 - spectral clustering algorithm, 406
 - SSE
 - linear regression, 590
 - stacking, 582
 - standard deviation, 36
 - standard score, 36
 - statistic, 27
 - degrees of freedom, 39
 - robustness, 31
 - sample correlation, 43
 - sample covariance, 42
 - sample covariance matrix, 44, 48
 - sample inter-quartile range, 35
 - sample mean, 27, 31
 - bivariate, 41
 - multivariate, 46
 - sample median, 33
 - sample mode, 33
 - sample range, 35
 - sample standard deviation, 36
 - sample total variance, 41
 - sample variance, 36
 - standard score, 36
 - trimmed mean, 31
 - unbiased estimator, 31
 - z -score, 36
- statistical independence, 24
 - step function, 638
 - derivative, 641
 - Stirling numbers
 - second kind, 334
 - stochastic gradient ascent, 628, 633
 - stochastic gradient descent, 655, 667, 677
 - striding, 703
 - string, *see* sequence
 - string kernel
 - spectrum kernel, 155
 - subgraph, 283
 - connected, 283
 - support, 285
 - subgraph isomorphism, 284
 - substring mining, 269
 - suffix tree, 269
 - Ukkonen's algorithm, 272
 - suffix tree, 269
 - Ukkonen's algorithm, 272
 - sum of squared residual errors, *see* SSE
 - support vector machines, 517
 - bias, 517
 - canonical hyperplane, 521
 - classifier, 525
 - directed distance, 518
 - dual algorithm, 537
 - dual objective, 524
 - hinge loss, 528, 535
 - hyperplane, 517
 - Karush–Kuhn–Tucker conditions, 523
 - kernel SVM, 533
 - linearly separable, 518
 - margin, 520
 - maximum margin hyperplane, 523

- non-separable case, 527
- nonlinear case, 533
- primal objective, 523
- quadratic loss, 532, 535
- regularization constant, 528
- separable case, 523
- separating hyperplane, 518
- slack variables, 527
- soft margin, 527
- stochastic gradient ascent, 537
- support vectors, 520
- weight vector, 517
- SVD, *see* singular value decomposition
- SVM, *see* support vector machines
- swap randomization, 323

- tanh function, 640
 - derivative, 641
- tidset, 220
 - transaction identifiers, 220
 - tids, 220
- total sum of squares, 722
- total variance, 11, 41
- transaction, 220
- transaction database, 220
- true negative, 551
- true positive, 551
- TSS, *see* total sum of squares

- Ukkonen's algorithm
 - computational cost, 273
 - implicit extensions, 274
 - implicit suffixes, 273
 - skip/count trick, 274
 - space requirement, 272
 - suffix links, 275
 - time complexity, 278
- univariate analysis
 - categorical, 61
 - numeric, 29

- vanishing gradient, 668, 682
- variance, 36
- variation of information, 433
- vector
 - dot product, 8
 - Euclidean norm, 8
 - length, 8
 - linear combination, 6
 - L_2 -norm, 8
 - L_p -norm, 8
 - normalization, 8
 - orthogonal decomposition, 12
 - orthogonal projection, 12
 - orthogonality, 9
 - perpendicular distance, 12
 - scalar projection, 13
 - standard basis, 6
 - unit vector, 8
- vector kernel, 144
 - Gaussian, 147
 - polynomial, 144
- vector random variable, 25
- vector space
 - basis, 15
 - column space, 14
 - dimension, 15
 - linear combination, 14
 - linear dependence, 14
 - linear independence, 15
 - orthogonal basis, 15
 - orthonormal basis, 15
 - row space, 14
 - span, 14
 - spanning set, 14
 - standard basis, 15

- Watts–Strogatz model
 - clustering coefficient, 121
- weight tying, *see* parameter sharing

- z-score, 36