Index

a
Active learning, 61
Actor-critic algorithm, 907
Ad-clicks, 459
Adaptive learning algorithms, 202
as sequential decision, 202
designing policies, 209
learning problems, 202
Affine function, 747
Aggregation, modeling, 125
Aleatoric uncertainty, 564
Algorithm
 approximate value iteration
 linear model, 886
Benders for two-stage, 950
indifference zone selection, 383
optimal computing budget allocation, 383
temporal-difference learning
 infinite horizon, 833
actor-critic, 907
ADP with pre-decision state, 872
approximate policy iteration,
  901, 907
 linear model, 903
LSTD, 905
backward ADP
 lookup tables, 800
parametric model, 800
backward dynamic programming, 749
bias-adjusted Kalman filter stepsize, 297
direct lookahead policy simulator, 994
double-pass ADP, 876
Gauss-Seidel variation, 758
hybrid value/policy iteration, 764
least squares temporal differencing, 905
MCTS algorithm
 backup step, 1014
 roll-out policy, 1011
 simulation step, 1012
tree policy, 1012
policy evaluation, 825
policy iteration, 763
Q-learning
 finite horizon, 875, 876
relative value iteration, 758
shortest path algorithm, 76
value iteration, 757
American option, 894
Anytime problems, 41
Apparent convergence, 300
Applications
 ad-clicks, 320, 370, 459
aid in Africa, 813
American option, 894
American options, 55
asset selling, 54, 662
asset valuation, 437
bidding, 992
blood management, 450, 715, 926, 1070
blood sugar, 662
budgeting problem, 752
building management, 1067
business, 5
carbon monoxide, 567
cash management, 186
chemical diffusion, 570
clinical trials, 55, 810
computer simulations, 320
cost-minimizing newsvendor, 226
COVID pandemic, 575
currency exchange, 513
demand estimation, 567, 569
diabetes medication, 490
diabetes medications, 376
diseases, 114
drones, 580, 611
dynamic assignment, 447
dynamic resource allocation, 926
dynamic shortest path, 707
e-commerce, 5, 227
economics, 5
electric utility, 931
electricity contracts, 509
electricity price forecasting, 565
electricity prices, 589
energy, 926
energy forecasting, 489
energy investments, 186
energy modeling
  crossing times, 593
  jump-diffusion, 590
  mean reversion, 590
  quantile distributions, 591
  regime shifting, 592
energy planning, 986
energy storage, 471, 523, 611, 662, 811
  active learning, 526
  passive learning, 525
  rolling forecast, 717
  rolling forecasts, 526
  time series price, 525
energy storage policies, 627
CFA, 628
comparison, 629
DLA, 629
hybrid, 629
PFA, 628
VFA, 628
engineering, 5
engineering design, 185
European options, 55
field experiments, 320
finance, 5
financial trading, 611, 711, 986
fleet management, 78
  multiple drivers, 80
  nomadic trucker, 79
flu mitigation, 1044
  controllable truth, 1048
  designing policies, 1054
  resource constrained, 1049
  spatial model, 1049
  static model, 1045
  time-varying, 1047
  two-agent learning, 1050
  with drift, 1047
freight transportation, 5
gambling problem, 751
games, 514
graph problems, 433
grid model, 570
homeland security, 55
HVAC controller, 1067
  model, 1067
  policies, 1069
information collection, 5
inventory, 6, 13, 73, 226, 439, 611
investment banking, 932
laboratory experimentation, 458
load matching, 705
logistics, 800
machine replacement, 55
machine scheduling, 992
manufacturing, 328, 932
materials science, 5, 320, 376, 986
medical, 514
medical decisions, 81, 114, 320, 406, 457, 580, 986
medical diagnostics, 567
medical research, 5
money, 926
nested newsvendor, 226
newsvendor, 70, 189, 406, 929
nomadic trucker, 124, 434
oil inventories, 580
online bookseller, 932
police patrolling, 610
power grid, 567, 572
power transformers, 509
pricing, 80, 114, 572
public health, 5, 114
rat diet, 572
resource allocation, 446
multi-period, 933
two-stage, 930
ride hailing, 991, 1008
rockets, 47
scientific exploration, 82
shortest path, 76, 189, 992, 1001
deterministic, 76
dynamic, 78
robust, 78
stochastic, 77
shuttle bus, 662
SpaceX, 47
sports, 114, 320
stochastic shortest path, 433
stopping problems, 54
supply chain, 5, 6
tic-tac-toe, 898
transformer replacement, 435
transportation, 320, 799
trucking, 489, 799, 926, 991
two-agent newsvendor, 1062
water reservoir, 611
Approximate dynamic programming, 50
single state, 855
LSPE, 857
LSTD, 856
recursive least squares, 857
Approximate linear programming method
linear model, 912
Approximate policy iteration, 900
linear model, 903
lookup tables, 901
Approximate value iteration, 828
backward pass, 876
bounding $1/n$ convergence, 859
geometric view, 850
gradient-based methods, 845
linear models, 845, 886
lookup tables, 871
post-decision state, 875
pre-decision state, 872
slow backward learning, 832
TD(0), 830
Approximation architectures, 165
Asset valuation, 437

b
Backgammon, 492
Backpropagation, 879
Backward ADP benchmarking, 810
aid in Africa, 813
clinical trials, 810
energy storage, 811
finite horizon, 797
lookup table, 800
parametric model, 800
VFA approximation, 805
  linear model, 806
  monotone function, 807
Backward dynamic
  programming, 749
Base model, 528
Base models vs. lookahead
  models, 528
Basis functions, 940
  American option, 896
  approximate linear
    programming, 913
  tic-tac-toe, 899
Bayesian prior, 112
Bayesian updating
  correlated normal beliefs, 113
  independent beliefs, 113
  normal-normal, 113
  precision, 113
Behavior policy, 208, 890
Belief MDP, 1060
Belief state, 408, 495
Bellman error, 829
Bellman’s equation, 741
  curse of dimensionality, 47
  deterministic, 741
  expectation form, 742
  linear model, 837
  linear operator, 747
  matrix form, 743
  operator form, 746
  standard form, 742
Bellman’s optimality equation, 46
Benders decomposition, 949
  asymptotic analysis, 951
  two-stage, 946
Benders’ decomposition
  regularization, 951
Beta distribution, 587
Bias and variance, 118
Bias in max operator, 909
Blood management problem, 450
Boltzmann exploration, 337, 627
Budgeting problem, 752
C
Chance constrained
  programming, 61
Classification of problems, 529
  state-dependent, cumulative
    reward, 531
  state-dependent, final reward, 532
  state-independent, cumulative
    reward, 530
  state-independent, final
    reward, 530
Competitive analysis, 391
Conditional value at risk, 523
Contextual bandit, 406
Control law, 49
Correlated beliefs, 113
Cost function approximation
  constraint-modified, 714
    blood management, 715
    dynamic trading, 711
    energy storage, 717
    parameterized policy, 722, 1004
    SPSA, 722
    tuning, 723
  general form, 703
  heat maps, 723
  interval estimation, 702
  uncertainty bonus, 702
  objective-modified, 704
    cost function correction, 705
    dynamic assignment, 705
    pure exploitation, 702
    shortest path, 707
Cost-to-go function, 48
COVID pandemic, 575
Crossing times, 593
Cumulative reward, 42
  adaptive policy, 401
bandit problems, 323
cumulative loss, 393
examples, 59
expectation
  state-dependent, 530
expensive experiments, 349
knowledge gradient, 362, 369, 370, 400
learning, 206
multiarmed bandit, 57
newsvendor, 71
policy evaluation, 641, 996
simulated
  state-dependent, 533
  state-independent, 533
  state-independent, 530
stochastic gradient algorithm, 224, 319, 327
stochastic search, 187
two-agent, 1051
Curse of dimensionality, 47, 162
Cutting planes, 945
d
Data driven, 70, 82
Decision tree, 44, 738
Decision variables, 65
  constraints, 504
  execution decisions, 503
  policies, 505
  strategic decisions, 503
  tactical decisions, 503
  types, 502
Decisions, 500
Deep neural networks, 154
Derivative-free stochastic search
  CFA policies, 335
    Bayes greedy, 336, 337
    greedy policy, 335
    Interval estimation, 337
    interval estimation, 337
    UCB policy, 337
  designing policies, 394
Direct lookahead
  multiperiod deterministic, 355
  multiperiod stochastic, 357
  restricted multiperiod, 353
  single period lookahead, 350
direct lookahead, 348
  exogenous state information, 405
  large choice sets, 403
  learning in batches, 380
  policy function
    approximation, 333
  scaling, 396
tuning, 398
VFA policy, 338
  backward ADP, 342
  Gittins indices, 343
VFA-policy
  Beta-Bernoulli, 340
Designing a policy, 631
Diagnostic uncertainty, 561
Direct lookahead
  approximations, 616
Direct lookahead model
  approximation strategies, 980
  horizon truncation, 981
  latent variables, 982
  policy approximation, 984
  sampling, 982
  stage aggregation, 982
lookahead objectives, 985
  chance-constrained, 989
  managing risk, 985
  model discounting, 992
  risk measures, 986
  risk-adjusted policies, 989
  robust optimization, 990
  utility functions, 991
notation, 980
Direct lookahead policy
  approximate dynamic
    programming, 978
  approximate lookahead, 980
backward ADP lookahead policy, 1007
creating lookahead model, 978
deterministic, 1000
discounted, 992
DLA lookahead policy, 1008
fixed lookahead policy, 994
lookahead CFA, 1007
lookahead PFA, 1005
lookahead policy, 977
MCTS, 978, 1009
model predictive control, 977
optimal, 974
policy-within-a-policy, 977
robust optimization, 978
rolling horizon procedure, 977
rollout policy, 978
shortest path, 1001
stochastic programming, 977
why use, 997
Distributional uncertainty, 587
Double-pass ADP algorithm, 876
Dynamic assignment problem, 447
Dynamic program
  budgeting problem, 752
gambling problem, 751
Dynamic programming
  model-free, 516, 875
Episodic, 825
Epistemic uncertainty, 564
Epsilon-greedy, 626
Evaluating policies, 385
  competitive analysis, 391
dynamic regret, 390
  empirical performance, 386
  indifference zone, 392
  opportunity cost, 391
  probability of correction
    selection, 393
quantiles, 386
  static regret, 387, 388
  subset selection, 394
excitation, 627
Excitation policy, 675
Exogenous policy, 65, 506
Exogenous information, 65, 506
  adversarial processes, 513
deterministic, 514
  lagged, 510
  modeling, 506, 511
  scenarios, 509
  supervisory processes, 514
Exogenous learning, 104
Expectation
  compact form, 40
  expanded form, 40
Expected improvement, 351
Experimental uncertainty, 561
Exploration
  Boltzmann exploration, 627
epsilon-greedy, 626
  excitation, 627
  Thompson sampling, 627
Exploration vs. exploitation, 626
Exponential smoothing, 110, 274
Final reward, 42, 187, 206, 329
  asymptotic convergence, 392
  bandit problems, 323
cost function approximation, 702
Index

expectation
  state-dependent, 530, 531
  state-independent, 529
expensive experiments, 349
knowledge gradient, 351, 362, 364
multiarmed bandit, 60
newsvendor, 70
policy search, 668
simulated
  state-dependent, 534, 885
  state-independent, 533
stochastic gradient algorithm, 235
stochastic search, 224
Finite-horizon approximations
  steady state, 915
Fitted value iteration, 804, 823
Fleet management problem, 78
  fleet, 80
  nomadic trucker, 79
Forecasts
  rolling, 75, 497, 526, 635, 707, 717

G
Gambling problem, 751
Games
  backgammon, 492
  chess, 514, 900
  computer Go, 514, 900
  maze, 52
  video, 154
  video games, 514
Gaussian process regression, 117
Gittins indices, 343
  normally distributed rewards, 346

H
Hamiltonian, 48
Hierarchical aggregation, 121
  correlations, 169
  estimation, 129
  estimation, 125
Hyperparameter, 229
Hyperparameters, 587, 589

I
Implementation policy, 208, 337
Inferential uncertainty, 561
Infinite horizon, 755
  policy iteration, 901
  temporal-difference learning, 832
Information acquisition, 456
Interval estimation, 337
Inventory problems, 73
  basic, 439
  batch replenishment, 444
  forecasts, 75
  general, 440
  lagged, 443
  with lags, 75
  without lags, 73
Inverse optimization, 105

J
Jump diffusion, 590

K
k-nearest neighbor, 150
Kalman filter, 293
Kernel regression, 151
Knowledge gradient, 362
  correlated beliefs, 375
  cumulative reward, 369
  final reward, 351, 364
  posterior reshaping, 354
  restricted lookahead, 353
  sampled belief, 370

L
Laboratory experimentation, 458
Lagged information, 510
Lagrangian relaxation, 768
Lasso, 134
Latent variable, 496, 563
Learning
  approximation strategies, 106
  Bayesian
    state variable, 113
bias and variance, 118
deep neural networks, 154
endogenous, 101, 105
exogenous, 101, 104
exponential smoothing, 110
frequentist
state variable, 111
Gaussian process regression, 117
inverse optimization, 105
k-nearest neighbor, 150
kernel regression, 151
linear regression, 132
local polynomial regression, 153
lookup tables, 106
Bayesian, 112
frequentist, 111
maximum likelihood estimation, 141
model-based, 52
model-free, 52
nonlinear parametric models, 140
nonparametric definition, 149
nonparametric models, 107
nonstationary, 159
learning process, 161
transient truth, 160
truth, 159
normal equations, 133
parametric models, 106
linear, 106
nonlinear, 107
sampled belief models, 141
Sherman-Morrison derivation, 168
supervised, 101
support vector machines, 155
Learning policy, 208, 337, 890
Learning process, 161
Learning rate, 235
Least squares policy evaluation, 841, 857
Least squares temporal difference, 840, 841, 856, 903, 905
Leveling algorithm, 938
Linear model
basis functions, 131
geometric view, 850
Lasso, 134
recursive least squares, 136
sparse additive, 134
Linear models, 131
Linear programming method, 766, 767
Linear quadratic regulation, 48, 767
Linear regression, 132
Longstaff and Schwartz, 894
Local polynomial regression, 153
Longstaff and Schwartz, 894
basis functions, 896
Lookahead dynamic programming, 885
Lookahead model, 528
Lookup tables
aggregation, 121
Bayesian updating, 112
state variable, 113
frequentist updating, 111
state variable, 111
hierarchical aggregation, 121, 169
LSPE, 841
single state, 844
LSTD, 840, 841
single state, 844
Markov decision process, 45, 737
average reward, 765
backward dynamic programming, 749
Bellman’s equation, 46, 741
cost-to-go function, 744
curse of dimensionality, 47
finite horizon, 747
infinite horizon, 755
linear programming method, 767
objective function, 740
optimality equation, 741
partially observable, 488
random contributions, 746
state transitions, 739
transition matrix, 745
value function, 744
value iteration
  stopping rule, 757
Martingale model of forecast evolution, 498
Maximum likelihood estimation, 141
MCTS
  backpropagation step, 1014
  expansion step, 1012
  selection step, 1010
  simulation step, 1012
  steps, 1010
Mean reversion, 590
Measure-theoretic view of information, 535
Medical decision making, 81, 457
MMFE, 498
Model
  base model, 528
  lookahead, 528
  nominal, 528
  true, 528
  universal, 467
Model first, then solve, 4, 7, 11, 27, 433, 470, 474, 505, 603
Model predictive control, 62
Model-free dynamic programming, 516, 875
Modeling frameworks, 68
Modeling time, 478
Monte Carlo, 582
Monte Carlo sampling, 582
Monte Carlo simulation, 560
  inverse cumulative method, 585
Monte Carlo tree search, 1009, 1010
  optimistic, 1016
Monte-Carlo simulation, 581
Multiagent systems
  blood management, 1070
    multiple agent, 1073
    communication, 1038
    active observations, 1039
    architecture, 1038
    receiving information, 1039
    sending information, 1039
    signal distortion, 1039
    controlling architecture, 1043
dimensions, 1036
  communication, 1036
  coordination, 1036
  learning, 1036
  the agents, 1036
flu mitigation, 1044
  controllable truth, 1048
designing policies, 1054
  multiarmed bandit, 1056
  neural network policy, 1055
  resource-constrained, 1049
  spatial model, 1049
  static model, 1045
time-varying model, 1047
time-varying with drift, 1047
HVAC controller, 1067
  model, 1067
  policies, 1069
  modeling, 1040
two-agent learning model, 1050
transition functions, 1052
two-agent newsvendor, 1062
types
  controlling, 1037
  environment, 1037
  learning, 1037
  replenishment, 1037
Multiarmed bandit, 57
  active learning, 61, 325
backward ADP, 342
best hitter, 59
best medication, 59
best path, 59
beta-bernoulli belief, 340
Boltzmann exploration, 337
computer science, 321
dynamic programming, 338
flu mitigation, 1056
Gittins, 323
Gittins indices, 343
history, 60
interval estimation, 337
interval estimation policy, 59
nonstationary, 399
objective function, 58
sequential decision problem, 325
state variable, 58
Thompson sampling, 337
transient learning, 401
two-armed, 321
upper confidence bounding,
value function, 50, 734, 907,
value functions, 608, 624
Newsvendor
final reward, 70
Newsvendor problem, 70
cumulative reward, 71
final reward, 70
multidimensional, 72
Nomadic trucker, 79, 434, 491
Nonanticipativity constraint, 57, 201
Nonparametric model
definition, 149
Nonparametric models, 149
k-nearest neighbor, 150
kernel regression, 151
local polynomial regression, 153
Nonparametric statistics
tree regression, 156
Nonstationary learning, 159
learning process, 161
martingale truth, 159
transient truth, 160
Normal equations, 133

Objective function, 66
conditional value at risk, 523
cumulative reward, 206
final reward, 206
performance metrics, 518
probability of correct
selection, 393
regret, 208
robust optimization, 523
static regret, 388
value at risk, 523
Observational uncertainty, 560
Off-policy learning, 890
Offline learning, 882
On-policy learning, 890
Online learning, 883
Opportunity cost, 391
Optimal computing budget allocation, 383
Optimal control, 47
control law, 49
cost-to-go function, 48
Hamiltonian, 48
Lagrangian relaxation, 768
linear decision rule, 769
linear quadratic regulation, 48, 767
state equation, 768
stochastic, 49
transition function, 47
Optimal stopping, 54
Optimality equations proof, 770
Optimistic policy iteration, 903
Optimum-deviation policy, 674

\textbf{p}
Parametric models
linear, 131
nonlinear, 140
Partial policy evaluation, 903
Partially observable Markov decision process, 563, 1059
Partially observable states, 495
Performance elbow, 997
Policy
affine, 612
behavior, 208, 667, 890
Boltzmann, 627
choosing policy class, 638
complexity tradeoffs, 636
cost function approximation, 613
definition, 604
designing, 631
direct lookahead, 616, 974
energy storage problem, 627
CFA, 628
DFA, 629
Hybrid, 629
PFA, 628
VFA, 628
epsilon-greedy, 626
evaluation, 532, 641
examples, 604
excitation, 627
four classes, 209
four classes of policies, 606
hybrid, 620
implementation, 208, 337, 644, 667
learning, 208, 337, 644, 667, 890
linear decision rule, 612
measurability, 538
optimistic, 903
partial policy evaluation, 903
performance-based, 668
policy function approximation, 610
randomized, 626
roll-out heuristic, 1009
sampling the shoulders, 673
soft issues, 644
supervised, 668
target, 208, 667, 890
Thompson sampling, 627
tree search, 1009
tuning, 642
value function approximation, 614
Policy evaluation, 641
infinite horizon, 826
infinite horizon projection, 828
partial simulations, 827
recurrent visits, 826
temporal difference update, 828
Policy function approximation, 659
  affine, 660
  affine policy, 660
  Boltzmann, 659
  contextual, 665
  limitations, 632
  locally linear, 663
  monotone policies, 661
  nonlinear, 662
Policy gradient theorem, 683, 684
  computation, 684
Policy iteration, 763
  hybrid value/policy iteration, 764
  infinite horizon, 901
Policy search
  correlated knowledge
    gradient, 677
  derivative-free methods, 670
  exact derivatives
    continuous, 677
    discrete, 680
    policy gradient theorem, 683
  excitation, 675
  interval estimation, 675
  knowledge gradient, 677
  numerical derivatives, 669
  optimum-deviation, 674
  sampled belief, 676
  supervised learning, 686
Policy-within-a-policy, 620
POMDP, 488, 563, 1059
  belief MDP, 1060
Post-decision state, 79, 490
Pricing, 80
Prognostic uncertainty, 560

q
Q-learning, 875
  algorithm, 875, 876
  augmented state, 493
  deep neural network, 53
  expected rewards, 688
  linear model, 886
  lookup tables, 873
  off-policy, 845
  recursive least squares, 887
  temporal-difference updating, 52, 735, 845
  updating, 51, 277, 615, 735
Quantile distribution, 591

r
Random number seed, 582
Randomized policies, 626, 781
Receding horizon procedure, 63
Recursive least squares, 136
  derivation, 166
  multiple observations, 139
  nonstationary data, 138
  stationary data, 136
Regret, 208
Reinforcement learning, 50
  Q-learning, 52
Resource allocation, 446, 929
  blood management, 450
  dynamic assignment, 447
  exogenous information state, 957
  general resource allocation, 933
  information acquisition, 456
  newsvendor, 929
  two-stage resource allocation, 930
Response surface method, 335
Risk
  robust optimization, 523
Risk measures, 523
  conditional value at risk, 523
  value at risk, 523
Risk operator, 987
Robust optimization, 63, 523
  as a policy, 991
  optimization problem, 991
  robust policy, 64
  uncertainty set, 64
    box constraints, 990
Roll-out heuristic, 1009
Rolling forecast, 635
Rolling forecasts, 497
Rolling horizon procedure, 63

S
Sample average approximation, 194
   convergence rate, 197
Sampled belief models, 141
Sampled models, 193, 595
Sampling the shoulders, 673
SARSA, 875
Scientific exploration, 82
Sequential kriging, 352
Sherman-Morrison formula,
   115, 168
Shortest path
   information collecting, 459
   stochastic, 433
Shortest path problem, 76
   dynamic, 78
   robust, 78
   stochastic, 77
   deterministic, 76
Simulation optimization, 60, 382
   indifference zone algorithm, 383
   OCBA, 60, 383
Simultaneous perturbation
   stochastic approximation
   see SPSA, 240
Soft max, 627
Software, 469
Specification error, 569
SPSA, 240
   mini-batch, 240
   tuning, 722
State variables, 65
   asset pricing, 437
   belief, 408
   belief state, 485, 495
   definition, 481
   dynamic assignment problem, 450
   factored, 498
   flat representation, 498
   gambling problem, 751
information state, 485
   lagged, 490
   latent variable, 496
   physical state, 408, 485
   post-decision state, 79, 490, 491
   pre-decision state, 491
   rolling forecasts, 497
   shortest path, 493
   three types, 485
Stepsizes policies
   apparent convergence, 301
   BAKF proof, 303
   bias-adjusted Kalman filter, 297
   convergence conditions, 276
   deterministic, 276
      constant, 278
      harmonic, 279
      McClain, 280
      polynomial learning rate, 280
      search-then-converge, 281
   guidelines, 301
   infinite horizon
      bounds, 859
   introduction, 233
   optimal
      approx. value iteration, 297
      BAKF, 295
      nonstationary - I, 293
      nonstationary - II, 294
      stationary, 291
   optimal stepsizes, 289
   stochastic, 282
      AdaGrad, 287
      ADAM, 287
      convergence conditions, 283
      Kesten’s rule, 285
      RMSProp, 288
      stochastic gradient adaptive
         stepsize, 286
      Trigg, 286
Stochastic approximation procedure
- see stochastic gradient algorithm, 233
Stochastic decomposition, 949, 951
Stochastic gradient algorithm, 225, 233, 835
as a sequential decision problem, 247
finite differences, 238
for neural networks, 242
gradient smoothing, 237
martingale proof, 256
mini-batches, 240
older proof, 252
second order methods, 237
SPSA, 240
Stochastic modeling
crossing times, 593
electricity prices, 589
energy illustration
crossing times, 593
jump diffusion, 590
mean reversion, 590
quantile distribution, 591
regime shifting, 592
quantile distribution, 591
regime shifting, 592
Stochastic optimal control, 49
Stochastic optimization
asymptotic formulation, 42
cumulative reward, 42
final reward, 42
Stochastic programming, 56
Benders decomposition, 945
interstage independence, 947, 958
lookahead policy, 57
nonanticipativity, 57
nonanticipativity constraint, 57
recourse variables, 56
two-stage, 56
Stochastic search, 40
adaptive learning, 184
cumulative reward, 187, 224
derivative-based, 42
derivative-free, 43
deterministic methods, 184, 188
delayed-reward, 190
Markov decision process, 192
newsvendor, 189
optimal control, 191
shortest path, 189
final reward, 187, 224
newsvendor, 183
sampled approximations, 184
sampled models, 193
cumulative reward, 187, 224
delayed-reward, 187
state-dependent, 187
state-independent, 187
Stopping time, 54
Supervised learning, 101, 900
Support vector machines, 155

\textbf{t}
Target policy, 208, 890
TD(\lambda), 829
TD(0)
recursive least squares, 842
TD(0) updating
slow backward learning, 832
Temporal difference learning, 829
Temporal-difference learning, 828
infinite horizon, 832
Thompson sampling, 337, 627
Three curses of dimensionality, 47, 748
Tic-tac-toe, 898
Time, 478
Trajectory following, 890
Transformer replacement, 435
Transient learning model, 401
knowledge gradient, 402
Transition function, 65, 515
batch replenishment, 445
Transition matrix, 745
Tree regression, 156
Trucking, 491
Two-stage stochastic programming, 1018
Benders’ decomposition, 946

u
Uncertainty
adversarial, 564
aleatoric, 564
algorithmic, 573
algorithmic uncertainty, 561
coarse-grained, 564
communication, 572
communication uncertainty, 561
device control, 571
decision control, 561
diagnostic, 561, 567
distributional, 564
epistemic, 564
experimental, 561, 568
fine-grained, 564
goal, 574
goal uncertainty, 561
implementation, 571
inferential, 561, 567
model, 569
model specification, 569
model uncertainty, 561
observational, 560, 562
political, 574
political uncertainty, 562
prognostic, 560, 564
regulatory, 574
transitional, 571
transitional uncertainty, 561
Uncertainty quantification, 560
Universal modeling framework, 10, 467, 469
base model, 528
compact formulation, 68
decision variables, 65, 500
energy storage illustration, 523
exogenous information, 506
expanded formulation, 65
lookahead model, 528
objective function, 66, 518
risk measures, 523
uncertainty operators, 523
problem classifications, 529
software, 469
state variable, 481
state variables, 65
transition function, 65, 515
Upper confidence bounding, 337

V
Value at risk, 523
Value function approximation, 80
Bayesian learning, 852
correlated beliefs, 853
creating the prior, 855
parametric models, 854
cutting planes, 945
leveling, 938
regression methods, 940
separable piecewise linear, 943
tic-tac-toe, 898
Value function approximations
linear approximation, 955
piecewise linear, 937
Value iteration, 757
algorithmic update, 869
bound, 761
bounds, 760
error bound, 780
Gauss-Seidel variation, 758
monotone convergence, 778
monotonic behavior, 760
pre-decision state, 872
proof of convergence, 774
relative value iteration, 758
stopping rule, 757