

INDEX

Note: illustrations and captions are indicated by *italic page numbers*, footnotes by suffix 'n'

- AC2 algorithm 324
 - performance on various data-sets 338, 341, *352-5, 357-9*
- ACE computer 9-10, 11, 40-2
- ACT theory 437
- agents
 - coordination using blackboard 411
 - derivation by inductive learning 411-12
 - low-level skills implemented by 411
- aircraft flight simulators 400-1
 - inductive rule-learning experiments 404-8
 - results 406, 408, *409-10*
- ALGOL 14
- algorithm, use by Turing 7
- algorithmic information theory, hypothesis evaluation using 90, 91
- Alloc80 algorithm 317
 - performance on various data-sets 335, 336, 339, *352-9*
- α -helix formation 101n, 200
 - chemical principles 200, *212*
- ambient vision 258
- Angluin's learning algorithm
 - compared with Yokomori's algorithm 182-4, *185*
 - example operation 183, *185*
- animal learning 213-28
 - basic procedures 214-15
 - perspectives 215-16
- anticipation 225-7, *245-6*
- AQ algorithm 325
 - see also* CN2 algorithm
- artificial intelligence
 - and logic 6, 21-3
 - Turing's contribution 9, 24
 - von Neumann's contribution 24
- artificial neuron concept 4, 12, 28
 - see also* neural networks
- assembler program 23
- associative transfer (in conditioning) 219-20
- atoms, Boolean algebra 146
- attribute generalization 139
- attribute noise 124
 - CAFE's learning ability affected by 135, 136, *136*
 - COBWEB's learning ability affected by 134, *136*
- attribute-oriented induction 77-9
 - application of minimal multiple generalization 79-83
- automated hypothesis-formation 87, 88
- automatic programming 13-14, 23
- automatization, skill-learning 389
- autopilots 406, 408
 - degree of control achieved 408, *409-10*
 - heuristic vs causal models 413-14
- Babbage's Analytical Engine 42
- background knowledge
 - chess endgames 298, 299
 - drug design 196, 201
 - encoding of 94-7
 - protein structure 199-200

INDEX

- representation of 93, 196, 200, 201
- back-propagation 326-7
 - compared with other systems 313-14, 352-9
 - performance on various data-sets 336, 337, 338, 339, 341, 352-9
- Back-propagation Multi-Layer Perceptron 326-7
 - biological interpretation 327
- backtracking 442, 449
 - occurrence in planning 452
 - suppression of 443, 444, 446-7, 453
- bang-bang control 364, 396, 398
- batch processing, inductive inference from frequency data 159-60
- Bayesian networks 320-1
 - see also* CASTLE algorithm
- Bayesian posterior variance, as significance measure 102-3, 104
- Bayes rule algorithm 316
 - performance on various data-sets 335, 337, 339, 341, 352-9
- behavioural field, mechanism for generating 251, 252, 255
- beta-reduction rule 16, 20
- bimolecular modelling, aims 194
- bindings
 - meaning of term 286
 - minimum number in universal engineering knowledge base 286, 287
- blackboard-like model, coordination of agents using 411-12
- BMT expert system 390, 391
- Boolean algebra
 - atoms 146
 - construction of Euclidean space 146, 148
- Hasse diagram 147
- Boswell (patient with no episodic memory) 390, 392
- BOXES algorithm 364-9, 396
 - cleaning up 379, 379-80
 - comparison of variants 370
 - Cribb-Sammut variant 367-8
 - Cribb's version 367
 - decisions on action 365-9
 - Law-Sammut variant 368-9
 - properties 369, 370-3
 - Law's version 368
 - local merit criterion introduced 367
 - original Michie-Chambers algorithm 366-7, 396
 - patterns discovered 379-80
 - readability of code 379, 381-2
 - task addressed 363-4
- brain
 - left hemisphere 257
 - activities associated 258, 414
 - right hemisphere 257
 - activities associated 258, 414
- brain development, embryo-to-young-child 250-5
- brain stem systems, activities controlled 258-9
- brain systems 248, 249
- British Museum algorithm 30
- brittleness bottleneck (problem in knowledge bases) 278-9
- C4.5 algorithm 324
 - aircraft flight simulation 405, 406
 - performance on various data-sets 352-7, 359
 - pole-and-cart system control rules 401
- CAFE concept formation system algorithm 191
 - characteristics 125, 129-30

- classification and learning by 129-30
- compared with COBWEB 129-30, 134, 135-6, 137-8
- compared with LABYRINTH 125, 126, 129, 138-9
- concept-predictability used 139-40
- evaluation of system 134-9
 - artificial domain 134-6
 - natural domain 137-9
- knowledge representation formalism 126-9
- learning curves 134, 136, 138
- Cal5 algorithm 324-5
 - discrimination measure used 325
 - performance on various data-sets 352-9
- canonical discriminant correlations (of StatLog project data-sets) 346, 351
- canonical discriminants, variation (for StatLog project data-sets) 346-7, 351
- cardiology 306-7, 332-3
 - see also* heart disease data-set; KARDIO system
- CART system 321-3
 - compared with ID3-type algorithms 322, 323
 - compared with neural networks 314
 - impurity measure used 322
 - performance on various data-sets 333, 338, 352-3, 355-8
 - pruning used 323
 - versions available 323
 - see also* INDCART
- CASTLE algorithm 320-1
 - characterization measures for 334, 342
 - performance on various data-sets 335, 336, 338, 339, 341-2, 352-9
- category utility 130
- causal networks 320
 - approach to learning 321
 - determination of structure 320-1
 - see also* CASTLE
- Central Limit Theorem 336
- cerebral cortex
 - development 248
 - links between two halves 255, 257
 - specialized areas 260
- Cessna aircraft flight simulator 404-6
 - flight plan described 404
 - variables sampled 405
- CFIX concept formation system 125
- checkers (draughts)
 - knowledge acquisition 392
 - playing programs 37, 46, 47
- chess endgames
 - learning classification rules 292
 - learning optimal strategies 291-307
 - materials used 293-6
 - method used 296-8
 - results 298-306, 307
 - see also* KRK chess endgame
- chess playing machines, Turing's
 - discussions 39, 41
- chess program 291-2
- child development, comparison with machine learning 45
- choice-point encoding 98-9
- Church, Alonzo 10
 - solution of Decision Problem 10, 38
- Church machines 18-20
- Church's lambda calculus 5
 - beta-reduction rule 16, 20
 - mathematical interpretation 14
 - see also* lambda calculus

- classical conditioning 215
 - with feedback 216
- classical logic, new (Boolean algebra) model 150
- classification, meaning of term 311, 312
- classification algorithms
 - comparative study 311-59
 - representative collection 313
- classification noise model 101, 113
- CLASSIT conceptual clustering system 333
- clause sets, specialization of 445
- clean-up effects
 - aircraft flight simulator control 406, 408, 409-10
 - BOXES algorithm 379, 380
 - pole-balancing control system 379, 380, 402-3, 403
- closed-world specialization (CWS) 107, 297
 - algorithm 108
 - clauses produced, compression-based selection used 112
- CN2 algorithm 325-6
 - performance on various datasets 338, 352-5, 357-9
- COBWEB 124, 125
 - compared with CAFE 129-30, 134, 135-6, 137-8
 - evaluation function used 130, 132-4
 - learning curves 134, 136, 138
- Colmerauer, Alain 16, 21
- commissurotomy 255-6
 - effect on speech 258
 - effect on visual information 256, 258
- common sequence, sharing in learning algorithm 448, 449
- common substructures, extraction in explanations 455
- communication mechanisms, human infants 261
- compactness, with respect to containment 63-4
- competitive learning (in conditioning) 220
- compiler program 23
- compression
 - estimation from theory 111
 - meaning of term 91
 - as noise-meter 104-6
 - as significance measure 100-4, 115
 - compared with other measures 102-3, 104
 - see also* database compression
- compression-based clause selection 107-13
 - CWS-produced clauses 112
- compression-guided learning
 - case study 106-15
 - empirical evaluation 113-15
- compression model
 - characteristics 117-18
 - effect of noise 114-15
- computers, and logic 4-6, 6-12
- concept formation
 - structure information utilized 123-41
 - see also* CAFE; COBWEB; LABYRINTH
- concept hierarchy 77, 79
 - generation of 124, 129
 - inductive learning of 124
- concept-predictability, CAFE using 125, 129-30, 132, 139-40
- concept-predictability function 132, 141
 - examples of calculation 139
- conditioned stimulus/response 215
- conditioning 215
 - associative transfer in 219-20, 239
 - CSPUS interval studied 218, 236

- facilitation of remote associations 219, 237
- first-order 220, 239
- impairment of proximal associations 219, 238
- representation in 221-3
- second-order 220, 239
- serial conditioned stimuli 219
- summation in 221-2, 240
- SuttonPBarto model 220-1
- confirmation theories 90
- conformity (for knowledge base) 277-8
- Connection Machine 31
- consciousness, genesis of 247-50, 414
- consistent max trees 67
- constraint networks 273, 274
- context-free grammars, learnability 170
- context-free transformations with flat base 72-5
 - inference algorithm for 74
 - operation of 73-4
- contradiction backtracking 173, 176
- control decision nodes 425, 428
 - successfully applied rule 428
- correlation coefficients (for StatLog project datasets) 345-6, 351
- cost matrices, use with classification algorithms 337, 341
- counterexamples, learning using 171, 173
- credit assignment 217-21, 364
 - basic principles 219-20
- credit risk dataset (StatLog project) 333-4
 - characteristics 351
 - performance of various algorithms 338-9, 358
- Cribb-Sammut algorithm (for pole-balancing problem) 367-8
 - compared with other variants of BOXES algorithm 370
 - asymmetric pushing 373, 374
- Cribb's local merit criterion 367
- critical feedback 216
- curriculum construction, design of knowledge base similar to 276, 284
- CWS *see* closed-world specialization
- cyclopean 252
 - with jointed member 253
 - watching another cyclopean 254
- CYC project 276
- database compression 307
- databases, discovery of knowledge in, minimal multiple generalization algorithm used 77-83
- datasets (StatLog project) 328-35
 - characterization of datasets 334-5, 340-1, 351
 - measures described 345-7
 - performance of various algorithms 335-40, 352-9
 - see also* credit risk...; handwritten digits...; head injury...; heart disease...; Karhunen-Loeve digits...; satellite image...; Space Shuttle control...; vehicle silhouettes dataset
- Decision Problem 7
 - solutions 8, 10, 38
- decision rule algorithm 325
 - see also* CN2
- decision tree algorithms 321-2, 323, 324
 - advantages 322
 - see also* AC2; C4.5; Cal5; CART; ID3; INDCART; NewID
- decision tree induction 307
 - patterns discovered using 380

INDEX

- decision trees, aircraft simulation data 405, 406
- declarative knowledge 388-9
 - compared with procedural knowledge 393-4
- declarative memory, contrasted with procedural memory 388
- deductive knowledge base 25, 26
- Desch, Joseph 9n
- descriptive approach to programming 14-15, 16, 21
- deterministic finite automata (DFAs)
 - learning using 169
 - compared with NFA learning 184
 - see also* non-deterministic finite automata
- device malfunction data 165
 - description length of propositions obtained 164
- Diagram Configuration (DC) model, perceptual chunks used 420
- dialogic closure 264
- dialogue (knowledge) elicitation methods 390, 392, 393
- digits datasets (StatLog project) 329-30
 - characteristics 351
 - performance of various algorithms 336-7, 353-4
 - see also* handwritten...; Karhunen-Loeve digits dataset
- dihydrofolate reductase (DHFR)
 - binding of trimethoprim analogues 195-8
 - favourable interactions 197-8, 207
- Discrim algorithm 317-18
 - performance on various datasets 338, 340, 352-9
- discriminant analysis *see* linear...; quadratic discriminant analysis
- DNA promoters data set 137
 - evaluation of CAFE and other systems 137-9
- draughts *see* checkers
- drug design 100-1, 195-8
 - Golem used
 - compared with Hansch equation 196-8, 208-11
 - rules obtained for favourable drugs 197-8
 - significance measures compared 104
- dual personality, human brain born with 264
- EBL systems 117
 - see also* explanation-based learning
- ECG interpretation 306-7
- Eckert, J.P. 4, 11
- Eckert-Mauchly computers 11
- EDSAC computer 11
- EDVAC computer 11
- embryo brain 250
- emotional referencing 263
- emotions, infant-mother 260-5
- engineering, role in computer design 12
- engineering knowledge
 - classification 284-6
 - general 284, 285
 - specific 284-5, 285-6
 - see also* nonspecific knowledge
- engineering knowledge base
 - clusters of knowledge 283
 - segmentation of knowledge 285-6
 - size 286, 287
- ENIAC computer 11
- entropy gain 323
- entropy gain ratio 324
- episodic memory 388
- equivalence queries, learning using 169, 175
- error rate

- data for protein structure prediction 101, 102
- reduction/correction 110
- Escherichia coli* dihydrofolate reductase
 - modelling of trimethoprim analogues binding to 195-6
 - see also dihydrofolate reductase (DHFR)
- Euclidean space, construction using Boolean algebra 146, 148, 151-3
- evaluation of logical theories
 - encoding scheme for 93-100
 - comparison with FOIL's scheme 99-100
 - input tape encoding 93-9
 - background knowledge and hypothesis encoding 94-7
 - example 96
 - proof encoding 97-9
 - output tape encoding 99
- exclusive-OR (XOR) problem, solving of 221
- experimental machine learning 289-359
- expert systems 26-7
 - knowledge acquisition 392
 - typical rule-based systems 391
- Explanation-Based Generalization (EBG) 445, 447
- explanation-based learning (EBL) 420
 - compared with PCLEARN 431-6
 - in geometry domain 423-4
 - macro-operators 421
 - operationality in geometry domain 421, 431-3
 - use in speed-up learning 442
- eyelid closure
 - conditioning in rabbit 217, 226, 234-5, 245
 - real-time models 226-7, 246
 - as protective measure 217, 226
- F-16 aircraft flight simulator 406-8, 409-10
 - clean-up effect due to autopilot 409-10
 - flight plan described 407
 - variables sampled 407-8
- failure, learning from 456
- familiar objects, time taken for recognition 393
- feature construction technique (in CAFE) 124, 125, 141
- Fifth Generation Project 27, 51, 388
- figure-pattern strategies 420
- finite elasticity 71
- finite thickness 71
- fire-protection equipment configuration, expert system for 391
 - see also BMT
- first-order conditioning 220, 239
- flight control simulations
 - blackboard-like model for coordination among agents 411-12
 - Cessna aircraft 404-6
 - F-16 combat aircraft 406-10
- flowchart diagram
 - early use 13
 - learning algorithms (for NFAs) 179, 187
- FOIL learning system 51
 - compared with neural network approach 314
 - encoding scheme 99-100
- forebrain commissurotomy 258
 - see also commissurotomy
- Friedman, William 9
- functional constraints 272
 - see also higher-order constraint network
- functional programming 143
- GASOIL expert system 390, 391
- GCWS algorithm 297

INDEX

- chess endgame Prolog programs
 - learned by 293
- generalization relation 79
- generalized database 80
 - with conceptual hierarchy 79
 - LCHR algorithm used 77-8
- generalized linear models (GLIMs) 318
- generalizing-number technique 448
- geometry domain 422-3
 - EBL techniques used 423-4, 431-6
 - experimental results using EBL and PCLEARN 430-6
 - learning search control knowledge in 422-4
 - PCLEARN used 427, 431-6
 - recognition rules 427
- GINI criterion 322
- Glasgow Outcome Scale 331-2
- Gluck's category utility 130
- goal-oriented search control knowledge, learning 420
- goal reduction, Prolog computation via 444
- Gödel, Kurt, as first programmer 7
- Gödel's descriptive recursive function formalism 6-7, 15
- Gödel's Incompleteness Theorem 6-7, 38, 39
 - Turing's attempt to circumvent 38
- Golem learning program 194-5
 - algorithm used 194-5
 - applications 50
 - bimolecular modelling 193-212
 - drug design 196-8, 208-10
 - protein structure prediction 101n, 199-201, 212
 - approach 51, 193, 206
 - compared with Hansch equation 196-8, 208-11
 - inputs 194, 196-7, 199-200
 - RLGGs implemented 297, 307
- Good, Jack 9, 39
- Grisin's logic 151
- guards, speed-up learning using 443, 456
- handwritten digits dataset (StatLog project) 329-30
 - characteristics 351
 - performance of various algorithms 336-7, 353
- Hansch linear regression equation 195
 - compared with Golem 196-8, 208-11
- Hasse diagram 146, 147
- head injury dataset (StatLog project) 331-2
 - characteristics 351
 - performance of various algorithms 337-8, 356
- heart disease dataset (StatLog project) 332-3
 - characteristics 351
 - performance of various algorithms 337, 338, 357
- helicopter control 404
- heterogeneity of knowledge 277
- Hierarchical Image Processing System (HIPS) 331
- hierarchy tree of concepts 128
- higher-level computing machines 31
- higher-order constraint network 273, 274
- Hilbert Program 6, 7, 10, 38
- historical perspectives 1-52
- homogeneity of covariances (of StatLog project datasets) 345, 351
- Horn-clause-resolution predicate calculus 5, 16
 - implication on Church machines 20
- Huffman coding 94, 96

- human cognition, regulation of 247-65
- human experts
 - non-determinism used 171
 - rule-formulation not liked 392, 394
- human mind, working of 28-9
- human reasoning, logic rarely used 24, 25
- human short-term memory capacity 297
- hydrocarbon separation system configuration, expert system for 391
 - see also* GASOIL
- hypotheses
 - automated formation 88
 - construction 89-90
 - evaluation 90-2
 - see also* incremental hypothesis
- IAS computer 11
 - similar machines listed 11
- ID3 algorithm
 - application to KRKN chess endgame 292-3
 - compared with neural networks 313-14
 - other algorithms based on 323, 324
 - patterns discovered using 380
- image analysis datasets (StatLog project) 328-31
 - performance of various algorithms 335-7, 352-5
 - see also* handwritten digits...; Karhunen-Loeve digits...; satellite image...; vehicle silhouettes dataset
- imperative approach to programming 15, 21
- incremental freezing 376-7
 - difficulties with BOXES algorithm 377, 378
- incremental hypothesis
 - construction 89-90
 - evaluation 90-2
- incremental learning 44
 - frequency data 161
- INDCART system 323
 - performance on various datasets 352-9
 - see also* CART system
- induced rule-sets, performance 402-3
- inductive inference 57-189
 - from frequency data 159-65
 - algorithm used 159-61
 - batch processing 159-60
 - cloud/rain example 145, 161
 - incremental processing 161
 - see also* concept formation
- inductive logic programming (ILP) 45-6, 47-9
 - applications 50
 - chess endgame 291-307
 - minimal multiple generalization used 69-77
 - theory 49-50
 - see also* Golem
- inductive skill-grafting
 - applications 400-10
 - see also* skill-grafting experiments
- infant consciousness, development 260-5
- inference algorithm 70
- information of probability distribution 155
- inheritance calculus, use in language translation for large knowledge bases 279
- inner product, definition 151
- inner product space *see* Euclidean space
- input tape encoding (for evaluation of logical theories) 93-9
- instance, definition 444
- instance-based prediction 333

INDEX

- Institute for New Computer Technology (ICOT) 27, 31
- instructive feedback 216
- instrumental learning 215
- integrity, knowledge bases 278
- intelligence, theories of regulation 249
- interpreter, use in machine learning 42, 48
- intuitive motherese (babytalk) 261
- Inverse Resolution 49
- Ishizaka's learning algorithm 187
- ITrue algorithm 326
 - performance on various datasets 353-8
- justification of logical theories 87-118
- KARDIO system 306-7
- Karhunen-Loeve (KL) digits dataset (StatLog project) 330
 - characteristics 351
 - performance of various algorithms 336, 354
- Kemeny, John 14
- kernel density estimation method 317
 - see also* ALLOC80 algorithm
- k*-minimal multiple generalization 63
 - algorithm for 59
 - application to attribute-oriented induction 79-83
 - compared with use of LCHR algorithm 82-3
- k*-nearest-neighbour (K-N-N) algorithm 316
 - performance on various datasets 335, 336, 337, 339, 352-9
- knowledge
 - types 387-8
 - see also* declarative...; procedural knowledge
- knowledge acquisition, support of 276-7
- knowledge-based programming environments, knowledge representation in 270
- knowledge bases
 - applications 276
 - building analogous to teaching humans 276, 284, 287
 - changeability 278
 - conformity 277-8
 - engineering knowledge base
 - clusters of knowledge 283
 - segmentation of knowledge 285-6
 - size 286, 287
 - extendability 278
 - integrity 278
 - large knowledge bases 269-87
 - openness 278
 - processing users' queries 282-3
 - stratification of knowledge 280
 - uniting into large knowledge bases 276, 287
- knowledge engineers, aims 387
- knowledge representation
 - CAFE system 126-9
 - large knowledge bases 271-5, 279-80
- knowledge system, formalization of 280
- Knuth, Donald 14
- Kohonen net 328
 - performance on various datasets 336, 352-3, 355, 357-8
- Kowalski, Robert 16, 17, 21
- k*-pivot 66
- KRK chess endgame
 - BTM (black-to-move) WFW (won-for-white) positions 300, 301, 303-4
 - depth 0 induced Prolog definition 299, 300-1

- depth 1 induced Prolog definition 301, 302, 305-6
 - number of legal positions 295-6, 295
 - canonical positions 294
 - databases 291, 293-6
 - depths of win 293
 - number of positions available 295
 - illegality induction task 101
 - significance measures compared 104
 - legality learning task
 - compression-guided CWS procedure used 113-15
 - GCWS algorithm used 297-306
 - symmetrical translation 294, 295
 - KRKN chess endgame 292
 - Kullback-Leibler measure 335, 342
 - kurtosis (of StatLog project datasets) 347, 351
 - LABYRINTH 125, 126
 - attribute generalization used 139
 - compared with CAFE 125, 126, 129, 138-9
 - representation of instance 127
 - lambda calculus
 - on Landin machine 20
 - see also* Church's lambda calculus
 - Landin, Peter 20
 - LandSat image data 328-9
 - see also* satellite image dataset
 - Laplace's Law of Succession 325
 - large knowledge bases 269-87
 - amount of work needed to build 287
 - architecture 279-83
 - contents 283-6
 - knowledge representation for 271-5
 - operational parts described 280-2
 - referential part 282
 - requirements 276-9
 - size 278, 286, 287
 - Law-Sammut algorithm (for pole-balancing problem) 368-9
 - compared with other variants of BOXES algorithm 370
 - asymmetric pushing 373, 374
 - properties 369, 370-3
 - Law's version of BOXES algorithm 368
 - layered network model 222-3, 241
 - simulated learning curves using 223, 225, 242, 244
 - LCHR algorithm 77
 - generalized database obtained using 78
 - compared with use of k -mmg algorithm 82-3
 - leap-year example 109
 - learning, definitions 413
 - learning algorithm
 - LA (for NFAs) 177-8
 - correctness 178, 180-1
 - flowchart diagram 179
 - time analysis 181-2
 - speed-up learning 445-8
 - learning control 361-456
 - learning curves
 - animal learning 235, 237-40
 - simulations on layered network 242-4
 - CAFE system 134, 136, 138
 - COBWEB 134, 136, 138
 - LABYRINTH 138
 - Tsukimoto-Morita algorithm 163-4
- learning from failure 456
 - learning to learn 224-5
 - demonstration 243
 - simulation on layered network 225, 244

INDEX

- least general generalization 60
 - applications 69
 - generalization of 59-83
 - meaning of term 62
- least Hebrand model (of program)
 - 70, 72
- least-mean square rule 223
- linear discriminant analysis 317-18
 - performance on various data-sets 337, 338, 340, 341
 - see also* Discrim algorithm
- linear regression methods, compared with Golem 197-8, 208-11
- linguistic learning 263-4
- LISP 16
- living organisms, growth 247
- local merit criterion 367
- logic
 - and artificial intelligence 6, 21-3
 - and computers 4-6, 6-12
 - and neurocomputation 27-30
 - and programming 12-13
 - and programming languages 16-21
- logical entropy 154
 - and information of probability distribution 155
- logical function, information of 154
- logical theories
 - evaluation, encoding scheme for 93-100
 - justification 87-118
- logical vector 146, 149-53
 - definitions used 149
 - interpretation of nonclassical logical vectors 156, 157
 - meaning of term 153
 - relationship to probability vector 148, 153-9
 - correspondences 156-8
 - transformation from probability vector 146, 158-9
- Logic Theorist 29-30
- logistic regression (LogReg) method 318-19
 - performance on various data-sets 337, 338, 340, 352-9
- McCarthy, John 16, 21
 - Minsky's opinion on his approach to AI 24, 25
- McCarthy's logical deductive knowledge-base paradigm 24, 25, 26, 388-9
- McCulloch, W.S. 4, 12, 27
- machine learning
 - first program developed 37, 46-7
 - first shown by imitation of trained human 402
 - integral role in computers 51
 - memory size required 41
 - Turing's ideas 41, 51
 - see also* concept formation; inductive learning
- machine learning algorithms 321-6
 - performance on various data-sets 340, 352-9
 - time taken 341
 - see also* AC2...; C4.5...; Cal5...; CART...; CN2...; Golem; ID3...; INDCART...; ITrule...; NewID algorithm
- macro-operators
 - average sizes 433
 - increased matching cost 434
 - matching costs compared 435-6
 - operationality
 - experimental results 431-3
 - factors affecting 421
 - re-ordering effect 433-4, 435
 - unsuccessful application 434, 435
 - see also* PCLEARN; perceptual chunks

- Makarovic's rules (for pole-and-cart system) 380, 398
- Martin-Löf's Intuitionistic Type Theory 5
- Marvin program 44-5
- Mauchly, J.W. 4, 11
- MAXT algorithm 68
- max trees *see* consistent max trees
- medical datasets (StatLog project) 331-3
 performance of various algorithms 337-8, 356-7
 see also head injury...; heart disease dataset
- medical diagnosis expert system 391
- membership queries, learning using 169, 175
- memory
 accessibility to 388
 declarative 388
 procedural 388, 394
- metacntrl (in split brain) 259
- meta-interpreter 442-3, 456
- metaknowledge 282
- Michie, Donald, discussions with Turing 39, 40
- Michie-Chambers algorithm (for pole-balancing problem) 366-7, 396
 compared with other variants of BOXES algorithm 370
- microtheories, knowledge representation using 284
- minimally adequate teacher (MAT) learning model 169
 interactive learning in 170
 learnability of NFAs in polynomial time 171-2
- minimal multiple generalization (MMG)
 algorithm for 59, 65
 applications
 in inductive logic programming 69-77
 in knowledge discovery in databases 77-83
- Minimum Description Length (MDL) criteria 99, 144
 proposition obtained by Tsukimoto-Morita algorithm 162, 164
- Minsky, Marvin 21
 on McCarthy's approach to AI 24, 25
- MMG algorithm 65
- modularity of knowledge 277
- molecular modelling, Golem used 50
- mother love 261
- multigroup discriminant analysis 317
 see also Alloc80 algorithm
- multimodality measures 335, 339
- mutual induction, use by CAFE 125, 129, 141
- MYCIN expert system 391
- naive Bayes algorithm 316
 performance on various datasets 335, 337, 341, 352-9
- NASA
 LandSat image data 328-9
 Space Shuttle data 334
- National Cash Register (NCR) Company 9n
- National Physical Laboratory (NPL), Turing's work 9-10, 40, 43
- natural intelligence (NI) 25
- natural numbers, encoded using prefix codes 95
- nearest-neighbour algorithms 316, 340, 341
 see also *k*-nearest-neighbour (K-N-N) algorithm
- negative patterning 222, 240
 simulation by layered network model 223, 242
- neocortical commissurotomy 255

INDEX

- see also* commissurotomy
- nervous system 247
- neural networks 326-8
 - compared with ID3 algorithm 313-14
 - compared with other algorithms and systems 200, 201, 313-14, 352-9
 - drug design using 195
 - lack of diagnostic help criticized 342
 - performance on various datasets 336, 337, 340-1, 352-9
 - problems solved 223
 - protein structure prediction using 199, 200
 - see also* back-propagation...; Kohonen algorithm...; layered network model...; radial basis functions
- neurocomputation 27-30
- neurological data *see* head injury dataset
- neurone activity 250-1
- newborn baby's brain 249, 250
- NewID algorithm 323
 - performance on various datasets 338, 352-9
 - see also* ID3 algorithm
- Nim game 37, 46-7
- noise
 - effect in learning 44, 101
 - compressive theories 114, 115
 - estimation using compression 104-6
- noisy data, discovery of propositions in 143-66
- nonclassical logical vectors, interpretation 156, 157
- nonclassical logics 27
- non-determinism, use by human experts 171
- non-deterministic finite automata (NFAs)
 - application in pattern matching 171
 - query learning by 172-88
 - algorithm used 177-8
 - compared with DFA learning 184
 - construction of new candidate rules 176
 - diagnosis of set of transition rules 176-7
 - introduction of new states 175
 - learning protocol used 175
 - notation used 174-5
 - usefulness 170-1
 - non-monotonic inductive inference, closed-world specialization used 108
 - nonmonotonicity 278
 - nonspecific knowledge 284, 285
 - norm, definition 152
 - NTgrowth 333
 - NUT programming environment 275
 - Occam's razor 90
 - operant conditioning 215
 - with feedback 216
 - optimal chess strategies, learning 291-307
 - orthonormal systems 152-3
 - output tape encoding (for evaluation of logical theories) 99
 - overfitting 144
 - pruning to avoid 144, 323
 - PAC-learning 114, 117
 - parallel computation, Church machines used 20
 - partial structures of explanations, learning methods based on 443, 456
 - pattern-matching problem
 - DFA compared with DFA 171
 - non-determinism used 170-1
 - Pavlovian conditioning 215

- PCLEARN system 424-30
 chunking facility 426
 algorithm used 426-30
 method used 425
 specification 425
 compared with EBL learner 431-6
 experimental results
 learning performance results 433-6
 methods used 430-1
 operability of
 macro-operators measured 431-3
 future work proposed 437-8
 input 425
 learning concept 424-5
 output 425
 overview 426
 problem-solver module 426
 peak-to-peak heuristics 454-5
 perceptual chunks 420
 learning based on 421-2
see also PCLEARN
 perceptually chunked
 macro-operators 430
 perfect causality 456
 physical symbol system hypothesis 386, 386-7
 Pitts, W. 4, 12, 27
 planning
 example 449-50
 proof tree for 450
 experiments in robot domain 452-4
 Plotkin, Gordon 37
 pole-balancing problem 363-4
 asymmetric pushing problem 373
 comparison of various algorithms 374
 BOXES algorithm used 364-9, 396
 clean-up effects 379, 380
 C4.5 algorithm used 401, 402
 clean-up effects 402-3, 403
 determination of *DK* 372
 determination of *K* 369, 371
 induction of rules from behaviour 402
 line-and-pointers representation 401
 order in which dimensions checked 398
 reliable controllers 375-8
 role of problem representation 401
 pole-and-cart system 363-4, 396, 402
see also pole-balancing problem
 polynomial time *k*-mmg algorithm 62-9
 polynomial update time inference algorithm 70
 polytree algorithms 320, 342
see also CASTLE algorithm
 polytrees 321
 pooling of learning results 376
 predicate calculus 46, 48
 use in Golem 195
see also Inductive Logic Programming
 predicate invention 50, 307
 prefix codes 94
 natural numbers encoded using 95
 prefunctional morphogenesis 248
 primitive attribute 126
 primitive Prologs 75-7
 definition 75
 inference algorithms for 76-7
 probability vector
 relationship to logical vector 148, 153-9
 correspondences 156-8
 transformation to logical vector 146, 158-9
 problem-solving, involvement of both declarative and procedu-

INDEX

- ral knowledge 412
- procedural knowledge 389–93
 - compared with declarative knowl-
edge 393–4
- procedural memory 394
 - contrasted with declarative mem-
ory 388
- PRODIGY system 424, 425, 437,
456
- productivity 223–5
- programming
 - descriptive approach 14–15
 - imperative approach 15
 - and logic 12–13
 - Turing's attitude 13–14
 - see also* automatic program-
ming
- programming languages
 - development 5, 16
 - and logic 5, 16–21
- program-translation programs 23–
4
- projection pursuit regression 319–
20
 - see also* SMART algorithm
- Prolog 16
 - computation via goal reduc-
tion 444
 - as ideal language for symbolic
learning 48
- proof encoding 97–9
 - example 97–8
- proof trees 445
 - geometry problem 428
 - planning example 450
 - positive/negative instances 447
- propositional calculus theorem prov-
ing program 29
- propositions
 - acquisition of 144
 - discovery in noisy data 143–
66
- protein docking problem 201
- protein secondary structure pre-
diction task 100, 198–9
 - chemical rules governing 200,
212
 - error rate data 101, 102
 - Golem used 199–201
 - significance measures compared
104
- protoconversations (baby–mother)
261, 262
- proximal associations, impairment
in conditioning 219, 238
- psychological development 249
- psychology, in AI 24–5, 401
- PVM algorithm, compared with
backpropagation 314
- quadratic discriminant analysis 318
 - performance on various data-
sets 336, 337, 338, 340,
341, 352–9
- quantitative structure–activity re-
lationship (QSAR) mod-
elling 195–6
 - rules 202–3
- query learning 169–88
 - see also* minimally adequate
teacher (MAT) learning
model
- rabbit eyeblink, conditioning of 216–
17
- radial basis functions 327–8
 - performance on various data-
sets 335, 336, 337, 339,
352–8
- real-time skills, building symbolic
representations 385–415
- reclassification error rate 101
- recognize–act cycle, frequency 389–
90
- REDUCED algorithm 66
 - operation of 65–6, 67
- reinforcement learning, combined
with induction 380–2
- relational attributes 126
 - value distribution 129, 139–
40

- relative least general generalizations (RLGGs) 49, 297, 307
- remote associations, facilitation in conditioning 219, 237
- Rescorla-PWagner rule 223
- resubstitution error rate 101
- retrograde analysis method, chess endgame databases generated using 293-5
- rewriting rules 18-19
- Robinson's resolution logic 21
 - inversion 49
 - see also* Inverse Resolution
- rote learning programs 37, 47
- rotorcraft control systems 404
- rule-based control, experimental study 395-400
- rule-conjecture-and-test procedure 394
- Runge-Kutta algorithm 396-7
- run-time thinking 414
- Samuel, A.L. 37, 47
- satellite fault diagnosis, ILP approach used 307
- satellite image dataset (StatLog project) 328-9
 - characteristics 351
 - performance of various algorithms 335-6, 352
- scientific domains 191-287
- Scott, Dana 14
- search control knowledge
 - learning 420
 - in geometry domain 422-4
- SECD machine 20-1
- second-order conditioning 220, 239
- self, origin in brain 414-15
- self-reproduction 28
- semantic memory 388
- Shapiro's contradiction backtracking algorithm 176
- Shapiro's debugging system 42, 51
- Shapiro's Model Inference System 41-2
- Shuttle... *see* Space Shuttle...
- skewness (of StatLog project datasets) 347, 351
- skill, meaning of term 389
- skill acquisition, postulates for 393-4
- skill-grafting experiments 400-14
 - aircraft flight 404-10
 - pole-balancing 401-4
- SMART algorithm 319-20
 - performance on various datasets 337, 338, 352-9
- SOAR learning method 437
- software reusability 269-71
- somatotopic maps 251, 255
- spacecraft attitude control problem 382, 395-400
 - black-box simulator 396-7
 - constraints 395-6
 - decision arrays used 399
 - result of control strategy 398-400
 - rules used 397-8, 399
- Space Shuttle control dataset (StatLog project) 334
 - characteristics 351
 - performance of various algorithms 339-40, 359
- speed-up learning 40, 441-56
 - algorithm for 445-8
- split brain
 - anatomy 257
 - levels and directions of consciousness in 255-60
- state transition diagram, BOXES algorithm 376
- statistical algorithms 316-21
 - performance on various datasets 338, 340, 352-9
 - time taken 341
 - see also* Alloc80...; Bayes...; CASTLE...; Discrim...; K-N-N...; LogReg...; Quadra...; SMART algorithm

INDEX

- statistical analysis methods, protein structure prediction using 199
- StatLog project
 - aim 311, 312
 - algorithms studied 316–28
 - machine learning algorithms 321–6, 352–9
 - neural network algorithms 326–8, 352–9
 - statistical algorithms 316–21, 352–9
 - time taken 341
 - datasets studied
 - characterization of datasets 334–5, 340–1, 351
 - credit risk data 333–4, 338–9, 358
 - handwritten digits data 329–30, 336–7, 353
 - head injury data 331–2, 337–8, 356
 - heart disease data 332–3, 337, 338, 357
 - image analysis data 328–31
 - Karhunen-Loeve (KL) digits data 330, 336, 354
 - medical data 331–3, 337–8, 356–7
 - satellite image data 328–9, 335–6, 352
 - Space Shuttle control data 334, 339–40, 359
 - vehicle silhouettes data 330–1, 337, 355
 - key personnel listed 343–5
 - objectives 312
 - previous studies commented on 314–15
 - results 335–40, 352–9
 - credit risk dataset 338–9, 358
 - image analysis datasets 335–7, 352–5
 - medical datasets 337–8, 356–7
 - Space Shuttle dataset 339–40, 358
 - testing methodology described 315–16
 - stimulus compounding 221, 240
 - simulation by layered network model 223, 242
 - Strachey, Christopher 13, 14, 37, 46–7
 - STRIPS planning program 449–52
 - strong AI 26
 - criteria summarized 386
 - logico-neural version 27
 - supporters of thesis 26, 386
 - structure information, utilization in concept formation 123–41
 - supervised learning 124, 140
 - with instructive feedback 216
 - pruning methods used 144
 - support, knowledge acquisition 276–7
 - SuttonPBarto model 220–1
 - symbol description header, components 94
 - symbolic propositional algorithms 321
 - performance on various datasets 340, 352–9
 - see also* machine learning algorithms
 - symbolic representation of intuitive processes 387
 - compared with neural-net representations 414–15
 - Symbol-Level Learning (SLL) 442
 - classes 443
 - teaching of humans, building of knowledge bases analogous to 276, 284, 287
 - thought, and knowledge 387–94
 - tightening process 65

- touch-typing, declarative memory
 - disabled during 392-3
- training set accuracy/covery, as
 - significance measure 102, 103, 104
- transfer of training 224
 - see also* learning to learn
- transparency of knowledge 277
- tree pattern language 62
- trial-and-error learning 215
- trimethoprim analogues 207
 - binding with dihydrofolate reductase 195-8, 207
 - observed vs predicted activity 208-9
- Tsukimoto-Morita algorithm 145-6, 159-61
 - description length of proposition obtained 162, 164
 - experimental analyses 161-2, 163-4
 - incremental processing version 161
- Turing, Alan
 - attitude to programming 13-14
 - building of universal machine 8, 9-10
 - child development compared with machine learning 45
 - contributions to AI 24, 37, 43
 - on Hilbert's decidability question 7, 8, 38
 - on incremental learning 44
 - invention of universal machine 1-2, 6, 7-8
 - at Manchester 12
 - on neural-net vs logic-based learning 42
 - at NPL 9-10, 40, 43
 - at Princeton 10
 - and von Neumann 2, 3n, 8-9
 - war-time work 8-9, 39-40
- Turing machine 7-8
 - invention 1-2, 6
 - model for
 - evaluating logic programs 91
 - incremental hypothesis evaluation 92
 - von Neumann's practical version 11, 30
- Turing's imperative notation 15
- Turing Test 24, 26
- tying shoelaces, as example of procedural knowledge 390
- unconditioned response 215
- unconditioned stimulus 215
- unit clause programs, inference algorithm for 71-2
- univariate skewness (of StatLog project datasets) 347, 351
- universal computing machine 38
 - implications of concept 3-4, 8
 - invention by Alan Turing 1-2, 6, 7-8, 18, 38
 - practical exploitation 5, 8, 9-10, 30-1
- unsupervised learning 124-5, 140
- VAX computer configuration, expert system for 391
 - see also* XCON
- vehicle silhouettes dataset (Stat-Log project) 330-1
 - characteristics 351
 - performance of various algorithms 337, 355
- virtual other 264
- visual information uptake 256
- visual input (to brain hemispheres) 256
- von Neumann, John
 - contribution to AI 24, 27-8
 - practical development of Turing machine 11, 30
 - on practical implications of universal machine 3-4

INDEX

- propensity to develop other
 people's ideas 2, 4
- and Turing 2, 3n, 8-9, 10
- von Neumann architecture 2, 11
- voting, use in learning 376
- weak AI, criteria summarized 386
- weather (rain/cloud) example 145,
 161
- Widrow-Hoff rule 223
- Winograd, T. 21
- Womersley, J.R. 4n
- XCON expert system 391
- Yokomori's learning algorithm 177-
 8
 - compared with Angluin's al-
 gorithm 182-4, 185
 - correctness 178, 180-1
 - example operation 183, 185
 - flowchart diagram 179
 - practical variant 186-8
 - flowchart diagram 187
 - time analysis 181-2



