# Index


$P$ statistic, 903
$T$-squared sampling tests, 902
$\Gamma$ statistic, 870, 876
$\gamma$ statistic, 874
$\nu$-SVM, 135
*k-medoids* algorithms, 744–746
$l_p$ metric dissimilarity measures, 604–605
$t$-distribution, 272, 925
$t$-test, 273, 576
$k$ Nearest neighbor density estimation, 56–57

## A

Absolute moments, 414
Acceptance interval, 271
Activation function, 101
Active learning, 511
AdaBoost, 231
Adaline, 108
Adaptive fuzzy C-shells (AFCS) algorithm, 724
Adaptive momentum, 182
Adaptive resonance theory (ART2), 635
Agglomerative algorithms, 629
  definitions, 655–658
  generalized, 654–655
  graph theory and, 667–676
  implementation issues, 667
  matrix theory and, 658–664
  monotonicity and crossover, 664–667
  proximity matrix, 676–679
Agglomerative hierarchical algorithms, 654
Akaike Information Criterion, 306, 310, 887
Algebraic distance, 719, *see also* distance between a point and a quadratic surface
Alternating Cluster Estimation (ACE), 733
Alternating Optimization (AO), 716
Alternative hypothesis, 268
Analysis filters, 377
Ancillary unit, 161
Angular second moment, 416
Any path method, 536
Approximation properties, 194–196
Arithmetic average rule, 224
Audio classification, 451–466
Auto-associative networks, 298
Autocorrelation matrix, 104, 430, 433
Autoregressive-moving average (ARMA ($p$, $m$)), 429
Autoregressive processes (AR), 428
  one-dimensional, 428–429
  two-dimensional, 431
Average expected quantization error, 750
Average partition density (PA), 891
Average partition shell density, 892
Average proximity function, 616, 621
Average Risk, 16
  minimizing, 16–18

## B

Backpropagation algorithm, 162, 167
  batch and pattern modes, 169
  momentum term, 170
Bagging (bootstrap aggregating), 220, 226
Barlow's hypothesis, 342
Barycentric correction procedure, 101
Basic sequential algorithmic scheme (BSAS), 633–634
  modified, 637–638
Basis images, 325
Basis images (matrices), 325
Basis sequences, 365, 378
Basis vectors, 324
Batch mode, 168
  algorithms, 780
Baum–Welch reestimation, 540
Bayes classification rule, 14
  naive, 59–61
  normal distributions and, 20–33
Bayes decision theory, 13
Bayes rule, 13
Bayesian Inference, 39
Bayesian Information Criterion (BIC), 310, 887
Bayesian learning, 175–176
Bayesian networks, 66
Bellman's optimality principle, 484
Bending energy, 442
Best path method, 539
Beta distribution, 925
Between-class scatter matrix, 281
Bhattacharyya distance, 278
Bias–Variance Dilemma, 114
Binary morphology clustering algorithms (BMCAs), 789
  algorithmic scheme, 796–798
  cluster determination in a discretebinary set, 794–795
  discretization, 790–791

Binary morphology clustering algorithms (BMCAs) (*continued*)
feature vector assignment, 795–796
morphological operations, 791–794
Bioinformatics, 632
Biorthogonal expansion, 378
Biorthogonality condition, 381
BIRCH algorithm, 690
BLAST, 838
Boosting, 230
Bootstrap, 220, 572
Bootstrap aggregating, 220, 226
Bootstrap method, 572
Bootstrapping techniques, 866
Boundary detection algorithms (BDA), 630
Branch and bound, 288
Branch and bound clustering (BBC) algorithms, 630, 805
Branch and bound methods, 803
Brownian motion, 446–447
BUBBLE and BUBBLE-FM algorithms, 690

## C

c-means algorithm, *see* isodata algorithm
Cascade correlation, 181
Cauchy–Schwarz inequality, 499
Cell-based clustering method (CBF), 832
Center of mass, 442
Central limit theorem, 20, 923
Central moments, 413
Centroid condition, 750
Cepstrum, 456
Ceptral coefficients, 494
Chain codes, 439
Chaining effect, 660
Chameleon algorithm, 686
Channel equalization, 527
Character recognition, 1–2
Characteristic functions, 601
Chernoff bound, 278
Chi-square distribution, 274, 924–925
Circular backpropagation model, 184
CLARA algorithm, 748
CLARANS, 748
Class imbalance problem, 237
Class separability measures, 276–283
Classification error probability, 6, 15
minimizing, 15
Classification task, 5
Classification tree, 804
Classifiers, 4
Bayes decision theory and, 13–71
design stage, 6
generalization performance of, 120
linear, 91–141, 185–187
nonlinear, 151–239
polynomial, 189–190
soft margin, 209
CLIQUE algorithm, 825
Closing, 793
CLTree algorithm, 832
Cluster, 595, 600
compact, 701
linear-shaped, 701
ring-shaped, 701, 708
shell-shaped, 701
spherical, 728
Cluster-based graph formulation (CBGF), 841
Cluster detection algorithm for discrete-valued sets (CDADV), 794
Cluster validity, 863
hypothesis testing and, 866–876
of individual clusters, 893–896
Cluster variation, 889
Clustering algorithms, *see also under name of*
categories of, 629–632
Clustering algorithms, cost function optimization, 629, 701, 930, *see also under name of*
branch and bound, 630, 803–807
deterministic annealing, 765, 808–810
fuzzy, 630, 712–733
genetic, 630, 810–811
hard, 630, 739–749
mixture decomposition schemes, 703–712
possibilistic, 630, 733–739
simulating annealing, 765, 807–808
vector quantization, 749–751
Clustering algorithms, hierarchical, *see also under name of*
agglomerative, 629, 654–679
applications, 653
cophenetic matrix, 679–680
divisive, 629, 680–682
for large data sets, 682–690
selecting the best, 690–692
Clustering algorithms, sequential, 629, 633–641, *see also under name of*
basic, 633–635
modified basic, 637–638
neural network implementation, 643–646
number of possible, 627–628, 635–636
refinement stages, 641–643
role of, 629
two-threshold scheme, 638–641
Clustering/clusters, 7–8, 600–601, 627
applications, 598–599

compact and hyperellipsoidal, 705–709
criterion, 595–597
definitions, 600–601
proximity measures, 602–622
Clustering criterion, 595–596, 597
Clustering hypothesis, 896
Clustering tendency, 597, 863, 896–905
applications, 896
tests for spatial randomness, 900–905
Co-association matrix, 841
Code vector, 749, *see also* reproduction vector
Combination of clusterings, 839
Combining classifiers, 10, 222–230
Compactness and separation validity function, 8, 890, *see also* Xie–Beni index
Competitive learning algorithms, 631, 780, 789
applications, 782
basic, 782–783
conscientious, 784–785
cost functions and, 785–786
generalized, 781
leaky, 783–784
self-organizing maps, 786–788
vector quantization, 788–789
Competitive learning associated with cost functions, 785
Complementary slackness, 122, 124, 939, 944
Complete link algorithm, 659, 671–672
Computer-aided diagnosis, 2
Computer storage utilization, 749
Concordant pair, 875
Confidence intervals, 271–272
Confusion matrix, 573, 707, 718
Conjugate gradient algorithm, 171, 934
Conscientious competitive learning algorithms, 784
Constant-Q filter banks, 382
Constrained problems, optimization for, 935–945
Constraint-based clustering algorithms, 839
Constraint clustering, 839
Constraints, 497
Constructive techniques, 160, 178, 181
Content-Based Image Retrieval (CBIR), 509
Content-based retrieval, 508
Context-dependent classification, 521
Bayes classifier, 521–522
channel equalization, 527–532
hidden Markov models, 532–551
Markov chain models, 522–523
Markov random fields, 554–556
neural networks and training, 552–553
Viterbi algorithm, 523–527
Contingency table, 608
Continuous observation HMM, 543
Continuous speech recognition (CSR), 491
Contrast, 416
Convex function, 809, 937
Convex hull, 898
Convex programming, 123, 126, 943–944
Convex set, 126, 138, 938
Co-occurrence matrices, 414–417
Cophenetic correlation coefficient (CPCC), 874
Cophenetic distance, 679
Cophenetic matrix, 680
"Corrected" statistic, 872
Correlation, 104
Correlation matrix, 108, 928–929
Correlations, template matching and measures based on, 498–504
Cosine similarity measure, 606
Cost function(s), 701, *see also* clustering algorithms, cost function optimization
backpropagation algorithm and, 162–165
competitive learning algorithms, 780–781
convex, 123
cross-entropy, 173–174
least square, 172
penalty terms, 180
quadratic error, 174
selecting, 172–176
Co-training, 589
Covariance matrix, 21
Cox–Lewis test, 902
Cramer-Rao lower bound, 36, 923
Crisp clustering algorithm, *see* hard/crisp algorithms
Critical bandwidth, 458
Critical interval, 269
Cross-correlation, 104
Cross-correlation coefficient, 499
Cross-entropy, 173, 545
Cross-entropy cost function, 173–174
Cross-validation with active pattern selection, 572
Crossover, 665
Cumulants, 344–345, 917–918
CURE algorithm, 683
Curse of dimensionality, 55
Curvature features, 440

## D

Data
missing, 263–265, 614–615
reduction, 598
Data compression, 598, 749

Data normalization, 263
Daubechies' low-pass filters, 379
Davies–Bouldin (DB) index, 883–884
Davies–Bouldin like indices, 883–884
DBCLASD, 818
DBSCAN, 815
Decision surfaces, 19–20
  hyperplanes, 26–30, 91–92
Decision trees, 215
Decomposition layers, 902
Decomposition technique, sparse, 902–905
Deformable template models, 504–508
Degrees of freedom, 273
Delta-bar-delta, 171
Delta-delta rule, 171
DENCLUE, 819
Dendrogram, 656, *see also* threshold dendrogram
Density-based algorithms, 631
  for large data sets, 815–821
Density function, 819
Deterministic annealing, 808–810
Diagnostic methods, 583
Diameter of a cluster, 882
Dilation, 791
Dimensionality reduction, 331, 350
Directed acyclic graph (DAG), 66
Directed graphs, 667
Directed path, 770
Directed tree, 770
Direction length features, 440
Discordant pairs, 875
Discrete binary (DB) set, 790
Discrete cosine transform (DCT), 366
Discrete Fourier transform (DFT), 363
  one-dimensional, 364–366
  short time, 454
  two-dimensional, 366
Discrete observation HMM models, 539
Discrete sine transform (DST), 367
Discrete time wavelet coefficients, 378
Discrete time wavelet transform (DTWT), 375
  many bands case, 380–384
  two-band case, 376–380
Discrete wavelet frame, 394
Discriminant functions, 19
Discriminative learning, 174
Dispersion of a cluster, 883–884
Dissimilarity matrix, 655
Dissimilarity measure (DM)
  between discrete-valued vectors, 608–609
  between points, 620, 622
  between real-valued vectors, 604–607
  between sets, 603, 620
  defined, 602
Distance between a point and a quadratic surface, 719
  Algebraic distance, 719
  normalized radial, 721
  perpendicular distance, 720
  radial, 720
Distances
  between two sequences, 483
  Bhattacharyya, 278, 279
  classifiers, 30–33
  Euclidean, 30, 603
  Mahalanobis, 30, 32
  Minkowski, 716
Distortion function, 750
Distortion measure, 750
Distributed clustering, 846
Divergence, 276, 278, 283
Divisive hierarchical algorithm(s), 653, 680, 690
DNA microarray analysis, 632
DNA sequencing, 838
Downward closure property, 825, 827, 831
Dunn index, 882
Dunn-like indices, 882
Dynamic Bayesian networks, 545
Dynamic programming, 287, 484
Dynamic Similarity Measures, 610
Dynamic time warping in speech recognition, 491–498

## E

Eccentricity, 442
Edge connectivity, 670
Edge cut set, 687
Edgeworth expansion, 346, 918–919
Edit distance, 487
EM-algorithm, 703
Empirical classification error, 299
ENCLUS algorithm, 830
End point constraints, 494
Entropy, 43, 330, 414, 417
Entropy estimation, maximum, 43–44
Epochs, 169, 175, 182
Erosion, 791
Error counting approach, 568
Euclidean dimension, 445
Euclidean distance, 30, 603
Exclusive OR (XOR) problem, 151–153
Expectation maximization (EM) algorithm, 45
  description of, 45–46, 703, 751–752
  mixture modeling problem and, 46–47
Extended self similar (ESS) processes, 451
External criteria, 864, 867–873
External energy, 505–507

## F

FASTA, 838
Feature generation, audio analysis and, 412
  cepstrum, 455–457
  example, 463–466
  mel-cepstrum, 457–460
  short time processing of signals, 452–455
  spectral features, 460–462
  time domain features, 462–463
Feature generation, image analysis and
  extraction, 421–423
  fractals, 444–451
  local linear transforms for texture, 421–423
  moments, 423–427
  parametric models, 427–435
  shape and size characterization, 435–443
  texture characterization, 412–421
Feature generation, linear transforms and
  applications, 390–396
  basis vectors and images, 324–326
  discrete cosine and sine transforms, 366–368
  discrete Fourier transform, 363–366
  discrete time wavelet transform, 375–384
  Haar transform, 369–375, 382–384
  Hadamard transform, 368–369
  independent component analysis, 342–349
  Karhunen-Loève transform, 326–334
  multiresolution interpretation, 384–386
  singular value decomposition, 335–341
  two-dimensional generalizations, 388–390
  wavelet packets, 387–388
Feature generation methods, 822
Feature selection, 6, 261, 596
  Bayesian information criterion, 309–310
  class separability measures, 276–282
  neural networks and, 298–299
  optimal generation, 288–297
  preprocessing, 262–265
  receiver operating characteristics curve, 275–276
  scalar, 283–284
  statistical hypothesis testing and, 268–275
  subset, 283–288
  Vapnik-Chervonenkis learning theory, 300–309
  vectors, 284–288
Feature selection methods, 822
Features, 4
  defined, 5
  interval scaled, 599
  maps, 184
  nominal, 599
  ordinal, 599
  ratio-scaled, 599, 676
  vectors, 6
  types of, 599
Filter approach, 285
Filter model, 823
Finite impulse response (FIR), 376, 947
Finite state automaton, 535
First-order statistics features, 412–414
Fisher's discriminant ratio (FDR), 282, 289
Fisher's linear discriminant, 204, 294
Floating search methods, 286–287
Fourier descriptors, 437–439
Fourier features, 436–439
Fourier transform, 455
Fowlkes and Mallows index, 870
Fractal dimension, 444–446
Fractals, 444–451
Fractional Brownian motion, 446–451
Frequency ratio, 50
Frobenius norm, 337, 350
Fukuyama-Sugeno index, 890
Fundamental frequency, 461–462
Fuzzifier, 713
Fuzzy approaches, 702, *see also* fuzzy clustering algorithms
Fuzzy C ellipsoidal shells (FCES) algorithm, 726
Fuzzy C plano-quadric shells (FCPQS), 727
Fuzzy C quadric shells (FCQS) algorithm, 727
Fuzzy clustering algorithms, 630, 712–716
  alternating cluster estimation, 733
  convergence aspects, 732–733
  generalized, 715
  geometrical interpretation, 732
  hyperplane representatives, 728–732
  internal criteria, 894–896
  point representatives, 617–619
  quadric surface representatives, 718–724
  relative criteria, 877–880
  shell, 724–728
Fuzzy c-Means (FCM) algorithm, 717
Fuzzy decision trees, 221
Fuzzy density, 891
Fuzzy hypervolume, 891
Fuzzy k-means algorithm, 717
Fuzzy measures, 613
Fuzzy proximity measures, 613–614
Fuzzy shell clustering algorithms, 724–728
Fuzzy shell density, 892

## G

Gabor filter, 394
Gabriel graphs (GG), 770, 882, 884
Gap statistic, 885
Gauss–Newton method, 717
Generalization error probability, 300

Generalization performance of classifiers, 120
Generalized agglomerative scheme (GAS), 654
Generalized competitive learning scheme (GCLS), 781
Generalized divisive scheme (GDS), 681
Generalized fuzzy algorithmic scheme (GFAS), 715
Generalized hard algorithmic scheme (GHAS), 740
Generalized linear classifiers, 185
Generalized mixture decomposition algorithmic scheme (GMDAS), 704
Generalized possibilistic algorithmic scheme (GPAS), 735
Generalized XB index, 890
Generative models, 579
Genetic algorithms, 630, 765, 810, 811
  crossover, 810
  mutation, 810
  reproduction, 810
Geometric average rule, 223–224
Geometric features, 442–443
Geometric moments, 423
Gibbs random fields, 554
Global constraints, 485, 495
Global convergence theorem, 732
Grade of membership, 601
Gradient descent algorithm, 95, 172, 930–933
Graph
  complete, 766
  edges, 667, 766
  inconsistent edges, 766
  vertices, 667
Graph embedding, 361
Graph theory, 765
  agglomerative algorithms, and, 667–676
  directed trees, 770–772
  minimum spanning tree, 675–676, 766–770
  regions of influence, 768–770
Graph theory–based algorithmic scheme (GTAS), 670
Gray level run lengths, 417
Grid-based subspace clustering algorithms (GBSCAs), 825
Gustafson–Kessel (G-K) algorithm, 728

## H

Haar transform, 369, 384
Hadamard transform, 368
Hammersley-Clifford theorem, 554
Hamming distance, 608–609
Hamming window, 454
Hard clustering, 702
Hard/crisp clustering algorithms, 630, 739
  CLARA and CLARANS algorithms, 748–749
  generalized, 740–741
  internal criteria, 867, 873–876
  isodata, k-means, and c-means algorithms, 741–745
  *k*-medoids algorithms, 745–749
  PAM algorithm, 746–748
  relative criteria, 877–880
Hermitian operation, 324
Hessian matrix, 123, 171, 933–934
Hidden layer, 156
Hidden Markov models (HMMs), 532
  coin tossing example, 532–533
  continuous observation, 543–545
  discrete observation, 539–543
  recognition, 536–539
  state duration modeling and, 545–551
  training, 539
Hierarchical clustering algorithms, 629
  agglomerative algorithms, 629
  divisive algorithms, 629
Hierarchical search, 502
High-dimensional data sets, clustering algorithms for, 821–837
Higher order, 184
Hilbert space, 199
Histogram approximation, 50
Ho-Kashyap algorithm, 109
Holdout method, 570
Hopkins test, 901
Hu moments, 425–426
Hubert's $\Gamma$ statistic, 870, *see also* $\Gamma$ statistic
Hughes phenomenon, 303
Hurst parameter, 447
Hybrid bipartite graph formulation (HBGF), 842
Hyperbolic tangent, kernels, 200
Hypercube, 709, 732
Hyperellipses, 30, 719
Hyperparameters, 202
Hyperplanes, 26–30, 91–92
  margin, 106, 182, 266
  representatives, 619, 728–732
Hyperquadrics, 25
Hypersphere, 867
Hyperspherical representatives, 619–620
Hypothesis generation, 598
Hypothesis testing, 598, 864–876
  basics, 268–269
  cluster analysis and, 598
  cluster validity and, 866–876
  feature selection and, 268–275
  known variance case, 269–271
  null, 268
  *t*-test, 273–275
  unknown variance case, 272–273

## I

Images, basis, 324–326
Incomplete data set, 45
Independent component analysis (ICA), 342–343
  based on mutual information, 345–348
  based on second- and fourth-order cumulants, 344–345, 917–918
  identifying condition for, 343–344
  simulation example, 348–349
Inequality constraints, 937–945
Influence function, 819
Information theory based criteria, 886–887
Inner product, 606
Input layer, 156
Instance-based graph formulation (IBGF), 841
Internal criteria, 864, 867, 873–876
Internal energy, 505–506
Interpoint distances, 900
Interpolation functions, 185–186
  polynomial classifiers, 189–190
  radial basis, 190–194
Interpretation of clustering results, 597
Intersymbol interference, 527
Interval scaled, 599
Intrinsic dimension, 331
Invariance, transformation, 183
Inverse difference moment, 417
Isodata algorithm, 742
Isolated word recognition (IWR), 491
ISOMAP, 359
Itakura constraints, 495, 496
Itakura global constraints, 495
Itakura–Saito distortion, 782
Iterative function optimization schemes, 629

## J

Jaccard coefficient, 869–870
Jaccard measure, 842

## K

k-means algorithm, *see* isodata algorithm
Kalman filtering approach, 171
Karhunen–Loève transform, 326, 822
Karush-Kuhn-Tucker (KKT) conditions, 121, 126, 939–941
Kernel clustering methods, 811
Kernel Euclidean distance classifier, 203
Kernel PCA, 351
Kernel perceptron algorithm, 205
Kernels, 51
Kesler's construction, 101
Koch curve, 445–446
Kohonen self-organizing maps, 787
Kronecker's delta, 18
Kruskal's algorithm, 676
Kruskal-Wallis statistic, 274
Kullback-Leibler (KL) probability distance measure, 223, 277, 350, 919–920
Kurtosis, 413

## L

Label propagation, 586
Lagrange multipliers, 122, 124, 940–941
Lagrangian duality, 123, 943–945
Laplacian eigenmaps, 353
Laplacian matrix, 355
Laplacian regularized kernel least squares, 585
Latent semantics indexing, 337
LBG algorithm, 751
Leaky learning algorithm, 783
Learning machines, perceptrons, 101, 114
Learning subspace methods, 333
Learning theory, Vapnik-Chervonenkis, 231, 300–309
Least squares methods (LSM), 103
  algorithm, 105–107
  cost function, 172
  mean square error estimation, 103–105, 110–117
  stochastic approximation, 105–108
  sum of error squares estimation, 108–110
Leave-one-out method, 570
Levenberg–Marquardt (L-M) method, 717
Levenberg–Marquardt algorithm, 171, 717
Levenstein distance, 490
Levinson's algorithm, 430
Lexicographic ordering, 324
Lifetime of a cluster, 690
Likelihood function, 24, 35
Likelihood ratio, 18
Likelihood ratio test, 57
Linear classifiers, 91
  functions and decision hyperplanes, 91–92
  generalized, 185–187
  least squares methods, 103–110
  logistic discrimination, 117–119
  mean square estimation, 110–117
  perceptron algorithm, 93–103
  support vector machines, 119–133
Linear dichotomies, 187
Linear discriminant analysis (LDA), 33, 288
Linear discrimination, 288, 294
Linear time invariant (LTI) systems, 946–947
Linear transforms, *see* feature generation, linear transforms and
Link graph, 686
Lloyd's algorithm, 751

LMS algorithm, 105
Local constraints, 484, 495
Local feature extractor, 421
Local linear embedding, 357
Local linear transforms for texture extraction, 421–423
Log-likelihood function, 36–37, 46, 118
Logarithmic search, 501–502
Logistic discrimination, 117
Logistic function, 162
Long run emphasis, 420
Loss matrix, 17
Loss minimization algorithm, 101

## M

Machine intelligence systems, 1
Machine vision, 1
MAFIA algorithm, 831
Mahalanobis distance, 32
Majority Voting Rule, 225
Manhattan norm, 605
Manhattan segmental distance, 833
Many-to-one mapping, 45
Marginalization, 916
Markov chain models, 522–523, *see also* hidden Markov models
Markov chain Monte Carlo (MCMC) techniques, 40
Markov edit distance, 490
Markov model, 522
Markov random fields, 434, 554
Masks, 422–423, 471–472
Matching score generator, 643
Matrix theory, 658–664
Matrix updating algorithmic scheme (MUAS), 658, 659
Max proximity function, 616, 620
Maximally complete subgraph, 668
Maximally connected subgraph, 668
Maximum a *posteriori* probability estimation, 38
Maximum entropy estimation, 43
Maximum likelihood, 34
Maximum margin classifiers, 308
Maximum variance unfolding, 361
Maxmin algorithm, 638
MaxNet, 643, 644
McCulloch-Pitts neuron, 101, 162
Mean center, 617
Mean proximity function, 621
Mean square error (MSE)
  bias-variance dilemma, 114–117
  estimation, 103–104
  Karhunen-Loève transform, 326–329
  multiclass generalization, 104–105
  posterior class probabilities, 112–113
  regression, 110–111
Mean square error estimation, 103
Mean square error regression, 110
Mean value, 106, 182, 266
Mean vector/point, 617
Median center, 618
Median clustering, 843
Mel Cepstrum, 457
Mellin transforms, 500
Membership functions, 601
Mercer's theorem, 199
Merging procedure, 641
Metric multidimensional scaling (MDS), 332
Min proximity function, 616, 620
Minimum cut bisector, 687
Minimum cut set, 687
Minimum Description Length (MDL), 887
Minimum distance classifier, 30
Minimum spanning tree (MST), 766, 882, 884, 900, 902
Minimum variance algorithm, 662, *see also* ward's algorithm
Minkowski distance, 716
Min-max duality, 942–943
Mixture decomposition, 702
Mixture decomposition schemes, 703–712
Mixture models, 44–45
  expectation maximization (EM) algorithm and, 45–49
Mixture of Experts, 230
Mixture scatter matrix, 281
Model sparsification, 205
Mode-seeking algorithms, 737, 772
Mode-seeking property, 737
Modified BSAS, 637
Modified fuzzy C quadric shells (MFCQS) algorithm, 726, 727
Modified Hubert $\Gamma$ statistic, 880
Moment-based features, 441–442
Moment generating function, 917
Moments, 423–427
  absolute, 414
  central, 413, 424
  geometric, 423–425
  of Hu, 424
  Zernike, 425–427
Momentum factor, 170
Monothetic algorithms, 682
Monotonicity, 496, 665
Monte Carlo techniques, 865
Moore machine model, 535
Morphological operations, 791–794
  closing, 793
  dilation, 791

erosion, 791
opening, 793
translation, 791
Morphological transformation techniques, algorithms based on, 631
Motion compensation, 498
Motion estimation, 498
Mountain method, 837
Multiclass generalization, 104–105
Multi-class tasks, 127
Multifractional Brownian motion, 451
Multiple additive regression trees (MART), 236
Multiresolution analysis, 386
Multiresolution decomposition, 375
Multiresolution interpretation, 384
Multispectral remote sensing, 7
Multivariate Gaussian, 920–921
Mutation, 810, 811
Mutual information, 545
ICA based on, 345–348

## N

Naive-Bayes classifier, 59–61
Natural clusters, 600
Natural gradient, 347
Nearest Neighbor, 901–902
*k* Nearest Neighbor (*k*NN) density estimation, 56
Nearest neighbor condition, 750
Nearest neighbor distances tests, 901
Nearest neighbor rule, 61–64, 161
Nested clusterings, 653
Nesting effect, 286
Network size selection, 176–181
Networks, 183
Neural networks, 552
basic sequential algorithmic scheme and, 644–646
feature selection and, 298–299
training Markov models, 552–553
Neurocomputers, 169
Neurons, *see* perceptrons
Newton's algorithm, 934
Neyman-Pearson decision rule, 18
Neyman–Scott procedure, 898
Noble identity, 372, 380
Node connectivity, 670
Node degree, 670
Node impurity, 218, 219, 220
Nonlinear classifiers
algorithms based on classification of training set, 160–162
approximation properties, 194–196
backpropagation algorithm, 162–169
combining, 222–235
cost function, selecting, 172–176
decision trees, 215–222
linear dichotomies, 187–189
polynomial classifiers, 189–190
radial basis functions, 190–194
support vector machines, 198–203
three-layer perceptrons, 158–160
two-layer perceptrons, 153–158
XOR (exclusive OR) problem, 151–153
Nonlinear optimization iterative techniques, 44
Nonnegative matrix factorization (NMF), 349
Nonparametric estimation, 49–59
Normal distributions, Bayesian classification for, 20–33
Normalization, 790
data, 263
Normalized central moments, 424
Normalized radial distance, 721–723, *see also* distance between a point and a quadratic surface
Normalized Γ statistic, 870
Null hypothesis, 268, 864, 866

## O

Octave-band filter banks, 382
One-dimensional DFT, 364
One-tailed statistical test, 865
left-tailed statistical test, 866
right-tailed statistical test, 866
Online mode, 168
Opening, 793
Optical character recognition (OCR), 435, 789
OPTICS, 818
OptiGrid, 801
Optimization, 935
ORCLUS algorithm, 834
Ordinal proximity matrices, 894
Ordinary binary classification trees (OBCTs), 216
Orientation, 442
Orthogonal projection, 104
Orthogonality condition, 104
Outliers, 262, 710, 732, 734
Output layer, 156
Overtraining, 177

## P

PAC bounds, 301
PAC learning theory, 301
Packing density, 899
PAM algorithm, 746
Parallel connection, 948

Parametric models, 427
one-dimensional, 428–431
two-dimensional, 431–433
Paraunitary, 379
Partial clusterings, 804
Partition algorithmic schemes, 737
Partition coefficient (PC), 888, 891
Partition density, 891–893
Partition density (PD) index, 891
Partition entropy (PE), 888, 891
Partition entropy coefficient, 888
Partition shell density, 892
Parzen windows, 51, 802
Pattern mode algorithms, 780
Pattern recognition
importance of, 1–4
supervised versus unsupervised, 7–9
syntactic, 12
Peaking phenomenon, 265
Pearson's correlation coefficient, 606
Penalty terms, 180
Perceptron(s), 100
description of, 100–101
three-layer, 158–160
two-layer, 153–158
Perceptron algorithm, 93
convergence proof, 95–97
pocket algorithm, 101
variants of, 98
Perceptron cost, 93
Perfect reconstruction, 378
Perpendicular Distance, 720, *see also* distance between a point and a quadratic surface
pFCM, 717
Pitch, 461
signal, 461, 462
Pocket algorithm, 101
Point representatives, 617–619, 716–718, 888
Point spread function, 948
Point-based subspace clustering algorithms (PBSCA), 832
Poisson distribution, 899
Poisson process, 899
Polyhedra, 157–158
Polynomial classifiers, 189–190
kernels, 199
Polythetic algorithms, 682
Positive definite/symmetric matrices, 927–928
Possibilistic algorithms, 630, 738
Possibilistic clustering algorithms, 630, 733–739
Potential functions, 51
Power function, 864
Precision, 573
Prediction based on groups, 599
Prim's algorithm, 676
Principal Component Analysis (PCA), *see* Karhunen-Loève transform
Probabilistic clustering algorithms, 630
Probabilistic neural networks, 58, 196
Probability chain rule, 64
Probability density functions (pdf), 13
Edgeworth expansion, 346, 918
estimating unknown, 34–61
mixture models, 44–47
nonparametric estimation, 49–59
normal density function, 20–33, 920–921
PROCLUS algorithm, 832
Projection pursuit, 186
Proximity dendrogram, 656
Proximity function
between a point and a set
average proximity function, 616
maximum proximity function, 616
minimum proximity function, 616
between two sets
average proximity function, 621
maximum proximity function, 620
mean proximity function, 621
minimum proximity function, 620
Proximity graph, 669, 894
dissimilarity graph, 669
Proximity matrix, 655
Proximity measure, 597
between a point and a set, 616–620
between two points, 604–616
between two sets, 620–622
definitions, 602–604
Pruning techniques, 178
Pseudolikelihood function, 556
Psychoacoustics, 461
Pythagoras' theorem, 333

## Q

QROCK, 686
Quadrat analysis, 900
Quadratic classifier, 25
Quadratic discriminant analysis (QDA), 33
Quadric surface, 719
hyperellipses, 719
hyperparabolas, 719
representatives, 718–728, 731–732
Quasistationary signals, 452
Quefrency, 456
Query point method (QPM), 513
Quickprop, 171

## R

Radial basis functions (RBFs), 190–194
  kernels, 201
Radial distance, 720, *see also* distance between a point and a quadratic surface
Rand statistic, 869
Random fields, 428
  Gibbs, 554
  Markov, 434, 554–556
Random Forests, 221
Random graph hypothesis, 867
Random label hypothesis, 868
Random position hypothesis, 867
Random projections method, 823
Random variables, 5
Random walk, 446
Randomness hypothesis, 866, 896
Ratio-scaled proximity matrices, 895
Reassignment procedure, 642
Recall, 573
Receiver operating characteristics (ROC) curve, 275
Receptive field, 184
Recognition, hidden Markov model, 536–539
Reduced Convex hull (RCH), 138
Redundancy reduction, 342
Regions of influence, 768
  algorithms based on, 768–770
Regression, mean square error, 110–111
Regularity assumption, 935
Regularity hypothesis, 896
Regularization parameter, 180
Relative closeness, 688
Relative criteria, 864, 877–893
Relative edge consistency, 769
Relative interconnectivity, 687–688
Relative neighborhood graph (RNG), 770, 882, 884
Relevance Feedback (RF), 509
Relevance vector machines, 208
Representatives
  hyperplane, 728
  point, 716–717
  quadric surfaces, 718
Reproducing kernel Hilbert space (RKHS), 199
Reproduction, 810
Reproduction set, 750
Reproduction vector, 749
Resubstitution method, 570
Reward and punishment schemes, 98
Ridge regression, 214
Right-tailed statistical test, 866
Risk or loss, 17
Risk, minimizing average, 16–18
Robbins–Monro iteration, 783
Robust statistics, 211
ROCK algorithm, 685
Rotation forest, 230
Roundness ratio, 442
Run length nonuniformity, 420
Run percentage, 421

## S

Sakoe and Chiba, 497
Saliency, 179
Sample complexity, 302
Sample correlation matrix, 108
Sample mean, 270
Sampling frame, 897
Sampling window, 897
Scalar feature selection, 283–284
Scan test, 900
Scatter matrices, 280
Searching techniques
  optimal, 287–288
  suboptimal, 285–287
Second characteristic function, 917
Second moment structure, 900
Second-order statistics features, 414–417
Segment modeling, 551
Segmental k-means training algorithm, 542
Self-affine, 448
Self-Organizing Maps (SOM), 786
Self-similarity, 444
Self-training, 588
Semi-supervised learning, 577
Sequential backward selection, 285
Sequential clustering algorithms, *see* clustering algorithms, sequential
Sequential decomposition, 902
Sequential forward selection, 286
Sequential Minimal Optimization (SMO), 129
Sequential search, 503
Serial connection, 948
Shannon's information theory, 43, 218
Shape and size characterization, 435
  chain codes, 439–441
  Fourier features, 436–439
  geometric features, 442–443
  moment-based features, 441–442
Shell hypervolume, 892
Shell partition density, 892
Shell-shaped clusters, 892–893
Short run emphasis, 420
Short-time autocorrelation, 455
Short-time DFT, 454
Short-time Fourier transform, 386

Short time processing of signals, 452–455
Sigmoid functions, 162
Significance level, 269
Silhouette index, 885
Similarity matrix, 655
Similarity measure (SM)
  between discrete-valued vectors, 607–608
  between real-valued vectors, 606
  defined, 602
Simple sequential inhibition (SSI), 899
Simulated annealing, 807–808
Sine transforms, 366
Single link algorithm, 659, 670–671
Singular value decomposition, 335, 822
Skewness, 413
Slack variables, 125
Small sample size problem, 295
Soft margin classifiers, 133
Softmax activation, 241
Softmax activation function, 174
Spanning tree, 675
Sparse decomposition technique, 902–905
Sparsification, *see* model sparsification
Spatial dependence matrix, *see* co-occurrence matrices
Spatial randomness, tests for, 900–905
Speaker-dependent recognition, 491
Speaker-independent recognition, 491
Spectral clustering, 772
Spectral features, 460–462
Spectral representation, 335
Speech recognition, 2–3, 451, 789
  dynamic time warping in, 491–498
Squashing functions, 162
Stacking, 226
State duration HMM, 545
Statistical hypothesis testing, 268–275
  one-tailed, 865
  two-tailed, 865
Statistical independence, 916
Statistically self-affine, 447
Step function, 154
Stirling numbers of the second kind, 628
Stochastic approximation, 105
Stochastic relaxation methods, 630–631
String patterns, 482
Structural graph tests, 900
Structural risk minimization, 303
Structuring element, 791–794
Subgraph, 668
  complete, 668
  connected, 668
Subsampling, 371
Subspace classification, 333
Subspace clustering algorithms (SCAs), 631, 824
  grid-based, 825–832
  point-based, 832–836
Sum of error squares estimation, 108
Support vector machines (SVMs)
  geometric viewpoint, 136–138
  nonlinear, 198–203
  nonseparable classes, 124–127
  reduced convex hulls, 138–142
  separable classes, 119–124
  Vapnik-Chervonenkis learning theory, 301–303
  v-SVM, 133–136
Support vectors, 122
Surface density criterion, 895
SVMs, *see* support vector machines (SVMs)
Sweep, 807
Synapses, 101
Synthesis filters, 377
System evaluation, 6, 567
  error counting, 568–569
  finite size of data set, 569–573
  medical imaging example, 573–577

## T

Tabu search method, 837
Tanimoto measure/distance, 606–607
Tanimoto measure for discrete valued, 609
Template matching, 481
  Bellman's principle, 484–487
  defined, 481
  deformable models, 504–508
  dynamic programming, 484–487
  dynamic time warping in speech recognition, 491–498
  edit distance, 487–491
  measures based on correlations, 498–504
  measures based on optimal path searching, 482–498
Templates, 481
Test statistic, 269
Texture, 393, 412
Texture characterization, 393, 412–421
Texture classification, 393–396
Thermal perceptron algorithm, 101
Three-layer perceptrons, 158–160
Threshold dendrogram, 674
Threshold graph, 668, 904
Tied-mixture densities, 544
Tiling algorithm, 160
Time domain features, 462–463
Toeplitz matrix, 430, 433
Topological sorting, 66
Total fuzzy average shell thickness, 892

Total probability, 915
Total variation, 889
Touching clusters, 767–768
Training Markov models, 552–553
Training patterns/feature vectors, 6
Transductive inference, 577
Transductive SVM, 586
Transfer function, 947
Transformation invariance, 183
Transformed divergence, 278
Tree-structured filter bank, 374
Triangular inequality, 603
Truth tables, 152, 154
Two-dimensional AR models, 431
Two-dimensional DFT, 366
Two-dimensional parametric models, 431–433
Two-tailed statistical test, 865
Two-threshold sequential algorithmic scheme (TTSAS), 639–640

## U

Uncertainty principle, 395
Undirected graphs, 667
Universal approximators, 194
Universum, 590
Unsupervised learning, 595
Unsupervised pattern recognition, 7
Unweighted graphs, 667
Unweighted pair group method average (UPGMA), 661
Unweighted pair group method centroid (UPGMC), 661
Utility function, 843

## V

Validation of clustering results, 597
Valley-seeking algorithms, 631, 801–803
Vapnik–Chervonenkis learning theory, 299
Variables, slack, 125
Variance, 916
Variance dilemma, bias-, 114–117, 228
Variational similarity, 487
Vector quantization, 749
Vector quantizer, 749
  decoder, 750
  encoder, 750
Vectors, 4
  basis, 324–326
  discrete-valued, 607–610
  mixed-valued, 610–612
  noisy feature, 710
  quantization, 749–751, 788
  real-valued, 604–607
  selecting, 284–288
  support, 122
Video coding, 481
Viterbi algorithm, 523
Viterbi reestimation, 542
Voronoi tessellation, 64

## W

Ward's/minimum variance algorithm, 662–663
WaveCluster algorithm, 838
Wavelet packets, 387
Web content mining, 632
Web data mining, 632, 838
Web mining, 590
Web usage mining, 632
Weierstrass theorem, 195
Weight elimination, 180
Weight sharing, 183–184
Weighted graphs, 667
Weighted pair group method average (WPGMA), 661
Weighted pair group method centroid (WPGMC), 661
Well-formed functions, 172
Widrow-Hoff algorithm, 107
Within-class scatter matrix, 280
Within scatter matrix, 741
Wolfe dual representation form, 123, 200, 214, 812, 944–945
Wrapper approach, 285
Wrapper model, 823

## X

Xie–Beni (XB) index, 890
XOR problem, 151

## Z

Zernike moments, 425
Zero-crossing rate, 462