

Index

- a* dimension, 77, 80, 86, 87, 94
 χ^2 -divergence, 389, 499, 524
 f -divergence, 524
- data compression, 72
- adaptive composition, 326
- ADX, 46
- ADX
- achievability, 57
 - converse, 56
- Akaike information criterion, 366
- algorithmic stability, 304, 308
- aliasing, 65
- almost exact recovery, 386, 394–401
 - necessary conditions, 398–401
 - sufficient conditions, 396–398
- analog-to-digital
 - compression, 46
 - conversion, 46
- approximate message passing (AMP), 74, 200, 202, 206–207
- approximation error, 338, 365
- arithmetic coding, 266
- artificial intelligence, 264
- Assouad’s method, 513
- asymptotic efficiency, 363
- asymptotic equipartition property, 6
- barrier-based second-order methods, 110
- basis pursuit, 74
- Bayes decision rule, 333
- Bayes risk, 337
- Bayesian information criterion, 366
- Bayesian network, 281
- belief propagation, 200
- bias, 365
- bilinear problem, 223
- binary symmetric community detection problem, 386–393
- binary symmetric community model, 385
- binary symmetric community recovery problem, 386–393
- bitrate, 45
 - unlimited, 57
 - unrestricted, 53
- Blahut–Arimoto algorithm, 283, 346
- boosting, 327
- Bradley–Terry model, 231
- Bregman divergence, 277
- Bregman information, 277
- Bridge criterion, 371
- capacity, 267, 325, 346
 - channel, 346
- cavity method, 393, 414
- channel, 3, 325
 - binary symmetric, 8
 - capacity, 10, 11
 - discrete memoryless, 10
 - Gaussian, 200, 203, 210
- channel coding, 339
 - achievability, 9
 - blocklength, 10
 - capacity-achieving code, 15
 - codebook, 10, 13
 - codewords, 3, 10
 - converse, 11
 - decoder, 3
 - encoder, 3
 - error-correction codes, 10
 - linear codes, 12
 - generator matrix, 13
 - parity-check matrix, 13
 - maximum-likelihood decoding, 10
 - polar codes, 14
 - sphere packing, 10
 - syndrome decoding, 12
- characteristic (conflict) graph, 459
- characteristic graph, 459

- chromatic entropy, 462
 circular Nyquist density, 168, 169
 classification, 303, 318, 333, 376–379
 clustering, 264, 271, 303
 clustering memoryless sources, 276
 clustering sources with memory, 279
 correct, 272
 dependence-based clustering, 280
 distance-based clustering, 276
 Gaussian mixture clustering, 415
 hard clustering algorithm, 277
 hard clustering methods, 272
 hierarchical clustering algorithm, 284
 independence clustering, 281
 information-based clustering, 284
 k -means clustering, 266, 276
 soft clustering algorithm, 277
 soft clustering methods, 272
 coherence, 164, 170, 174
 mutual, 177
 coloring connectivity condition, 457, 464
 communication, 267–268
 community detection, 31
 community detection problem, 223
 compressed sensing, 163, 237
 compression-based, 75
 compressible phase retrieval (COPER), 93
 compressible signal pursuit, 77
 compression, 266–267
 compression codes, 75
 compression-based gradient descent, 81
 compression-based (projected) gradient
 descent, 91
 computational limits, 401–412
 concentration inequality, 171, 175, 179–181
 conditional T -information, 306
 conditional chromatic entropy, 462
 conditional covariance matrix, 208
 conditional graph entropy, 466
 conditional mutual information, 282, 306
 conjugate gradient method, 110
 consistency, 361
 constrained least-squares problem, 110,
 113, 119
 constraint
 bitrate, 45
 contiguity, 389
 convex analysis, 105, 108
 convex function, 109
 convex optimization, 105, 108, 109
 convex set, 108
 correlated recovery, 385–388, 390–393
 lower bound, 390–393
 upper bound, 387–388
 CoSaMP, 74
 covariance matrix, 208
 covering, 499, 514
 number, 514
 covering number, 185, 317
 cross-entropy, 334
 cross-entropy loss, 335, 341–342
 cross-entropy risk, 335
 cross-validation, 367
 v -fold, 367
 leave-one-out, 367
 multi-fold, 90
 paradox, 377
 Dantzig selector, 87
 data
 empirical data distribution, 336
 labeled data, 264
 training data, 264
 true-data generating distribution, 303, 337, 360
 unlabeled data, 263
 data acquisition, 18, 72
 analog signal, 18
 digital signal, 18
 quantization, 18
 rate distortion, 18
 scalar, 18
 vector, 18
 sampling, 18
 Nyquist rate, 18
 sampling frequency, 18
 sampling period, 18
 shift-invariant, 18
 uniform, 18
 sampling theorem, 18
 task-ignorant, 20
 task-oriented, 20
 data analysis, 26
 statistical inference, *see* statistical inference
 statistical learning, *see* statistical learning
 data compression, 3, 76, 456
 lossless data compression, 76
 lossy data compression, 76, 95
 data-processing inequality, 492, 496
 data representation, 20, 134–135, 137
 analysis operator, 21
 data-driven, 20, 135
 linear, 21, 22
 nonlinear, 21, 23
 supervised, 21
 unsupervised, 21
 deep learning, *see* neural networks
 dictionary learning, *see* dictionary learning
 dimensionality reduction, 20
 features, 20
 independent component analysis, *see* independent
 component analysis
 kernel trick, 24
 linear, 135
 manifold learning, 23

- matrix factorization, 25
 model based, 20, 135
 nonlinear, 136
 principal component analysis, *see* principal component analysis
 sparse, *see* dictionary learning
 subspace, 22
 synthesis operator, 21
 union of subspaces, 24
 data science, 15
 fairness, 35
 phase transitions, 17
 pipeline, 16
 privacy, 35
 training label, 21
 training sample, 21
 training samples, 21
 De Bruijn identity, 209
 decoder, 52, 335, 339
 empirical decoder, 336
 maximum mutual information, 267
 minimum partition information, 268
 optimal decoder, 335
 randomized decoder, 334
 soft decoder, 335
 deep convolutional network, 264
 deep learning, *see* neural networks, 264, 326
 deep neural network, 331
 multilayer feed-forward network, 331
 deep recurrent network, 264
 deep representation, 335
 denoising, 269
 density estimation, 303, 519
 dependence-based clustering, 280
 detection, 201, 386–390, 401–412
 lower bound, 389–390
 upper bound, 387–388
 dictionary, 177
 incomplete, 177
 redundant, 177
 dictionary learning, 24, 134–159
 algorithms, 159
 coherence measures, 142, 147, 155
 complete, 136, 140
 distance metric, 138–140, 153
 Kronecker structured, 152
 mean-squared error, 138
 minimax lower bound, 24
 overcomplete, 136, 140
 sample complexity, 138
 supervised, 24
 tensor data, 149–157
 achievability results, 155–157
 coefficient tensor, 152
 coordinate dictionary, 152
 empirical risk minimization, 153
 model, 151
 random sparsity, 152
 separable sparsity, 152
 statistical risk minimization, 153
 unsupervised, 24
 vector data, 137–149
 achievability results, 142–148
 empirical risk minimization, 139
 identifiability, 143–144
 K-SVD, 139, 145
 minimax risk, *see* minimax risk
 model, 138
 noisy reconstruction, 144–147
 outliers, 147–148
 sparsity, 141, 142
 statistical risk minimization, 139
 differential entropy, 209
 differential privacy, 125, 323
 digital
 representation, 46
 distance
 Lempel–Ziv distance, 274
 Bhattacharya, 279
 edit distance, 274
 Euclidean distance, 274
 Hamming distance, 274
 information distance, 275
 Itakura–Saito, 276
 Mahalanobis distance, 277
 normalized compression distance, 280
 normalized information distance, 275
 distance-based clustering, 276
 distortion function, 339
 distortion measure, 339
 distortion rate, 77
 analog-to-digital, 52
 function, 77, 95
 analog, 55
 indirect, 55
 indirect/remote, 48
 remote, 55
 sampling, 52
 vector Gaussian, 48
 distributed detection, 427
 distributed estimation, 427
 distributed hypothesis testing, 427
 cascaded communication, 427,
 438–443
 interactive communication, 427, 443–445
 non-interactive communication, 427–437
 distributed identity testing, 427, 445–450
 distributed inference, 35
 distributed pattern classification, 427
 distributed source coding, 430
 distributed statistical inference, 425
 feature partition, 426
 sample partition, 425
 distribution over permutations, 232

- divergence
 - Bregman divergence, 274, 277
 - f -divergence, 274, 278
 - I-divergence, 277
 - Kullback–Leibler divergence, 11, 277
- empirical loss, 303, 361
- empirical risk, 303, 336
- empirical risk minimization, 139, 153, 304, 337
 - noisy, 324
- encoder, 51, 335, 339
 - randomized, 335
 - capacity, 351
- entropy, 5, 6
 - conditional entropy, 7
- entropy rate, 96
- erasure T -information, 306
- erasure mutual information, 306
- error
 - absolute generalization error, 313
 - approximation error, 338, 365
 - Bayes error, 333
 - estimation error, 338
 - expected generalization error, 304
 - generalization error, 303
 - prediction error, 364, 366, 368, 377, 379
 - testing error, 331
 - training error, 331
 - error probability, 493
 - approximate recovery, 494
 - estimation, 105, 201, 265
 - Bayesian, 489
 - continuous, 510
 - minimax, 489, 510
 - estimation bias, 365
 - estimation error, 338
 - estimation phase, 378
 - estimation variance, 365
 - exact recovery, 386, 394–401
 - necessary conditions, 398–401
 - sufficient conditions, 396–398
 - expectation propagation, 200
 - exponential family of distributions, 232
 - false-positive rate, 201
 - Fano’s inequality, 7, 11, 116, 134, 137, 141, 491
 - applications, 488, 499, 516
 - approximate recovery, 494
 - conditional version, 495
 - continuum version, 524
 - generalizations, 524
 - genie argument, 503
 - limitations, 523
 - standard version, 493
 - Fano’s lemma, 352
 - fast iterative shrinkage-thresholding algorithm, 74
 - filter
 - pre-sampling, 63
 - Wiener, 54, 64
 - first-moment method, 387–388
 - Fisher information, 209
 - frame, 136, 138, 148
 - tight, 144, 145
 - frequency $H(f)$, 51
 - frequency response, 51
 - functional compression
 - distributed functional compression with distortion, 474
 - feedback in functional compression, 473
 - functional compression with distortion, 456
 - lossless functional compression, 456
 - network functional compression, 456
 - fundamental limit, 201, 265, 330
 - Gauss–Markov process, 62
 - Gaussian process, 50
 - vector, 48
 - Gaussian complexity, 111
 - localized Gaussian complexity, 105, 111
 - Gaussian mixture clustering, 415
 - generalization error, 303
 - generalization gap, 337
 - bounds, 348–349
 - information bounds, 348
 - generalized information criterion, 368
 - generalized likelihood ratio, 268
 - generalized likelihood ratio test, 268
 - generalized linear model, 203
 - Gibbs algorithm, 320
 - good codes, 210
 - graph
 - forest, 504
 - tree, 504
 - weighted undirected graph, 108
 - graph coloring, 461
 - graph entropy, 459
 - graph sparsification via sub-sampling, 108
 - graphical model, 281
 - Gaussian, 281
 - graphical model selection, 503
 - adaptive sampling, 507
 - approximate recovery, 506
 - Bayesian, 509
 - graphons, 413
 - group testing, 500
 - adaptive, 502
 - approximate recovery, 501
 - non-adaptive, 501
 - Guerra interpolation method, 384, 393, 414

- Hamming distance, 9
 Hamming weight, 13
 hash function, 108
 Hellinger distance, 499, 524
 Hoeffding's lemma, 307
 Huffman coding, 266
 hyperparameter, 345
 hypothesis, 303, 333
 hypothesis class, 304
 hypothesis space, 302
 hypothesis testing, 117, 268
 composite hypothesis testing, 268
 M -ary hypothesis testing, 117
 outlier hypothesis testing, 278
 universal outlier hypothesis testing, 279
- I-MMSE, 209, 391–392
 ICA, *see* independent component analysis
 illum information, 289
 image inpainting, 164, 181
 image registration, 269
 image super-resolution, 164, 181
 independence clustering, 281
 independent component analysis, 22, 281
 inference, 198, 265, 268–271, 362
 Bayesian inference, 198
 high-dimensional inference, 198
 information bottleneck, 25, 283, 342–347
 algorithm, 349
 method, 345–347
 information dimension, 97, 98
 k -th order information dimension, 98
 Rényi information dimension, 97
 information dropout, 349
 information sequence, 212
 information source, 2
 binary, 4
 information theory, 265, 330
 information-based clustering, 284
 information-theoretical limits, 386–401
 instance, 303
 instance space, 302
 inter-cluster similarity, 274
 intra-cluster similarity, 274
 invariant representations, 353
 Ising model, 281
 iterative hard thresholding (IHT), 74
 iterative sketching, 120
- JPEG, 75
 JPEG2000, 75
- Karhunen–Loève
 expansion, 57
 kernel function, 122
 kernel matrix, 122
 kernel ridge regression, 122
- Kolmogorov complexity, 274
 Kullback–Leibler divergence, 277, 498,
 499, 524
- labeled data, 264
 large sieve, 165, 168, 170, 187–189
 inequality, 187
 large-system limit, 202
 Lasso, 119
 lautum information, 289
 Le Cam's method, 513
 learning algorithm, 303
 learning distribution over permutations of options,
 229, 230
 learning from comparisons, 229
 learning from first-order marginals, 230
 learning model, 337
 learning problem, 303
 learning rate, 347
 least absolute shrinkage and selection operator
 (LASSO), 74
 least-angle regression (LARS), 74
 least-squares optimization problem, 110
 Lempel–Ziv algorithm, 266
 likelihood ratio, 268
 likelihood ratio test, 268
 linear model
 generalized, 203
 standard, 200
 linear regression, 489
 loss, 360–363
 average logarithmic loss, 334
 cross-entropy loss, 334, 335, 341–342
 empirical loss, 303, 361
 function, 302
 in-sample prediction loss, 364
 out-sample prediction loss, 364
 prediction loss, 360–363
 validation loss, 367, 376
 loss function, 488, 510
 lossless compression, 266
 lossless source coding, 339
 lossy compression, 267
 lossy source coding, 339–341
 low-rank, 73, 120
 low-rank matrix estimation, 119
 Luce model, 231
- machine learning, 27
 Markov chain, 496
 matrix factorization problem, 223
 maximum *a posteriori*, 333
 maximum-likelihood estimator (MLE),
 361, 387
 maximum mutual information decoder, 267
 McDiarmid's concentration inequality, 354
 mean-squared covariance, 219

- mean-squared error, 52, 201
 - estimator, 54, 64
 - minimal, 53
 - minimum, 201
 - prediction, 201
- metric entropy, 116
- min-cut max-flow bounds, 476
- minimal mean-squared error, 47
- minimax estimation
 - global approach, 513, 519
 - local approach, 513
- minimax regret, 352
- minimax risk, 116, 510
 - dictionary learning
 - lower bounds, 140–142, 154–155
 - tightness, 141, 155, 158
 - upper bounds, 142, 155
- minimax rate, 368
- minimum description length, 350, 368
- minimum mean-squared error (MMSE), 202, 208, 219
- minimum mean-squared error sequence, 214
- minimum partition information, 285
- minimum partition information decoder, 268
- Minkowski dimension, 164, 185
 - modified, 165, 185
- misclassification probability, 333, 341–342
- model selection, 326, 348, 350, 359–363
 - asymptotic efficiency, 363
 - consistency, 361
 - criteria, 363, 365–375
 - for inference, 362
 - for prediction, 362
 - model class, 338, 360
 - model complexity, 365
 - model parameters, 360
 - non-parametric, 360
 - optimal model, 360–365
 - parametric, 360
 - restricted model class, 337
 - true model, 360–365
- modeling-procedure selection, 375–379
- MPEG, 75
- multi-task learning, 331
- multicast networks, 476
- multiinformation, 287
- multinomial logit model, 231, 235
 - mixture of multinomial logit models, 232, 256
 - nested multinomial logit model, 232
- multiple hypothesis testing, 490, 511
- mutual information, 11, 124, 202, 208, 306, 341–342, 390–393
 - bounds, 492, 495
 - bounds via KL divergence, 492, 498
 - conditional mutual information, 282
 - empirical mutual information, 348
- normalized mutual information, 275
- tensorization, 492
- nearest neighbor, 276, 360
- network coding, 476
- network functional compression, 456
- neural networks, 25, 326, 376
 - deep convolutional network, 264
 - deep neural network, 331
 - deep recurrent network, 264
 - feed-forward neural network, 376
 - gradient descent, 25
 - model, 232
 - non-parametric, 121, 360
 - rectified linear unit, 25
 - regression, 121
 - sparse model, 232, 234, 236
- null-space property, 165, 185
 - set-theoretic, 165, 184–187
- Nyquist
 - rate, 50
- Nyquist–Shannon sampling theorem, 73
- objectives of data analysis, 362
- Occam’s principle, 78
- optimization, 105, 515
 - convex, 521
- orthogonal matching pursuit (OMP), 74
- orthonormal basis, 172
- outlier hypothesis testing, 278
- overfitting, 232, 345, 370, 374
- PAC-based generalization bounds, 312, 348
- packing, 513, 514, 518
 - number, 513
- packing number, 116
- parametric, 360
- parametric model, 231
- parsimony, 164, 180
 - fractal structure, 184
 - manifold structure, 184
- partition information, 284
- pattern classification, 333
- PCA, *see* principal component analysis
- peeling algorithm, 239
- phase diagram, 198
- phase retrieval, 92
- phase transition, 202, 219, 384, 401
 - channel coding, 12
 - computational, 17
 - source coding, 8
 - statistical, 17
- Plackett model, 231
- planted-clique problem, 402
- planted dense subgraph (PDS) problem, 404
- polynomial-time randomized reduction, 403

- population risk, 302
positive-rate data compression, 431, 437, 439, 442, 444, 445, 447, 449
posterior distribution, 197
 posterior marginal distribution, 206
power graph, 461
power spectral density, 19
prediction, 201, 268, 362
prediction error, 364, 366, 368, 377, 379
prediction loss, 360–363
principal component analysis, 22
 sparse, 23
 spiked covariance model, 22
 statistical phase transition, 23
prior distribution, 197
 Bernoulli–Gaussian prior, 203
 Gaussian mixture model, 209
privacy, 105, 124
 differential privacy, 125
 privacy attacks, 125
probabilistic method, 10
probit model, 231, 235
process
 continuous-time, 49
 cyclostationary, 64
 stationary, 49
projected gradient method, 110
proximal gradient method, 110
PSD
 unimodal, 66
- QR decomposition, 110
- Rand index, 275
random projections, 104
random utility maximization, 231
random utility model, 231, 235, 247
randomized algorithm, 104
rank centrality algorithm, 247
ranking, 233, 236
rate
 Landau, 54
rate region
 achievable rate region, 458
rate distortion, 77, 267
 achievable distortion, 339
 achievable fidelity, 339
 achievable rate, 339
 distortion-rate function, 340
 rate-distortion function, 77, 95, 340
 sampling, 60
rate-distortion coding, 3
rate-distortion dimension, 98, 99
rate-relevance function, 343
receiver operating characteristic (ROC) curve, 201
reconstruction algorithm, 73
recovery algorithm
 phase-retrieval recovery algorithm, 93
regression, 110, 200, 303, 359, 362–366, 368, 370, 375–377, 379
 linear, 200, 359, 362, 363, 368
 polynomial, 363
regularization, 91, 113, 320, 326, 345
reinforcement learning, 264
relative entropy, 207, 305
replica method, 199
replica-symmetric formula, 202, 215
representation learning, 333–338
 information complexity, 349
reproducing kernel Hilbert space, 121
restricted isometry property, 237
risk
 Bayes, 208, 337
 cross-entropy risk, 335
 empirical cross-entropy risk, 350
 empirical risk, 303, 336
 excess, 304, 338
 expected empirical risk, 325
 expected excess risk, 304
 expected risk, 302
 induced risk, 338
 minimum cross-entropy risk, 337
 population risk, 302
 risk function, 335
 true risk, 335
- sample, 360
 sample size, 360
sample complexity, 314, 317, 322, 323
sampler, 50
 bounded linear, 50
 linear
 time-invariant, 51
 multi-branch, 51
 uniform, 51, 63
- sampling
 adaptive, 489
 bounded linear, 50
 branch, 51
 density
 symmetric, 51
 multi-branch, 67
 non-adaptive, 489
 optimal, 54, 57
 rate, 63
 set, 50
 source coding, 49
 uniform
 ADX, 63
- sampling rate, 73
 optimal, 59
 sub-Nyquist, 59

- sampling theorem, 73
- second-moment method, 389–390
- Shannon–Fano–Elias coding, 266
- signal
 - analog, 49
 - bandlimited, 50
 - noise, 50
 - stationary, 49
- signal alignment, 269
- signal declipping, 164, 181
- signal recovery, 164
 - noisy signal recovery, 172, 176
 - sparse signal recovery, 164
 - stable signal recovery, 172
- signal representation, 177
 - sparse signal representation, 177
- signal separation, 164, 180–184
 - noiseless signal separation, 181
 - sparse signal separation, 164
- signal structure, 164
- signature condition, 238
- similarity
 - function, 274
 - inter-cluster similarity, 274
 - intra-cluster similarity, 274
 - score, 274
- similarity criterion, 273
- single-community detection problem, 401–412
- single-community model, 384
- single-community recovery problem, 394–401
- single-crossing property, 209, 218
- singular value decomposition, 110
- sketches
 - Gaussian sketches, 105
 - Rademacher sketches, 106
 - randomized orthonormal systems
 - sketches, 106
 - sketches based on random row sampling, 107
 - sparse sketches based on hashing, 108
 - sub-Gaussian sketches, 106
- sketching, 104
 - iterative sketching, 120
 - sketching kernels, 123
- Slepian–Wolf, 467, 473
 - Slepian–Wolf compression with feedback, 473
- source, 339
- source coding, 339–341
 - achievability, 5
 - blocklength, 8
 - converse, 7
 - decoder, 3
 - encoder, 3
 - indirect, 54
 - lossless, 3, 4
 - lossless source coding, 339
 - lossy, 3
 - lossy source coding, 339–341
 - near-lossless, 5
 - noisy lossy source coding, 342–344
 - remote, 54
 - sampling, 49
 - variable length, 5
- source coding with a single helper, 432
- source coding with multiple helpers, 432
- space
 - Paley–Wiener, 54
- sparse coding, 24
- sparse linear regression, 516
- sparse PCA, 414
- sparse representation, 73
- sparsity, 73, 119, 164, 180, 201, 234
 - level, 181
 - block-sparsity, 73
- spectral methods, 258
- spectrum
 - density power, 50
 - density power
 - rectangular, 61
 - triangular, 61
 - occupancy, 50
- square-root bottleneck, 184
- stability, 304, 308, 311
 - in T -information, 310
 - in erasure mutual information, 309
 - in erasure T -information, 308
 - in mutual information, 310
- standard linear model, 200, 213–215
- statistical inference, 26, 28
 - covariate, 28
 - estimation, 27, 28
 - estimation error, 29
 - generalized linear model, 29
 - hypothesis testing, 26, 30
 - alternate hypothesis, 30
 - binary, 30
 - maximum *a posteriori* test, 31
 - multiple, 30
 - Neyman–Pearson test, 30
 - null hypothesis, 30
 - Rényi information measures, 31
 - type-I error, 30
 - type-II error, 30
 - I-MMSE relationship, 29
 - model selection, 26, 28
 - Akaike information criterion, 28
 - minimum description length, 28
 - non-parametric, 26
 - parametric, 26
 - prediction error, 29
 - regression, 27, 28
 - regression parameters, 28

- standard linear model, 29
- statistical learning, 27, 33
 - clustering, 27
 - density estimation, 27
 - empirical risk minimization, 34
 - generalization error, 27, 33
 - loss function, 33
 - supervised
 - training data, 33
 - supervised learning, 27
 - test data, 27
 - testing, 27
 - testing error, 27
 - training, 27
 - training data, 27
 - training error, 27
 - unsupervised learning, 27, 34
 - validation data, 27
- statistical model
 - model class, 360
 - model complexity, 365
 - model parameters, 360
 - non-parametric model, 360
 - parametric model, 360
- statistical risk minimization, 139, 153
- stochastic block model (SBM), 385
 - k* communities, 416
 - single community, 385
 - two communities, 385
- stochastic complexity, 351
- stochastic gradient descent algorithm, 356
- sub-Nyquist sampling, 49
- sub-Gaussian, 307
- subject, 493
- submatrix detection model, 385
- submatrix detection problem, 401–412
- subspace pursuit, 74
- sum-of-squares relaxation hierarchy, 412
- supervised learning, 264
- supervised learning problem, 303
- support vector machine, 276
- synchronization problem, 393
- system
 - linear
 - time-varying, 50
- T*-information, 306
- tangent cone, 109
- tensor PCA, 414
- tensors
 - CP decomposition, 151
 - data, 134–136
 - multiplication, 151
 - Tucker decomposition, 151
 - unfolding, 151
- test phase, 375
- test set, 375
- test statistic, 201
- testing against independence, 431–433, 439–441, 444
- Thurstone–Mosteller model, 231
- total variation, 499, 524
- total variation distance, 305
- training data, 264
- training phase, 336, 375, 377
- training samples, 346
- training set, 264, 303, 367, 375, 377
- tree networks, 458
 - general tree networks, 457, 467
 - one-stage tree networks, 457, 463
- true positive rate, 201
- type 1 error probability, 429
 - constant-type constraint, 429
 - exponential-type constraint, 429
- type 2 error probability, 429
 - type 2 error exponent, 429
- typical sequences, 6, 430–431
- typical set, 509
- uncertainty principle, 163
- uncertainty relation, 163, 165–181, 189
 - coherence-based, 164, 170, 174–175, 177–179
 - in $(\mathbb{C}^m, \|\cdot\|_1)$, 174–181
 - in $(\mathbb{C}^m, \|\cdot\|_2)$, 165–173
 - infinite-dimensional, 189
 - large sieve-based, 168–170
 - set-theoretic, 164
- unconstrained least-squares problem, 110, 112, 118
- underfitting, 371, 373, 374
- universal classification, 270
- universal clustering, 270, 273
- universal communication, 267–268
- universal compressed sensing, 98
- universal compression, 266–267
- universal denoising, 269
- universal inference, 268–271
- universal information theory, 265
- universal outlier hypothesis testing, 279
- universal prediction, 268
- universal similarity, 275
- unlabeled data, 263
- unsupervised learning, 264, 331
- unsupervised learning problem, 303
- utility, 235
 - random utility, 235
- validation loss, 367, 376
- validation phase, 375, 377, 378

- validation set, 367, 375, 377
Vapnik–Chervonenkis theory, 318
 Vapnik–Chervonenkis class, 318
 Vapnik–Chervonenkis dimension, 318
variance, 365
variational auto-encoder, 332, 348,
 349
water-filling, 55, 58, 65
water-filling
 ADX, 56
 reverse, 48
zero-rate data compression, 434, 441, 445, 448
zigzag condition, 463