

SUBJECT INDEX**A**

absorbed state, 57-58
accuracy illusions, 243-244
accuracy rate, 6, 9, 21, 111
acyclic graph, 205
adenoma, 304
adenosis, 304
Agrawal, Imielinski, and Swami
(AIS) algorithm, 251
altered RA1 algorithm (ARA1),
253-257
American College of Radiology
(ACR) Breast Imaging Lexicon,
312, 324
antecedent part, 250
antichain, 204
application areas, 2
Apriori algorithm, 251, 258-259
AprioriHybrid, 251
AprioriTid, 251
AQ system, 133
artificial mammographic cases,
193
associated findings (in breast
cancer), 303
association rules (negative), 260
association rules, 23, 249-265
attributes, 20

B

back propagation algorithm, 297
back propagation of errors, 181
basic sciences, applications in, 3
battle planning, 3
Bayesian models, 178
beam search, 46, 60
benign case, 185
bias, 21, 115-116

biceps brachii muscle, 290
binarization, of data, 24
binary attributes, 24, 28
binomial test, 284
biopsy (of breast cancer), 177
bipolar sigmoid transfer function,
297
BI-RADS (in breast cancer
diagnosis), 312
Boolean function inference, 19-
54
Boolean function (definition of),
28
border (between classes), 184,
189
border area, 244-245
border vectors (number of), 206
border vectors, 205, 235
borders of monotone Boolean
function, 206
borders, 243
branch-and-bound (B&B), 35,
38-45, 55-61
breast cancer diagnosis, 177-196
breast cancer diagnosis, 240,
301-310, 311-324
business (applications in), 4

C

calcifications density, 302
calcifications irregularity, 302
calcifications, 179, 194, 203
candidate item set, 250
candidate list, 74
cardinality of a set, 24, 37, 46
census data analysis, 252
centroids (negative), 275
centroids (positive), 275

- chain partition, 204
 - chain poset, 233
 - chain, 204
 - characteristic frequencies, 292
 - characteristics, 20
 - circuit design, 22
 - calcifications variation, 302
 - classes, 1
 - classification modeling, 252
 - classification of examples, 106
 - classification rules, 12, 14
 - classification scenarios, 104-105
 - clause inference, 32, 39
 - clause satisfiability problem, 110
 - clauses, 31
 - clinical applications, 179
 - clinical management problem, 240
 - clique based decomposition, 170
 - clique cover, 163-167, 170
 - clique, 160
 - cluster analysis, 178
 - cluster centroids, 274-275
 - clustering, 273
 - coastal erosion, 3
 - common logic approaches, 23
 - compact system (Boolean function), 67
 - comparable vectors, 199
 - complement of example, 151
 - complement (of the R-graph), 162
 - computer-aided diagnosis (CAD), 178
 - concept learning, 182
 - confidence interval, 111
 - confidence level (of association rule), 250
 - conjunctive normal form (CNF), 29, 151-155
 - connected component (of the R-graph), 163, 165-167
 - connected component based decomposition, 170
 - connectionist machine, 23
 - concept learning system (CLS), 133
 - consequent part, 250
 - content descriptors, see *keywords*
 - continuous attributes, 24
 - convex polyhedral, 13, 51
 - corrupted data, 8
 - cost (of classification) 103, 111
 - crime control, 4
 - criteria, 20
 - cyclic directed graph, 204
 - cycling, 255
 - cysts, 304
- D**
- data analysis, 2
 - data mining equation
 - for the future (the), 330
 - data collection, 6
 - data mining of EMG signals, 289-299
 - data mining process, 5
 - data mining (definition of), 1
 - data pre-processing, 8
 - decision trees, 12, 23, 178, 183
 - declassification process (of documents), 267-268
 - decomposition (of problem), 163
 - decomposition structure, 23
 - degree of confidence, 323-324
 - degree of lobularity, 319-324
 - degree of microlobularity, 319-324
 - density of parenchyma, 302
 - dense data, 252
 - depth of undulation, 321-324
 - design of a new product, 4
 - design of new drugs, 3
 - design problems, 241-242
 - diagnosis problem, 241
 - diagnostic classes (in breast cancer), 303

diagnostic classes, 189, 247
 diagnostic problems, 239-241
 diagnostic rules, 301-310
 digital mammography, 311-324
 directed graph, 204
 discriminant analysis, 178
 discriminant functions, 246
 disjunctive normal form (DNF),
 29, 151-155
 DNF clause, 61
 document classification, 270
 document clustering process,
 270-271
 document indexing, 270
 document representation, 270
 document surrogate, 270
 documents (by the Associated
 Press), 277
 documents (of text),
 documents (by the U.S.
 Department of Energy (DOE)),
 277
 documents (by the Wall Street
 Journal), 277
 documents (ZIPPS class), 277
 domain expert, 10
 dominated state, 41
 dual (of LP), 151
 dual problem, 109
 dual problem, 151
 duality relationship, 151-155
 ductal orientation, 179
 ductal orientation, 302
 dynamic programming approach,
 213

E

electrode orientation, 292
 electrodes, 291, 292
 electromyography (EMG), 289
 EMG signals, 292, 293
 engine of car, analysis of, 2

engineering (applications in), 2
 error minimization problem, 212
 error minimizing functions, 216
 ESPRESSO-MV system, 22
 “eureka” moment, 10
 evaluative function, 74
 exhaustive enumeration, 40
 explanatory capabilities, 269
 extra-observer variability (in
 digital mammography)

F

false-negative, 9, 177, 243
 false-positive, 9, 177, 243
 feasible solution, 168
 features, 20
 feed forward neural network,
 181, 297
 feedback mechanism, 10
 fibrocystic change, 304
 fibrosis, 304
 finance, 4
 FIND-BORDER algorithm, 210
 finite needle aspirates (FNAs),
 124
 Fisher’s linear discriminant
 analysis, 184-185, 293, 295, 299,
 Fourier transform, 292
 fractile frequencies, 292
 fraudulent transactions, 8
 frequent item set, 250
 full wave rectified integral
 (FWRI), 292
 fuzzy *c*-means, 293, 299
 fuzzy *k*-nearest neighbor, 293,
 299
 fuzzy logic, 311-324
 fuzzy models, 293
 fuzzy set approaches (FSA), 269
 fuzzy sets, 313-315

G

GEM system, 133, 135
 general Boolean functions,
 237-238
 generalability power, 137
 genetic algorithms, 187
 gradient-based optimization
 procedure, 297
 graph theory, 176
 GRASP approaches, 78
 greedy algorithm, 35
 greedy approach, 164, 168, 175,
 217
 guided learning, 17, 79, 103-128,
 199, 234, 268, 279-280,
 284-288

H

Hansel chains, 209-210
 Hansel's algorithm, see *Hansel's
 lemma*,
 Hansel's lemma, 209, 225
 Hansel's theorem, see *Hansel's
 lemma*,
 hard cases (diagnosis of), 244
 hashing trees, 259
 hashing, 259
 heterogeneous data, 10
 heuristics for clause inference,
 71-101
 "hidden" logic, 48, 64, 116
 hierarchical approach, 194
 hierarchical decomposition, 234
 hierarchy, 11
 hierarchical decomposition of
 variables, 202-204
 histogram, 261
 homogeneous data, 10
 horizontal end points, 217
 Horn clauses, 69
 Horvitz-Thompson estimator,
 223, 225

hyperplasia, 304
 hypothesis space, 21, 113

I

ID3 algorithm, 116
 IF-THEN rules, 10
 illusions, see *accuracy illusions*
 incomplete data, inference with,
 79-83
 incremental algorithms, 135-140
 incremental learning from
 examples (ILE), 129-150
 incremental learning, 116, 276
 incremental OCAT (IOCAT),
 129, 141
 incremental updates, 216
 index of actual reliability, 181
 index of potential reliability, 181
 indexing terms, see *keywords*
 inductive inference problem, 106
 inference objectives, 214-223
 infiltrating ductal carcinoma, 303
 information retrieval (IR), 140,
 269, 273
 integer programming (IP), 34, 47,
 53-54
 intelligent systems, 106, 314
 inter-observer variability, 178
 interpretable properties, 247
 interpretation of results, 10
 interviewing process, 108
 intra-ductal carcinoma, 303,
 305-310
 intra-ductal comedo type, 303
 intra-observer variability (in
 digital mammography), 178, 311
 investment opportunities, 4
 Irvine repository of learning
 databases, 83
 isometric loading, 289
 isomorphic mapping, 205
 isomorphic posets, 205

isotonic block class with stratification algorithm (BCS), 213
 item domain, 250
 judgmental rules, 314

K

Karnaugh maps, 22
 k -CNF, 21
 k -decision rules, 21
 k -DNF, 21
 keywords, 269, 270
 K -nearest neighbor models, 187
 knowledge based learning, 23
 knowledge discovery, see *data mining*
 knowledge validation, 198
 knowledge, 12
 knowledge acquisition, 106

L

large problem, 69
 large-scale learning problems, 155, 167, 173-175
 layer partition, 204
 layer, 204
 LCIS, 304
 Le Gal type, 302
 learning algorithm, 21
 learning from examples, 20, 268
 learning theory, 182, 183
 least generalizing clause (LGC), 136-137, 141
 likelihood function, 219
 likelihood ratio, 212, 214, 219, 221-222
 likelihood value, 212, 220
 linear discrimination analysis, see *Fisher's analysis*
 linear programming, 178
 lobular, 312-313

lobulation, 311-324
 local optimal point, 74
 logic, see *Boolean function*
 logical discriminant functions, 195
 logistic regression, 293, 296-297, 299
 lower bound, 162, 165-167, 173-175
 lower limit, 61, 63, 66, 205

M

machine learning, 140
 malignancy, 181
 malignant case, 185
 mammographic image, 182
 mammography, 177, 187, see also *digital mammography*
 market basket analysis, 249
 market basket data, 249
 marketing methods, 4
 marketing opportunities, 4
 marketing strategies, 249
 MATH WORKS, Inc, 291
 mathematical logic, 183
 MATLAB software, 291
 maximum cardinality, 61
 maximum clique, 162, 175
 maximum closure problem, 213
 maximum flow problem, 213
 maximum likelihood estimates, 294
 maximum likelihood problem, 212, 229
 maximum likelihood ratio, 221, 229, 231-232
 maximum likelihood, 211
 maximum simplicity, 36, 52, see also *Occam's razor*
 McCluskey algorithm, 22
 mechanical system, 239
 medical diagnosis, 178

medical sciences (applications in),
 2
 membership questions, 200
 membership functions, 314-324
 metaplasia, 304
 microlobulation, 311-324
 military operations, 3
 MineSet software, 260
 MINI system, 22
 minimal CNF representation, 206
 minimal DNF representation, 206
 minimum cardinality problem
 (MCP), 39, 40, 41, 45
 minimum cardinality, 57
 minimum clique cover, 160, 162
 minimum cost SAT (MINSAT),
 23
 minimum number of clauses, 66
 minimum size Boolean function,
 24
 mining explorations, 3
 min-max algorithm, 213
 misclassification probabilities,
 212, 214, 216, 228, 231
 missing values, inference with,
 90-93
 modal value (of fuzzy numbers),
 315
 model of learnability, 183
 moment of truth, 10
 monotone Boolean function
 (single), 233
 monotone Boolean functions
 (applications of), 237-248
 monotone Boolean functions
 (number of), 207-208
 monotone Boolean functions (pair
 of independent), 233
 monotone Boolean functions,
 197-236
 2-monotone functions, 236
 monotonicity constraints, 200
 monotonicity property
 (identification of), 245-247

monotonicity property (in data),
 189, 190, 194-195, 197
 monotonicity, 18, 178
 most generalizing clause (MGC),
 136-137, 141
 Moore's law, 330
 multicast, 22
 multi-criteria decision making,
 314
 muscle condition, 292
 musculoskeletal injuries, 290

N

narrow vicinity (NV) hypothesis,
 183-190, 193-196, 235
 national security, 267, 268, 288
 9/11 events in the U.S., 4
 nearest neighbor methods, 23,
 178, 269
 negative class, 20
 negative examples, 28
 negative model, 104-105
 negative rule, 272
 nested monotone Boolean
 functions, 200, 233, 234, 239-241
 neural networks, 178, 182, 187
 neural networks, 23, 246, 269,
 293, 297-299, 299, 311
 noise in data, 11
 non-decreasing monotone
 Boolean functions, 199
 non-increasing monotone Boolean
 functions, 199
 non-incremental learning from
 examples (NILE), 129, 133
 number of candidate solutions,
 113-116

O

OCAT (one clause at a time)
 approach, 34-38, 294-295, 299

Occam's razor, 12, 22, 35, 36, 52
 Okham's razor, see *Occam's razor*
 optimal clause, 60, 63
 optimal solution, 41, 42, 46, 58, 163
 optimal state, 60
 optimality, 66
 optimistic assessment (in diagnosis), 323-324
 optimization problem, 52
 optimization, 35
 oracle (deterministic), 200
 oracle (stochastic), 202
 oracle (ternary), 201, 215, 234
 oracle (three-valued), 201, 215, 234
 oracle (unrestricted), 201, 227, 234
 oracle, 15, 103, 200
 ordered list, 24
 outliers, 11
 overfitting problem, 31
 overgeneralization problem, 31

P

PAC (probably approximately correct), 21, 183
 PAC algorithm, 21,
 pair of nested monotone Boolean functions, 215
 papillomatosis, 304
 parallel algorithm, 163
 parallelism, 78
 parameters, 20
 parametric model, 178
 parsers, 268
 partial knowledge, 71
 partially defined Boolean function, 23
 partially defined examples, 69
 partially ordered set, see *poset*

partition algorithm, 251
 pattern complexity, 12, 17
 pattern recognition, 183, 187
 patterns, 11
 performance illusions, see *accuracy illusions*
 periodogram, 292
 permutation, 111
 pessimistic assessment (in diagnosis), 323-324
 political and social sciences (applications in), 4
 political campaign, 4
 pooled adjacent violators algorithm (PAVA), 213
 population (of examples), 189
 poset, 204-208
 positive examples, 28
 positive model, 104-105
 positive rule, 272
 positive class, 20
 power spectrum, 292, 293
 predictive modeling, 252
 PRESTO system, 22
 principle of maximum possible simplicity, see *Occam's razor*
 problem definition, 6
 problem pre-processing, 42, 45-47
 problem solving, 1
 proceeding vector, 199
 process diagnosis problem, 243
 processing elements (in NNs), 181, 297
 pruning, 41
 pseudo-classification, 274
 population (of examples), 180-181
p-value, 142

Q

queries (minimum number of), 217

queries (number of), 224
 query complexity, 210-211, 233
 query costs, 236
 query complexity, 200
 question-asking strategy, 199, see
 also *guided learning*

R

RA1 algorithm, 253
 randomized algorithm 1 (RA1),
 74-79, 100, 124
 randomized algorithms, 74, 78
 randomized heuristic 2 (RA2),
 79-83, 100
 receiver operator characteristic
 (ROC) analysis, 189
 record linkage, 198
 redundant precedence relations,
 205
 rejectability graph (R-graph), 63,
 157-176
 related vectors, 199
 reliability criteria, 179-183
 reliability of data mining systems,
 177-196
 repair of a Boolean function, 136,
 138-139
 representation hypothesis,
 183-190, 193-196
 root mean square (RMS), 292
 Round-Robin test, see *cross*
 validation
 rule induction, 23

S

sample complexity, 21, 79
 sample set, 180-181
 SAT approach, 63, 153, 166-167
 satisfiability (definition of), 110
 satisfiability (SAT), 22, 32
 screening mammography, 177

search stages, 40
 search states, 40
 search tree, 43, 59
 sensors, 8
 sequential design of experiments,
 211
 sequential monotone rule
 induction, 201, 213
 set covering problem, 13, 36, 39,
 set of all monotone Boolean
 functions, 206
 set oriented mining (SETM), 251
 sigmoid function, 294
 sign test, 139, 279
 signal analysis, 289-290, 291
 Silicon Graphics, Inc., 260
 similarity measures
 size of population (of examples),
 181
 size of state space, 180-181
 Sokolov's algorithm, 209, 225
 "smart" questions, 199
 space complexity, 113-116
 SQL system, 251
 state space, 180-181, 189
 states (of nature), 1
 statistical models, 12
 stochastic membership questions,
 200
 succeeding vector, 199
 supervised classification, 267
 support (of association rule), 250
 support vector machines (SVM),
 23
 semantic analysis (SA), 269
 symbolic machine, 23
 symmetric chain partition, 204
 symmetric system, 104, 134
 synthetic databases, 257
 system, 1

T

target identification, 3
tautologies, 170
telecommunication data analysis,
252
terminal nodes, 40
terminal state
terminal state, 57-58
testing set, 294
text classification, 269
text documents (mining of),
267-288
threshold functions, 236
TIPSTER collection, 130, 140,
277
training data, 30, 180, 183, 294
training rate, 297
transactions, 249
trapezoid fuzzy numbers,
314-315
traveler screening, 3
triangular fuzzy numbers,
314-315
tubular carcinoma, 303
two-class problems, 1

U

ultra optimistic diagnostic system,
243
undecidable cases, 244
undulations (size of), 318
undulations, 312-324
upper zero, 205

V

validation (cross), 276, 279,
280-284
validation (leave-one-out), 276,
278-279, 280-284
validation (of system), 127

Vapnic-Chernovenkis dimension,
21, 115-116
variables, 20
variance-covariance matrix, 185,
294
vector (example) selection
criteria, 216-223
vector space model (VSM), 269,
273-275
vertical middle, 217
visualization of decomposition of
general functions, 238
visualization of monotone
Boolean functions, 207
visualization of nested monotone
Boolean functions, 208
visualization of posets, 205-208
VLSI applications, 22

W

Web mining, 267
Wisconsin breast cancer database,
83, 85-92, 125-127
Woman's Hospital (in Baton
Rouge, LA) dataset, 181, 187,
243, 301

Z

ZIPPF collection, 130, 140

