

SUBJECT INDEX**A**

a priori algorithm, 461, 467, 468, 474, 478, 479, 480
abrupt pattern change, 256
absent value, 261
absorbed state, 60
accuracy, 395, 397
address components, 669
adjusted difference, 478
agglomerative hierarchical clustering, 702
algebraic lattice, 95, 11
approximate dependency, 484, 491
association rule(s), 32, 82, 94, 396, 575, 460
attractiveness of marker, 276
attribute evaluation, 329
attribute-value pair, 367, 368
automatch, 668

B

backward induction, 90
backward reasoning, 100
backward selection, 230
backward strategy, 236
base rate, 681
Bayesian approach to record linkage, 662, 665, 667, 669, 678-685
beam search, 61
bigmatch, 675, 676
binomial distribution, 664
blocking criteria, 675, 676
Boolean association rule, 461, 462, 466, 467, 471
border vectors, 158
boundary problem, 464, 469

branch-and-bound algorithm, 59, 28
breast cancer diagnosis, 156
Brooks-Gelman-Rubin (BGR) ratio, 670
b-scan, 588

C

C4.5, 329, 332, 336, 347, 350
canons of induction, 91, 104
cardinality, 577
certain rule, 387, 389
citeseer digital library, 698
clash, 263
clash condition, 263
class value, 256, 267
classification accuracy, 292, 3
classification quality, 287, 288
clause, 195, 2
cleansing, 564
clique cover, 75, 76, 77, 78
closed-loop methods, 280
clustering, 283, 566, 461, 468, 470, 486, 487, 488, 490, 491, 492
coarse-grained parallel GAs, 445
cobraMED, 563
common sense reasoning, 94, 95
comparison vector/Feature vector/Bit-pattern, 661, 669, 672, 674, 676
compatibility, 98, 1
complete disjunct, 32
comprehensibility, 397
computational complexity, 119, 121, 475, 488
concentration operator, 465
concept learning, 657

- condition attribute, 380, 381
 - conditional probability, 463, 475, 481
 - confidence, 285
 - confidence level, 4, 367, 385, 387
 - conjunctive matrix, 9
 - conjunctive normal form, 197
 - conjunctive normal form (CNF), 49, 52, 197, 603, 624
 - consistency, 236
 - context-based spelling error, 597, 601, 616
 - conversion, 258
 - correlation, 238
 - correspondences of Galois, 108
 - covariance, 209, 21
 - CPRI, 563
 - critical interval, 274
 - cropping, 565, 576
 - cross-correlation, 230
 - cross-entropy, 234
 - crossover, 400, 401, 413
 - cross-validation, 287, 289, 297, 419
 - crowding, 443
 - current population survey, 668
 - cutpoint method, 256
- D**
- data aggregation, 555, 562
 - data binarization, 54
 - data, Boolean, 575
 - data, categorical, 575
 - data evolution, 289
 - data farming, 280
 - data farming methods, 281
 - data farming process, 298
 - data fusion , 552, 582, 585
 - data matrix, 9
 - data mining, 2, 12, 48, 280, 306, 368, 369, 389, 396, 434, 460, 461, 497, 552, 597, 631, 696
 - data mining process, 553
 - data preprocessing, 555, 563, 585
 - data semistructured, 556
 - data structured, 556
 - data transformation, 282
 - data transformation, 552, 567
 - data unstructured, 556, 579
 - data visualization, 700
 - database modeling, 462, 483
 - Davies-Bouldin index, 566
 - Dconfidence, 461, 462, 463
 - DDSM, 567
 - decision attribute, 380, 381
 - decision forest , 333
 - decision function, 2
 - decision making, 523
 - decision rule, 360, 366, 377, 385, 656, 658-662, 672, 675, 681, 685
 - decision support system, 367, 377
 - decision tree, 51, 306, 307, 360, 363, 382, 385, 403, 435, 496
 - decomposition of good classification tests inferring, 122, 129
 - decreasing monotone Boolean functions, 152
 - defuzzification, 490
 - degenerate ternary matrix, 23
 - degree of confidence, 461, 467
 - degree of fuzziness, 20
 - degree of implication, 482
 - degree of satisfaction, 484, 485
 - degree of support, 461, 467, 473
 - Densest k-subgraph, 238
 - dependence, 230
 - dependence modeling, 395, 396, 397, 404, 415, 416, 417, 418, 429
 - diagnostic rule (operation), 98, 101, 103, 113, 114, 133
 - DICOM, 563
 - difference value, 272
 - digital mammogram, 565

dilation operator, 465
 Dimp, 482, 483
 Dirichlet distribution, 670
 discrete support vector machine,
 308, 313, 314, 324
 discretization, 258, 283
 discriminant analysis, 306
 disjunct, 9
 disjunctive normal form (DNF),
 2, 4, 10, 49, 52, 197, 199,
 256, 262, 603, 624
 disjunctive ternary matrix, 9
 distance, 230
 distributed algorithm, 462
 distributed data, 560
 distributed data mining, 703
 distributed population, 405, 412,
 413, 429
 diversity-preserving mechanisms,
 433
 domain-specific knowledge , 555
 dstrength, 477, 478
 Dsupport, 461, 462, 463
 DTW, 486, 487, 490
 dual lattice, 111, 112
 dual objects, 110
 duality of a good diagnostic test,
 109
 dynamic time warping, 486

E

ECG, 571
 efficiency, 462, 479, 483, 488
 elitism, 439
 EMG, 571
 empirical data, 359, 360, 372, 386
 entropy, 232, 283, 330, 332, 336,
 337, 342, 344, 365, 384, 478
 entropy-based approach, 258
 error control, 195, 196, 207
 essential example, 114, 127, 128,
 129, 130, 131

essential value, 113, 114, 127,
 128, 134
 evolutionary algorithms, 436,
 437, 438, 440, 441, 444
 evolutionary computation, 436,
 438, 440, 446, 447, 448
 expectation-maximization
 algorithm, 666
 expert knowledge, 656, 677
 expert's assertions, 97, 99
 expert's rules, 1, 97

F

false link and false non-link rate,
 360, 361, 378, 659, 685
 fault diagnosis, 228
 feature definition, 297
 feature evaluation, 282
 feature extraction, 287, 579
 feature functions, 227
 feature selection, 228
 feature selection methods, 280
 feature transformation, 286
 Fellegi-Sunter model, 658-660
 filter, 474, 475, 477
 filter methods, 232
 FIO, 481, 482, 483
 fitness, 400, 401, 410, 412
 fitness sharing, 441, 442, 443, 448
 forward induction, 90
 forward reasoning, 100
 forward selection, 230
 forward strategy, 236
 frequent itemset, 479, 48
 function approximation, 521
 functional dependency, 460, 462,
 483, 485, 459
 functional dependency with
 degree of satisfaction, 484,
 485
 future composition, 569
 fuzzification, 489

fuzziness, 460, 463, 464, 466, 469, 483, 488, 489
 fuzzy aggregation, 583
 fuzzy aggregation operators, 489
 fuzzy applications, 521
 fuzzy association, 460, 465, 466, 470
 fuzzy association rule, 460, 465, 466, 470
 fuzzy clustering, 631, 632, 638, 642, 646, 647
 fuzzy clustering, 496
 fuzzy C-means, 629, 630, 634, 635, 637, 638, 641, 644, 647
 fuzzy control, 522
 fuzzy data mining, 464, 489, 491, 492
 fuzzy decision tree, 517
 fuzzy functional dependency, 484
 fuzzy ID3, 518
 fuzzy inference, 489
 fuzzy logic, 629, 631, 632, 635, 638, 643
 fuzzy logic, 460, 463, 464, 483, 484, 492
 fuzzy membership functions, 488
 fuzzy modeling, 485
 fuzzy models, 486
 fuzzy neural networks, 508, 480, 493
 fuzzy relation, 359, 360, 361, 367, 370, 373, 375, 376, 377, 378
 fuzzy rules, 629, 630, 635, 638, 645, 646
 fuzzy set, 464, 465, 469, 470, 472, 473, 474, 490, 491, 493
 fuzzy taxonomies, 464, 466, 470, 471, 472, 473, 474

G

Gabriel graph, 293
 gain, 258

generalization, 434, 446, 447, 448, 453
 generalization operations, 108, 109, 130
 generalization rule, 112, 113, 115
 generalized Association Rule, 360, 462, 464, 470, 471, 474, 489, 492, 493
 genetic algorithm, 395, 402, 429
 genetic algorithms, 499
 Gini index, 283
 global search, 404
 good diagnostic (classification) test, 94, 95, 103, 104, 105, 107, 108, 109
 good irredundant test (GIT), 105, 106, 107, 110, 111
 good maximally redundant test (GMRT), 106, 107, 109
 goodness index, 470, 491
 granularity, 468
 graph connected components, 77
 GRASP approach, 64
 greedy, 229
 greedy random adaptive search procedure (see GRASP), 64
 greedy rule induction, 395, 400, 402, 404, 416, 429
 grid computing, 703
 grid partitioning, 490
 guided learning, 70

H

Hansel algorithm, 160, 175
 health care, 551, 559
 heaviest k-subgraph, 238
 hedge operator, 465
 hierarchical decomposition of variables, 155, 464
 hierarchy, 590
 HIPAA, 590
 histogram equalization, 565

history text, 599, 601, 602, 608,
610, 615
human thinking, 91
hundred-percent census unedited
file (HCUF), 668
hybrid fuzzy modeling, 517

I

ID3 , 329, 332, 344, 347, 350
identifiability (statistical), 662
if-then rules, 576, 577, 587
image transformation, 98, 1
implication, 460, 464, 466
implicative dependency, 104
implicative regularity, 5
implicative rules, 103
implicit fitness sharing, 443, 448,
450, 452, 453
imprecise data, 361
imprecision, 463, 464
inapplicable value, 261
incomplete data, 360, 367, 368,
369
inconsistency, 360, 381, 383
increasing monotone Boolean
functions, 152
incremental algorithm, 129, 132
incremental learning algorithm,
90, 132
incremental learning procedure,
136
incremental learning task, 90
independent component analysis
(ICA), 571
indiscernibility, 376, 378, 380
induction of decision trees, 328,
329, 536
inductive hypothesis, 91
inductive inference, 103
inductive learning, 360, 363, 364,
382, 383, 389
inductive rules, 91, 111

inference, 632, 635, 639
inference of monotone Boolean
functions, 165
inference of nested monotone
Boolean functions , 166
information gain (infogain), 330,
331, 335, 336, 337, 344, 407,
409
information loss, 469
information quantum, 10
information retrieval, 629, 631,
641
information system, 367, 368,
379, 380, 389
information theory, 462, 478
instances, 96
INT method, 493
integer program, 227
integer programming, 238
intelligent system, 360, 361, 362,
363, 371, 374, 377
interdiction, 98, 1
interestingness, 395, 397, 398,
399
interestingness measure, 462, 466,
475, 477, 491
interior item, 471, 472
irredundant collection of values,
105, 112
irrelevant value, 261
isomorphic posets, 158
iterative rule learning, 499

K

Karnaugh maps, 50
knowledge acquisition, 359, 360,
361, 362, 363, 364, 377
knowledge base, 98
knowledge discovery, 460, 461,
488, 489, 490, 491, 492, 493
knowledge matrix, 10
knowledge nuggets, 398, 418

knowledge representation, 362
 knowledge transformation, 290,
 293

L

label sequence, 257
 latent class assumption, 665, 667
 latent semantic indexing (LSI),
 582
 lattice chains, 110
 lattice theory, 94
 leaf item, 471, 472
 learning classifier systems, 292
 learning from examples, 363, 368
 Leibniz system, 205
 length of sequence, 269
 lightest k-subgraph, 238
 linguistic description, 386
 linguistic hedge, 465, 466, 473,
 474, 490, 493
 linguistic models, 486
 linguistic term, 464, 469, 474
 linguistic variable, 464
 link function, 404, 416, 663
 literal, 262
 local search, 194, 197
 logic data, 255, 261
 logic data, 242
 logic data miner, 247
 logic programming, 193, 205
 logic record, 218
 logic variable, 197, 260, 368, 603
 logic variable, 197
 logistic regression, 657, 661-665,
 672, 675, 681, 685
 log-linear model, 668
 lower approximation, 158, 375,
 379, 381, 384
 lower unit, 361, 362, 363, 364,
 377
 Ltest, 597, 599, 601, 602, 608,
 616, 625

M

machine learning, 94
 mapping, 465, 481, 482
 margin of separation, 311, 312,
 320, 322
 marker, 256
 Markov Chain Monte Carlo
 (MCMC), 662, 669, 671-674
 master instance, 205
 match weight, 659, 660, 665
 matching fields and records vs.
 linking records, 658
 maximally redundant collection of
 values, 105, 106, 107
 maximum likelihood, 171, 180
 mean, 209
 medical images, 557, 565
 medical signals, 571, 464, 465,
 472, 477, 482
 membership degree, 373, 374
 membership function, 464, 465,
 469, 470, 474
 merge, 656
 michigan approach, 435
 migration, 413, 414, 415, 419
 minimal confidence, 467
 minimal support, 467
 minimization of Boolean
 functions, 50
 minimum cost satisfiability
 (MINSAT), 51, 193, 197, 205
 minimum description length
 (MDL) , 331, 332
 minimum splits, 259
 mining process, 462, 469, 475,
 476, 479, 483, 488
 misclassification distance, 308,
 309, 312, 320, 322
 misclassification rate, 308, 309,
 312, 313, 320
 misclassification rates, 665, 681
 modality, 552, 583
 model of common reasoning, 96

modified item, 474
 monotone Boolean functions, 150, 157
 monotone phenomena, 149
 monotonicity property, 150
 m-probability, 660, 661
 multiattribute decision trees, 334
 multimedia patterns, 557, 561, 587
 multimodality, 583
 multivariate split, 307, 308, 313, 318
 mutation, 400, 402, 410, 411
 mutual information, 331, 337

N

naive Bayes decision algorithm, 259
 national health interview survey, 670
 natural language, 470, 487
 natural selection, 435, 437
 nested monotone Boolean functions, 153, 154
 neural fuzzy systems, 509
 neural network(s), 47, 306, 360, 378, 432, 434, 500, 507
 niching, 442, 443
 noise, 460, 484, 487
 nominal value, 255
 non-incremental algorithm, 115
 non-match weight, 659, 660, 661

O

oblique splits, 318, 323
 one clause at a time (OCAT) approach, 45, 54, 58
 on-line human reasoning, 92
 open-loop methods, 280
 optical character recognition (OCR), 703

optimal decision rule, 656
 optimal decision trees, 338
 optimization strategy, 475

P

PAC learning, 49
 parallel algorithm, 462, 489, 492
 PART, 329, 347, 350
 partial degree, 460, 464, 481, 483, 487
 partitioning, 460, 462, 466, 468, 469, 486, 487, 488, 491
 parts-of-speech, 597, 602, 621
 patient record, 551, 556, 561
 patient record standards, 560
 pattern association, 460, 461, 462, 463, 465, 466, 485, 486, 487, 488
 pattern classification, 657
 pattern matching, 466, 485
 Pittsburgh approach, 435, 436
 pitt-style, 436, 440, 441, 447, 448, 450, 452, 453
 Pitt-style evolutionary classifier systems, 433, 452
 plausibility of a hypothesis, 13
 p-median, 295
 p-median model, 294
 posets, 157
 possible rule, 385, 387, 389
 posterior distribution, 682, 683, 686-690
 posterior probability, 676, 678, 686-690
 potential of marker, 276
 principal component analysis (PCA), 571
 prior distribution, 670-674, 671, 672, 677, 679, 681, 685
 probability, 463, 475, 476, 481, 482, 488
 probability distribution, 196, 208

probability theory, 370, 371, 372
 problem decomposition, 77
 pruning of decision trees, 340,
 341

Q

qualified association rule, 462
 query complexity, 175

R

random variable, 208, 217
 rare events logit, 685
 rational data, 255, 261
 reasoning process, 102
 record, 261
 record and field, 658
 reducing subtasks, 125
 redundant collection of values,
 105
 rejectability graph, 72, 74, 79
 relational database, 462, 467, 484,
 489, 490, 492
 relevance count, 274
 resolvent, 35
 reverse engineering, 462
 rough set, 359, 361, 367, 375
 rule discovery, 639, 646, 647
 rule discovery, 360, 389
 rule induction, 259, 395, 399,
 402, 403, 429, 430, 431, 629
 rule of alternative, 98, 1
 rule strength, 285
 rule support, 285
 rules of the first type, 96, 97, 100,
 103, 105, 106
 rules of the second type, 96
 rule-structuring, 290

S

sampling, 462

sampling population , 12
 satisfiability (SAT) problem, 50
 satisfiable CNF, 23
 scheme, 471, 474, 486, 490
 scout operator, 470, 473
 screen effect, 19
 scrubbing, 564
 sectional Boolean vector, 10
 selection, 400, 405, 410, 412, 415,
 430
 semantic, 461, 462, 463
 semantic networks., 299
 semantic patterns, 575
 separating formula, 263
 separating hyperplane, 308, 311,
 313, 314, 315
 separating vector, 194
 sequent , 9
 set data, 255, 258, 261, 264
 sharp boundary, 469
 similarity, 463, 464, 484, 485,
 488
 simple association rule, 462, 483,
 489
 small disjuncts, 399, 43
 SNR, 589
 SODI , 329, 335, 339, 344, 347,
 350
 solution algorithm, 195
 SOM, 566
 sparsest k-subgraph, 238
 special reasoning operations, 112
 speckle model, 587
 speech recognition, 486
 spell checker, 597, 598, 605, 608
 Sperner system, 109, 110, 125
 statistical distribution, 589
 statistical properties of decision
 rule, 656
 statistics theory, 478
 strategies, 96
 strength of membership, 266
 structural risk minimization, 310
 subclass, 464

subtask of the first kind, 122, 127,
129, 132, 133, 134
subtask of the second type, 124,
127, 129, 133, 134
super-class, 464, 472, 476
support, 4
support vector machines, 308,
310, 313, 322, 324
syllogism, 91
syllogistic regularity, 15
syntax checker, 597, 599, 608,
611, 614
synthetic experiment, 475
system consumption, 475

T

target subset, 657-658
terminal state, 60
testing text, 597, 599, 600, 605,
608, 613, 615
text mining, 700
text structuring standards, 574
text transformation, 573
threshold, 467, 474, 476, 477,
479, 480, 484
threshold (“shift”), 658, 660, 663,
665, 675, 681
time complexity, 121, 122
time series, 460, 462, 463, 478,
485, 486, 487, 488, 489, 490,
491
time series prediction, 523
training data, 363, 364, 365, 366,
367, 369, 382, 383, 389
training text, 599, 601, 602, 609,
624
transaction, 462, 466, 467, 475,
484, 489, 491, 492
transformation, 258
transformation of set data, 264
transformation to logic data, 255
TSK models, 486

tuple, 467, 468, 471, 485, 586

U

ultrasound images, 359, 360, 361
unavailable value, 261
uncertainty, 463, 466, 483, 487,
488
undecided value, 262
unknown attribute values, 368
unsupervised generalization, 135
unsupervised learning, 655, 657,
658
upper approximation, 367, 375,
379, 381, 383
upper zero, 158
u-probability, 659, 66
user profiles, 631, 632, 640

V

Vapnik-Chernovenkis dimension,
50
variance, 209
variance in construction of
decision trees, 333
veto function, 23
visual data mining, 703
visualizing data mining models,
701
Voronoi diagram, 293
vote distribution, 193
vote-counts, 208, 209
vote-total, 193, 195, 196

W

web context mining, 697
web mining, 696
web structure mining, 699
web usage mining, 698
weight of evidence, 478
win-BUGS, 669, 675

WM method, 490

world model, 5

wrapper methods, 234