

INDEX

- absorbing state 163
- action-penalty representation 233, 238-239, 245
- adaptive heuristic critic (AHC) 12, 18, 286, 288
- admissible 234, 245
- advice-taking 252
- average-case reward 8, 159

- backgammon 61
- basis-function representations 81-84
- batch model 112, 126
- Bellman error 210
- bias 8, 135
- bias optimality 161, 164-165
- Blackwell optimality 168

- CMAC 143
- compact representation 60, 64
- consistent 234
- convergence results 8, 37, 42-45, 64, 71-72, 80, 111, 131
- cost-to-go 59
- credit assignment 13

- diameter 229
- Dyna 243
- dynamic programming 7, 59
- dynamic programming operator 64, 203, 206

- eligibility trace 8, 123, 286
- ergodic 163
- evolutionary algorithms 11
- expected-value criterion 198, 203
- experimentation-sensitivity 286, 288
- exploration 174, 180-181, 183, 227
- exploration 8

- features 61, 65, 125
- FOO 272

- gain optimality 161, 164
- generalization 23
- genetic algorithms 7, 11, 27
- GENITOR 12, 20, 27
- goal-directed reinforcement-learning 227
- goal-reward representation 233, 241, 245

- gradient descent 102
- greedy policy 63, 201, 203

- heuristic dynamic programming 42, 47, 159-160

- infinite-horizon discounted criterion 34, 72, 160, 198, 230
- interpolative representations 80-81
- irreducible 163

- KBANN 253, 256-257
- knowledge compilation 253

- learning automata 162, 174
- least-squares TD (LSTD) 40-42
- least-squares value iteration 68, 69
- limit cycles 181-184
- linear approximation 67, 96, 98
- linear least-squares approximation 33, 38-40
- linear reward-inaction 174
- look-up table 66
- loss 96, 107, 207
- lower bound 108

- Markov decision process 34, 62, 69, 74, 162, 197
- Markov games 200
- Markov process 109, 126, 133
- maximum-likelihood estimate 128, 132
- minimax criterion 8, 199, 205
- Monte Carlo algorithms 126
- mountain-car task 141
- multichain 163

- n-discounted optimality 161, 166, 168
- neural networks 11, 60, 67, 251, 255
- normalized TD (NTD) 38

- operationalization 252, 259

- parti-game algorithm 200
- partially observable Markov decision processes, 74, 268
- Pengo 261

periodic 163
pole-balancing 16
policy iteration algorithm 73, 169
premature convergence 13
priority-Dyna 288

$Q(\lambda)$ -learning, 283, 286-288
Q-hat learning, 199, 207, 231, 238-240, 245
Q-learning 12, 19, 38, 62, 68, 74, 123, 161,
175, 184, 204, 207, 230-231, 238-240,
256, 285

K-learning 161, 175-177, 181-184, 186-190
recurrent 163
recursive least-squares TD (RLSTD) 45, 47
resource constraints 200

safely explorable 229
SAMUEL 27
SANE 17-18
SARSA algorithm 142, 288
state aggregation 66, 70, 73
symbol-level learning 7-8

T optimality (see bias-optimality)
 $TD(\lambda)$ 100-102, 110
temporal difference (TD) learning 8, 33,
35-38, 47, 62, 68, 74, 96, 100, 123, 125,
283-284

Tetris 74
transient 163

unichain 163, 174
uniformly-initialized 233
uninformed 229
upper bounds 102, 107

value iteration algorithm 61, 63-64, 70,
76-77, 170-173
variance 48, 136

WHM 113-115
Widrow-Hoff algorithm 113-114
worst-case analysis 96