Index

0/1 loss function, 53
5 × 2
  cross-validation, 560
  cv paired F test, 575
  cv paired t test, 575

Active learning, 484
AdaBoost, 500
Adaptive resonance theory, 323
Additive models, 207
Agglomerative clustering, 177
AIC, *see* Akaike's information criterion
Akaike's information criterion, 86, 470
Alignment, 364
Analysis of variance, 576
Anchor, 329
Anomaly detection, 199
ANOVA, *see* Analysis of variance
Approximate normal test, 572
Apriori algorithm, 58
Area under the curve, 563
ART, *see* Adaptive resonance theory
Artificial neural networks, 267
Association rule, 4, 56
Attribute, 93
AUC, *see* Area under the curve
Autoencoder, 302

Backpropagation, 283
  through time, 306
Backup, 526
Backward selection, 117
Backward variable, 426
Bag of words, 108, 364
Bagging, 498
Base-learner, 487
Basis function, 241
  cooperative vs. competitive, 335
  for a kernel, 462
  normalization, 333
Basket analysis, 56
Batch learning, 285
Baum-Welch algorithm, 430
Bayes factor, 469
Bayes' ball, 399
Bayes' classifier, 53
Bayes' estimator, 72
Bayes' rule, 51, 597
Bayesian information criterion, 86, 470
Bayesian model combination, 494
Bayesian model selection, 86
Bayesian networks, 387
Belief networks, 387
Belief state, 535
Bellman's equation, 522
Beta distribution, 450
Beta process, 483
Between-class scatter matrix, 141
Bias, 69

Bias unit, 271
Bias/variance dilemma, 82
BIC, *see* Bayesian information criterion
Binary split, 215
Binding, 230
Binomial test, 571
Biometrics, 510
Blocking, 554
Bonferroni correction, 580
Boosting, 499
Bootstrap, 561

C4.5, 219
C4.5Rules, 225
Canonical correlation analysis, 145
CART, 219, 231
Cascade correlation, 299
Cascading, 507
Case-based reasoning, 206
Causality, 396
  causal graph, 388
CCA, *see* Canonical correlation analysis
Central limit theorem, 602
Chinese restaurant process, 478
Class
  confusion matrix, 564
  likelihood, 52
Classification, 5
  likelihood- vs. discriminant-based, 239
Classification tree, 216
Clique, 408
Cluster, 162
Clustering, 11
  agglomerative, 177
  divisive, 177
  hierarchical, 176
  online, 319
Code word, 164
Codebook vector, 164
Coefficient of determination (of regression), 80
Color quantization, 163
Common principal components, 127
Competitive basis functions, 335
Competitive learning, 318
Complete-link clustering, 177
Component density, 162
Compression, 8, 164
Condensed nearest neighbor, 194
Conditional independence, 388
Confidence interval
  one-sided, 567
  two-sided, 566
Confidence of an association rule, 57
Conjugate prior, 450
Connection weight, 271
Contingency table, 573
Convolutional neural network, 294
Correlation, 95
Cost-sensitive learning, 550
Coupled HMM, 436
Covariance function, 475
Covariance matrix, 94
Credit assignment, 518
Critic, 518
CRM, *see* Customer relationship management
Cross-entropy, 251
Cross-validation, 40, 83, 558
  K-fold, 559
  5×2, 560
Curse of dimensionality, 192
Customer relationship management, 173
Customer segmentation, 173

d-separation, 399
Decision forest, 234
Decision node, 213

Decision region, 55
Decision tree, 213
  multivariate, 230
  omnivariate, 233
  soft, 343
  univariate, 215
Deep learning, 308
Deep neural networks, 307
Dendrogram, 177
Density estimation, 11
Dichotomizer, 56
Diffusion kernel, 365
Dimensionality reduction
  nonlinear, 303
Directed acyclic graph, 387
Dirichlet distribution, 449
Dirichlet process, 479
Discount rate, 521
Discriminant, 5
  function, 55
  linear, 103
  quadratic, 101
Discriminant-based classification, 239
Distance learning, 197
Distributed vs. local representation, 174, 327
Diversity, 488
Divisive clustering, 177
Document categorization, 108
Doubt, 26
Dual representation, 381, 461
Dynamic classifier selection, 502
Dynamic node creation, 298
Dynamic programming, 523

Early stopping, 253, 292
ECOC, *see* Error-correcting output codes
Edit distance, 181, 364
Eigendigits, 125
Eigenfaces, 125
Eligibility trace, 530
EM, *see* Expectation -maximization
Emission probability, 421
Empirical error, 24
Empirical kernel map, 364
Ensemble, 492
Ensemble selection, 506
Entropy, 216
Episode, 521
Epoch, 285
Error
  type I, 569
  type II, 569
Error-correcting output codes, 368, 496
Euclidean distance, 104
Evidence, 52
Example, 93
Expectation-maximization, 168
  supervised, 337
Expected error, 548
Experiment
  design, 550
  factorial, 553
  strategies, 552
Explaining away, 393
Extrapolation, 35

FA, *see* Factor analysis
Factor analysis, 130
Factor graph, 409
Factorial HMM, 435
Feature, 93
  extraction, 116
  selection, 116
Feature embedding, 128
Finite-horizon, 521
First-order rule, 229
Fisher kernel, 365
Fisher's linear discriminant, 141
Flexible discriminant analysis, 127
Floating search, 118

Foil, 227
Forward selection, 117
Forward variable, 424
Forward-backward procedure, 424
Fuzzy *k*-means, 179
Fuzzy membership function, 333
Fuzzy rule, 333

Gamma distribution, 454
Gamma function, 449
Gaussian process, 474
Generalization, 24, 39
Generalized linear models, 263
Generative model, 396
generative model, 447
Generative topographic mapping, 344
Geodesic distance, 148
Gini index, 217
Gradient descent, 249
  stochastic, 275
Gradient vector, 249
Gram matrix, 361
Graphical models, 387
Group, 162
GTM, *see* Generative topographic mapping

Hamming distance, 197
Hebbian learning, 321
Hidden layer, 279
Hidden Markov model, 421
  coupled, 436
  factorial, 435
  left-to-right, 436
  switching, 436
Hidden variables, 59, 396
Hierarchical clustering, 176
Hierarchical cone, 294
Hierarchical Dirichlet process, 482
Hierarchical mixture of experts, 342
Higher-order term, 241
Hinge loss, 357
Hint, 295
Histogram, 187
HMM, *see* Hidden Markov model
Hotelling's multivariate test, 585
Hybrid learning, 329
Hypothesis, 23
  class, 23
  most general, 24
  most specific, 24
Hypothesis testing, 568

ID3, 219
IF-THEN rules, 225
iid (independent and identically distributed), 41
Ill-posed problem, 38
Impurity measure, 216
Imputation, 95
Independence, 388
Indian buffet process, 483
Inductive bias, 38
Inductive logic programming, 230
Infinite-horizon, 521
Influence diagrams, 411
Information retrieval, 564
Initial probability, 418
Input, 93
Input representation, 21
Input-output HMM, 433
Instance, 93
Instance-based learning, 186
Interest of an association rule, 57
Interpolation, 35
Interpretability, 225
Interval estimation, 564
Irep, 227
Isometric feature mapping, 148

Job shop scheduling, 541
Junction tree, 407

K-armed bandit, 519

K-fold
  cross-validation, 559
  cv paired *t* test, 574
k-means clustering, 165
  fuzzy, 179
  online, 319
k-nearest neighbor
  classifier, 193
  density estimate, 191
  smoother, 203
k-nn, *see* *k*-nearest neighbor
Kalman filter, 436
Karhunen-Loève expansion, 127
Kernel estimator, 189
Kernel function, 188, 360, 463
Kernel PCA, 379
Kernel smoother, 203
Kernelization, 361
Knowledge extraction, 8, 226, 333
Kolmogorov complexity, 86
Kruskal-Wallis test, 583
Kullback-Leibler distance, 473

Laplace approximation, 465
Laplacian eigenmaps, 153
Laplacian prior, 460
Large margin component analysis, 379
Large margin nearest neighbor, 197, 378
Lasso, 460
Latent Dirichlet allocation, 480
Latent factors, 130
Latent semantic indexing, 136
Lateral inhibition, 320
LDA, *see* Linear discriminant analysis
Leader cluster algorithm, 166
Leaf node, 214
Learning automata, 541
Learning vector quantization, 338
Least square difference test, 579
Least squares estimate, 79
Leave-one-out, 559
Left-to-right HMM, 436
Level of significance, 569
Levels of analysis, 268
Lift of an association rule, 57
Likelihood, 66
Likelihood ratio, 60
Likelihood-based classification, 239
Linear classifier, 103, 246
Linear discriminant, 103, 240
Linear discriminant analysis, 140
Linear dynamical system, 436
Linear opinion pool, 492
Linear regression, 79
  multivariate, 109
Linear separability, 245
Local outlier factor, 200
Local representation, 326
Locally linear embedding, 150
Locally weighted running line smoother, 204
Loess, *see* Locally weighted running line smoother
Log likelihood, 66
Log odds, 61, 248
Logistic discrimination, 250
Logistic function, 248
Logit, 248
Loss function, 53
LVQ, *see* Learning vector quantization

Mahalanobis distance, 96
Margin, 25, 351, 501
Marginal likelihood, 467
Markov chain Monte Carlo sampling, 461
Markov decision process, 521
Markov mixture of experts, 434
Markov model, 418
  hidden, 421

learning, 420, 429
observable, 419
Markov random field, 407
Matrix factorization, 135
Max-product algorithm, 410
Maximum a posteriori (MAP) estimate, 72, 447
Maximum likelihood estimation, 66
McNemar's test, 573
MDP, *see* Markov decision process
MDS, *see* Multidimensional scaling
Mean square error, 69
Mean vector, 94
Mean-field approximation, 473
Memory-based learning, 186
Minimum description length, 86
Mixture components, 162
Mixture density, 162
Mixture of experts, 339, 502
competitive, 342
cooperative, 341
hierarchical, 343
Markov, 434
Mixture of factor analyzers, 173
Mixture of mixtures, 174
Mixture of probabilistic principal component analyzers, 173
Mixture proportion, 162
MLE, *see* Maximum likelihood estimation
Model combination
multiexpert, 491
multistage, 491
Model evidence, 467
Model selection, 38
MoE, *see* Mixture of experts
Momentum, 291
Moralization, 408
Multi-view learning, 489
Multidimensional scaling, 137
nonlinear, 325
using MLP, 304
Multilayer perceptrons, 279
Multiple comparisons, 579
Multiple kernel learning, 366, 510
Multivariate linear regression, 109
Multivariate polynomial regression, 111
Multivariate tree, 230

Naive Bayes' classifier, 397
discrete inputs, 108
numeric inputs, 103
Naive estimator, 187
Nearest mean classifier, 104
Nearest neighbor classifier, 194
condensed, 194
Negative examples, 21
Neocognitron, 294
Neuron, 267
No Free Lunch Theorem, 549
Noise, 30
Noisy OR, 406
Nonparametric estimation, 185
Nonparametric tests, 580
Normal-gamma distribution, 454
Normal-Wishart distribution, 455
Novelty detection, 9, 199
Null hypothesis, 569

Observable Markov model, 419
Observable variable, 50
Observation, 93
Observation probability, 421
OC1, 231
Occam's razor, 32
Off-policy, 528
Omnivariate decision tree, 233
On-policy, 528
One-class classification, 199, 374
One-sided confidence interval, 567
One-sided test, 570
Online k-means, 319
Online learning, 275

Optimal policy, 522
Optimal separating hyperplane, 351
Outlier detection, 9, 199, 374
Overfitting, 39, 82
Overtraining, 292

PAC, *see* Probably approximately correct
Paired test, 573
Pairing, 554
Pairwise separation, 246, 497
Parallel processing, 270
Partially observable Markov decision process, 534
Parzen windows, 189
Pattern recognition, 6
PCA, *see* Principal component analysis
Pedigree, 435
Perceptron, 271
Phone, 437
Phylogenetic tree, 413
Piecewise approximation
  constant, 283, 338
  linear, 339
Policy, 521
Polychotomizer, 56
Polynomial regression, 79
  multivariate, 111
Polytree, 404
POMDP, *see* Partially observable Markov decision process
Positive examples, 21
Post hoc testing, 579
Posterior probability, 446
Posterior probability of a class, 52
Posterior probability of a parameter, 71
Postpruning, 222
Potential function, 242, 408
Power function, 570
Precision
  in information retrieval, 564
  reciprocal of variance, 453
Precision matrix, 455
Predicate, 229
Prediction, 5
Prepruning, 222
Principal component analysis, 121
Prior knowledge, 332
Prior probability, 446
Prior probability of a class, 52
Prior probability of a parameter, 71
Probabilistic networks, 387
Probabilistic PCA, 132
Probably approximately correct learning, 29
Probit function, 466
Product term, 241
Projection pursuit, 310
Proportion of variance, 124
Propositional rule, 229
Pruning
  postpruning, 222
  prepruning, 222
  set, 222

Q learning, 528
Quadratic discriminant, 101, 241
Quantization, 164

Radial basis function, 328
Random forest, 235
Random Subspace, 490
Randomization, 554
Ranking, 11, 260
  kernel machines, 373
  linear, 261
RBF, *see* Radial basis function
Real time recurrent learning, 306
Recall, 564
Receiver operating characteristics, 562
Receptive field, 326

Recommendation systems, 59, 156
Reconstruction error, 127, 164
Recurrent network, 305
Reference vector, 164
Regression, 9, 35
  linear, 79
  polynomial, 79
  polynomial multivariate, 111
  robust, 369
Regression tree, 220
Regressogram, 201
Regularization, 84, 301
Regularized discriminant analysis, 106
Reinforcement learning, 13
Reject, 34, 54
Relative square error, 80
Replication, 554
Representation, 21
  distributed vs. local, 326
Response surface design, 553
Ridge regression, 301, 460
Ripper, 227
Risk function, 53
Robust regression, 369
ROC, *see* Receiver operating characteristics
RSE, *see* Relative square error
Rule
  extraction, 333
  induction, 226
  pruning, 226
Rule support, 226
Rule value metric, 227
Running smoother
  line, 204
  mean, 201

Sammon mapping, 139
  using MLP, 304
Sammon stress, 140
Sample, 50
  correlation, 95
  covariance, 95
  mean, 95
Sarsa, 529
  Sarsa(λ), 531
Scatter, 141
Scree graph, 124
Self-organizing map, 324
Semiparametric density estimation, 162
Sensitivity, 564
Sensor fusion, 489
Sequential covering, 227
Sigmoid, 248
Sign test, 581
Single-link clustering, 177
Singular value decomposition, 135
Slack variable, 355
Smoother, 201
Smoothing splines, 205
Soft count, 430
Soft decision tree, 234
Soft error, 355
Soft weight sharing, 301
Softmax, 254
SOM, *see* Self-organizing map
Spam filtering, 109
Specificity, 564
Spectral clustering, 175
Spectral decomposition, 123
Speech recognition, 436
Sphere node, 231
Stability-plasticity dilemma, 319
Stacked generalization, 504
Statlib repository, 17
Stochastic automaton, 418
Stochastic gradient descent, 275
Stratification, 559
Strong learner, 499
Structural adaptation, 297
Structural risk minimization, 86
Subset selection, 116

Sum-product algorithm, 409
Supervised learning, 9
Support of an association rule, 57
Support vector machine, 353
SVM, *see* Support vector machine
Switching HMM, 436
Synapse, 268
Synaptic weight, 271

t distribution, 567
t test, 570
Tangent prop, 297
TD, *see* Temporal difference
Template matching, 104
Temporal difference, 525
  learning, 528
  TD(0), 529
  TD-Gammon, 541
Test set, 40
Threshold, 242
  function, 272
Time delay neural network, 305
Topic modeling, 480
Topographical map, 325
Transition probability, 418
Traveling salesman problem, 344
Triple trade-off, 39
Tukey's test, 584
Two-sided confidence interval, 566
Two-sided test, 569
Type I error, 569
Type II error, 569

UCI repository, 17
Unbiased estimator, 69
Underfitting, 39, 82
Unfolding in time, 306
Unit normal distribution, 565
Univariate tree, 215
Universal approximation, 283
Unobservable variable, 50
Unstable algorithm, 498

Validation set, 40
Value iteration, 523
Value of information, 535, 539
Vapnik-Chervonenkis (VC)
    dimension, 27
Variance, 70
Variational approximation, 472
Vector quantization, 164
  supervised, 338
Version space, 24
Vigilance, 323
Virtual example, 296
Viterbi algorithm, 428
Voronoi tesselation, 194
Voting, 492

Weak learner, 499
Weight
  decay, 298
  sharing, 295
  sharing soft, 301
  vector, 242
Wilcoxon signed rank test, 583
Winner-take-all, 318
Wishart distribution, 455
Within-class scatter matrix, 142
Wrapper, 117

z, *see* Unit normal distribution
z-normalization, 97, 602
Zero-one loss, 53