

## INDEX

- 1-of- $M$  coding, 247
- $N$ -max-product, 89
- $\alpha$ -expansion, 648
- $\alpha$ -recursion, 496
- $\beta$ -recursion, 498
- $\gamma$ -recursion, 498
  
- absorbing state, 91
- absorption, 104
  - influence diagram, 138
- acceptance function, 599
- active learning, 308
- acyclic, 23
- adjacency matrix, 25
- algebraic Riccati equation, 535
- ancestor, 23
- ancestral ordering, 592
- ancestral sampling, 592
- antifreeze, 273
- approximate inference, 123, 450, 617
  - belief propagation, 637, 639
  - Bethe free energy, 636
  - double integration bound, 650
  - expectation propagation, 639
  - graph cut, 646
  - Laplace approximation, 617
  - switching linear dynamical system, 548
  - variational approach, 618
  - variational inference, 621
- AR model, *see* auto-regressive model
- ARCH model, 526
- artificial life, 576
- asynchronous updating, 624
- asymmetry, 134
- auto-regressive model, 522
  - ARCH, 526
  - GARCH, 526
  - switching, 540
  - time-varying, 524
- auxiliary variable sampling, 601
- average, 166
  
- backtracking, 87
- bag of words, 249, 338
- batch update, 379
- Baum–Welch, 505
- Bayes information criterion, 291
- Bayes’
  - factor, 219, 284
  - model selection, 284
  - theorem, 5
- Bayes’ rule, *see* Bayes’ theorem
- Bayesian
  - decision theory, 314
  - hypothesis testing, 284, 291
  - image denoising, 622
  - linear model, 392, 413
  - mixture model, 442
  - model selection, 284
    - Occam’s razor, 287
  - outcome analysis, 291
- Bayesian Dirichlet score, 220
- BD score, 220
  - BDeu score, 220
- BDeu score, 220
- belief network
  - asbestos-smoking-cancer, 211
  - cascade, 37
  - chest clinic, 53
  - divorcing parents, 33
  - dynamic, 511
  - noisy AND gate, 34
  - noisy logic gate, 34
  - noisy OR gate, 34
  - sigmoid, 282
  - structure learning, 215
  - training
    - Bayesian, 208
- belief propagation, 81, 637
  - loopy, 634
- belief revision, 89, *see* max-product
- Bellman’s equation, 143
- Bessel function, 420
- beta
  - distribution, 173
  - function, 173, 202
- Bethe free energy, 637
- bias, 183
  - unbiased estimator, 183
- bigram, 503
- binary entropy, 624
- binomial coefficient, 171
- binomial options pricing model, 150
- bioinformatics, 513
- black and white sampling, 614
- black-box, 316
- Black–Scholes option pricing, 151
- Blahut–Arimoto algorithm, 633
- Boltzmann machine, 60, 74, 225, 651
  - restricted, 74
- bond propagation, 98
- Bonferroni inequality, 20
- Boolean network, 575
- Bradly–Terry–Luce model, 481
- bucket elimination, 94
- burn in, 597
  
- calculus, 663
- canonical correlation analysis, 354
  - constrained factor analysis, 473
- canonical variates, 362
- causal consistency, 134
- causality, 48, 134
  - do calculus, 51
  - influence diagrams, 51
  - post intervention distribution, 50
- CCA
  - see* canonical correlation analysis, 354
- centring, 179
- chain graph, 67
  - chain component, 68
- chain rule, 664
- chain structure, 87
- change-point model, 560
- checkerboard, 614

- chest clinic, 53  
   missing data, 280  
   with decisions, 157
- children, *see* directed acyclic graph
- Cholesky, 362
- chord, 23
- chordal, 113
- Chow–Liu tree, 222
- classification, 306, 381, 424  
   Bayesian, 401  
   boundary, 246  
   error analysis, 291  
   linear parameter model, 375  
   multiple classes, 382  
   performance, 291  
   random guessing, 300  
   softmax, 382
- clique, 23  
   decomposition, 455  
   graph, 103  
   matrix, 26, 455
- cliquo, 26
- cluster variation method, 638
- clustering, 307
- collaborative filtering, 344
- collider, *see* directed acyclic graph
- commute, 658
- compatibility function, 612
- competition model  
   Bradly–Terry–Luce, 481  
   Elo, 483  
   TrueSkill, 483
- competition models, 482
- concave function, 665
- conditional entropy, 632
- conditional likelihood, 241
- conditional mutual information, 252
- conditional probability, 4
- conditional random field, 234, 510
- conditioning, 179  
   loop cut set, 97
- conjugate distribution, **182**  
   exponential family, 181  
   Gaussian, 187, 189  
   prior, 181
- conjugate gradient, 381
- conjugate gradients algorithm, 670
- conjugate vector, 669
- conjugate vectors algorithm, 669
- connected components, 24
- connected graph, 24
- consistent, 109
- consistent estimator, 182
- control, 627
- convex function, 665
- correction smoother, 498, 532
- correlation  
   matrix, 168
- cosine similarity, 341
- coupled HMM, 511
- covariance, 168  
   matrix, 168
- covariance function, 414, 418  
    $\gamma$ -exponential, 420  
   construction, 419  
   Gibbs, 421  
   isotropic, 420  
   Matérn, 420  
   Mercer kernel, 422  
   neural network, 420  
   non-stationary, 420  
   Ornstein–Uhlenbeck, 420, 422  
   periodic, 420  
   rational quadratic, 420  
   smoothness, 421  
   squared exponential, 419  
   stationary, 419
- CPT, *see* conditional probability table
- CRF, *see* conditional random field
- critical point, 666
- cross-validation, 311
- cumulant, 589
- curse of dimensionality, 145, 373
- cut set conditioning, 96
- cycle, 23
- D-map, *see* dependence map
- DAG, *see* directed acyclic graph
- data  
   anomaly detection, 307  
   categorical, 165  
   dyadic, 454  
   handwritten digits, 381  
   labelled, 306  
   monadic, 455  
   numerical, 165  
   ordinal, 165  
   unlabelled, 306
- data compression, 447  
   vector quantisation, 447
- decision boundary, 376
- decision function, 308
- decision theory, 127, 203, 314
- decision tree, 128
- decomposable, 113
- degree, 26
- degree of belief, 5
- delta function, *see* Dirac delta function  
   Kronecker, 168
- density estimation, 432  
   Parzen estimator, 445
- dependence  
   map, 71
- descendant, 23
- design matrix, 399, 413
- determinant, 659
- deterministic latent variable model, 577
- differentiation, 664
- digamma function, 192
- digit data, 333
- Dijkstra’s algorithm, 92
- dimension reduction  
   linear, 329, 333  
   supervised, 359
- dimensionality reduction  
   linear, 336  
   non-linear, 336
- Dirac delta function, 168, 169
- directed acyclic graph, 37  
   ancestor, 23  
   ancestral order, **25**  
   cascade, 37  
   children, 23  
   collider, 41  
   descendant, 23  
   family, 23  
   immorality, 46  
   Markov blanket, 23  
   moralisation, 65  
   parents, 23
- direction bias, 510
- directional derivative, 665
- Dirichlet  
   distribution, 175
- Dirichlet process mixture models, 450
- discount factor, 143
- discriminative approach, 314, 316
- discriminative training, 316, 508
- dissimilarity function, 322
- distributed computation, 568
- distribution, 3  
   Bernoulli, 171, 436  
   beta, 173, 175, 192, 210, 286  
   binomial, 171  
   categorical, 171  
   change of variables, 166, 190  
   conjugate, 187  
   continuous, 5  
   density, 5  
   Dirichlet, 175, 193, 212, 218, 450, 451  
   discrete, 3  
   divergence, 169  
   double exponential, 174

- empirical, 169, 310
  - average, 166
  - expectation, 166
- exponential, 172
- exponential family, 180
  - canonical, 180
- gamma, 443
  - mode, 193
- Gauss-gamma, 188, 194
- Gauss-inverse-gamma, 188
- Gaussian
  - canonical exponential form, 180
  - conditioning, 179
  - conjugate, 187, 189
  - entropy, 180
  - mixture, 195
  - multivariate, 176
  - normalisation, 190, 191
  - partitioned, 178
  - propagation, 178
  - univariate, 174
- inverse gamma, 173
- inverse Wishart, 189
- isotropic, 177
- joint, 4
- kurtosis, 168
- Laplace, 174
- marginal, 4
- mode, 167
- multinomial, 172
- normal, 174
- Poisson, 172, 195
- Polya, 450
- scaled mixture, 175
- skewness, 168
- Student's  $t$ , 175
- uniform, 171, 172
- Wishart, 443
- domain, 3
- double integration bound, 650
- dual parameters, 374
- dual representation, 373
- dyadic data, 454
- dynamic Bayesian network, 511
- dynamic synapses, 580
- dynamical system
  - linear, 520
  - non-linear, 576
- dynamics reversal, 533
- early stopping, 381
- edge list, 25
- efficient IPF, 231
- eigen
  - decomposition, 177, 332, 662
- equation, 332
- function, **662**
- problem, 355
- spectrum, 422
- value, 332, **660**
- Elo model, 482
- emission distribution, 495
- emission matrix, 527
- empirical distribution, **169**, 185, 310
- empirical independence, 218
- empirical risk, 310
  - penalised, 311
- empirical risk minimisation, 311
- energy, 260
- entropy, 117, **170**, 195, 260, 632
  - differential, 170, 196
  - Gaussian, 180
- EP, *see* expectation propagation
- error function, 376
- estimator
  - consistent, 182
- evidence, *see* marginal likelihood
  - hard, 34
  - likelihood, 36
  - soft, 34
  - uncertain, 34
  - virtual, 36
- evidence procedure, 395
- exact sampling, 593
- expectation, *see* average
- expectation correction, 554
- expectation maximisation, 161, **260**, 346, 396, 504, 537
  - algorithm, **260**, 261
  - antifreeze, 273
  - belief networks, 264
  - E-step, 261
  - energy, 260
  - entropy, 260
  - failure case, 272
  - intractable energy, 270
  - M-step, 261
  - mixture model, 434
  - partial E-step, 270
  - partial M-step, 270
  - variational Bayes, 275
  - Viterbi training, 271
- expectation propagation, 639
- exponential family, 180
  - canonical form, 180
  - conjugate, 181
- extended observability matrix, 539
- face model, 469
- factor analysis, 465
- factor rotation, 464
- probabilistic PCA, 472
- training
  - EM, 468
  - SVD, 466
- factor graph, 69
- factor loading, 462
- family, *see* directed acyclic graph
- feature map, 352
- filtering, 496
- finance, 148
  - optimal investment, 151
  - option pricing, 148
- finite-dimensional Gaussian process, 414
- Fisher information, 193
- Fisher's linear discriminant, 360
- Floyd–Warshall–Roy algorithm, 94
- forward sampling, 593
- forward-backward, 498
- forward-filtering-backward-sampling, 500
- forward-sampling-resampling, 612
- gamma
  - digamma, 192
  - distribution, 173
  - function, 173, 192
- GARCH model, 526
- Gaussian
  - canonical representation, 176
  - distribution, 174
  - moment representation, 176, 530
  - sub, 168
  - super, 168
- Gaussian mixture model, 439
  - Bayesian, 442
  - collapsing the mixture, 553
  - EM algorithm, 439
  - infinite problems, 442
  - k-means, 446
  - Parzen estimator, 445
  - symmetry breaking, 443
- Gaussian process, 412
  - classification, 424
  - Laplace approximation, 425
  - multiple classes, 428
  - regression, 416
  - smoothness, 421
  - weight space view, 413
- Gaussian sum filtering, 549
- Gaussian sum smoothing, 554
- generalisation, 305, 310
- generalised pseudo Bayes, 557

- generative
  - approach, 315
  - model, 315
  - training, 315
- generative approach, 314
- Gibbs sampling, 594
- Glicko, 483
- GMM, *see* Gaussian mixture model
- Google, 491
- gradient, 663
  - descent, 666
- Gram matrix, 374
- Gram–Schmidt procedure, 669
- Gramm matrix, 353
- graph, 22
  - acyclic, 23
  - adjacency matrix, **25**
  - chain, 67
  - chain structured, 87
  - chord, 23
  - chordal, 113
  - clique, 23, 26, 103
  - clique matrix, 26
  - cliquo, 26
  - connected, 24
  - cut, 646
  - cycle, 23
  - decomposable, 113
  - descendant, 25
  - directed, 22
  - disconnected, 118
  - edge list, 25
  - factor, 69
  - loop, 23
  - loopy, 24
  - multiply connected, 24, 111
  - neighbour, 23
  - path, 23
  - separation, 65
  - set chain, 122
  - singly connected, 24
  - skeleton, 46
  - spanning tree, 24
  - tree, 24
  - triangulated, 113
  - undirected, 22
  - vertex
    - degree, 26
- graph cuts algorithm, 646
- graph partitioning, 453
- Gull–MacKay iteration, 397
- Hamilton–Jacobi equation, 143n3
- Hamiltonian dynamics, 602
- Hammersley–Clifford theorem, 62
- handwritten digits, 381
- Hankel matrix, 539
- harmonium, *see* restricted Boltzmann machine
- Heaviside step function, 456
- Hebb, 570
- Hebb rule, 570
- hedge fund, 299
- Hermitian, 658
- Hessian, 381, 664
- hidden Markov model, 117, 271, **495**
  - $\alpha$  recursion, 496
  - $\beta$  recursion, 497
  - coupled, 511
  - direction bias, 510
  - discriminative training, 508
  - duration model, 508
  - entropy, 117
  - filtering, 496
  - input–output, 509
  - likelihood, 499
  - most likely state (MAP), 500
  - pairwise marginal, 499
  - Rauch–Tung–Striebel smoother, 498
  - smoothing
    - parallel, 497
    - sequential, 498
  - viterbi, 89
  - Viterbi algorithm, 500
- hidden variables, 256
- HMM, *see* hidden Markov model
- Hopfield network, 568
  - augmented, 578
  - capacity, 573
  - Hebb rule, 570
  - heteroassociative, 578
  - maximum likelihood, 572
  - perceptron, 573
  - pseudo inverse rule, 571
  - sequence learning, 569
- hybrid Monte Carlo, 602
- hyper Markov, 238
- hyper tree, 117
- hyperparameter, 181, 211, 395
- hyperplane, 657
- hypothesis testing, 284
  - Bayesian error analysis, 291
- i.i.d., *see* identically and independently distributed
- I-map, *see* independence map
- ICA, 475
- identically and independently distributed, 199
- identifiability, 537
- identity matrix, 658
- IM algorithm, *see* information-maximisation algorithm
- immorality, 46
- importance
  - distribution, 607
  - sampling, 607
    - particle filter, 610
    - resampling, 609
    - sequential, 609
  - weight, 607
- incidence matrix, 26
- independence
  - Bayesian, 218
  - conditional, 7, 30
  - empirical, 218
  - map, 71
  - Markov equivalent, 47
  - mutual information, 171
  - naive Bayes, 243
  - parameter, 208
  - perfect map, 72
- independent components analysis, 477
- indicator function, 16
- indicator model, 449
- induced representation, 112
- inference, 77
  - bond propagation, 98
  - bucket elimination, 94
  - causal, 48
  - cut set conditioning, 96
  - HMM, 496
  - linear dynamical system, 528
  - MAP, 98
  - marginal, 77
  - Markov decision process, 146
  - max-product, 86
  - message passing, 77
  - mixed, 94
  - MPM, 98
  - sum-product algorithm, 83
  - transfer matrix, 79
  - variable elimination, 77
- influence diagram, 131
  - absorption, 138
  - asymmetry, 134
  - causal consistency, 134
  - chest clinic, 157
  - decision potential, 137
  - fundamental link, 134
  - information link, 131
  - junction tree, 137
  - no forgetting principle, 133
  - partial order, 132
  - probability potential, 137

- solving, 136
- utility, 132
- utility potential, 137
- information link, 131
- information maximisation, 633
- information retrieval, 340, 491
- information-maximisation algorithm, 633
- innovation noise, 522
- input–output HMM, 509
- inverse modus ponens, 12
- IPF, 226, *see* iterative proportional fitting
  - efficient, 231
- Ising model, 66, *see* Markov network, 98
  - approximate inference, 621
- isotropic, 177, 441
- isotropic covariance functions, 419
- item response theory, 480
- iterated conditional modes, 643
- iterative proportional fitting, 226
- iterative scaling, 232
  
- Jeffrey’s rule, 34
- Jensen’s inequality, 651, 666
- Joseph’s symmetrised update, 532
- jump Markov model, *see* switching linear dynamical system
- junction tree, 106, 111
  - absorption, 104
  - algorithm, 102, 115
  - clique graph, 103
  - computational complexity, 117
  - conditional marginal, 119
  - consistent, 109
  - hyper tree, 117
  - influence diagram, 137
  - marginal likelihood, 118
  - most likely state, 121
  - normalisation constant, 117
  - potential, 103
  - running intersection property, 107
  - separator, 103
  - strong, 138
  - strong triangulation, 138
  - tree width, 117
  - triangulation, 112
- k-means, 446
- Kalman filter, 527n2
- Kalman gain, 531
- KD-tree, 323
  
- kernel, 373, 374, *see* covariance function
  - classifier, 382
- kidnapped robot, 502
- Kikuchi, 638
- KL divergence, *see* Kullback–Leibler divergence
- KNN, *see* nearest neighbour
- Kronecker delta, 168, 658
- Kullback–Leibler divergence, 169, 260, 619
- kurtosis, 168
  
- labelled data, 306
- Lagrange
  - dual, 674
  - multiplier, 673
- Lagrangian, 673
- Laplace approximation, 403, 425, 617
- latent ability model, 479
- latent Dirichlet allocation, 451
- latent linear model, 462
- latent semantic analysis, 338
- latent topic, 339
- latent variable, 256
  - deterministic, 577
  - model, 256
- lattice model, 65
- LDA regularised, 366
- LDS, *see* linear dynamical system
- leaky integrate and fire model, 581
- Leapfrog discretisation, 603
- learning
  - active, 308
  - anomaly detection, 307
  - Bayesian, 200
  - belief network, 204
  - belief networks
    - EM, 264
  - Dirichlet prior, 212
  - inference, 199
  - nearest neighbour, 322
  - online, 307
  - query, 308
  - reinforcement, 308
  - semi-supervised, 308, 447
  - sequences, 504, 536, 569
  - sequential, 307
  - structure, 215
  - supervised, 305
  - unsupervised, 306, 462
- learning rate, 379
- likelihood, 118, **182**, 499, 534, 553
  - bound, 260
  - marginal, 84, 182
  - model, 182, 215
    - approximate, 290
    - pseudo, 237
  - likelihood decomposable, 215
  - line search, 667, 668
  - linear algebra, 655
  - linear dimension reduction, 329, 333
    - canonical correlation analysis, 354
    - latent semantic analysis, 338
    - non-negative matrix factorisation, 349
  - probabilistic latent semantic analysis, 345
    - supervised, 359
    - unsupervised, 329
  - linear discriminant analysis, 359, 360
    - as regression, 365
    - penalised, 365
    - regularised, 365
  - linear dynamical system, 179, 527
    - cross moment, 534
    - dynamics reversal, 533
    - filtering, 530
    - identifiability, 537
    - inference, 528
    - learning, 536
    - likelihood, 534
    - most likely state, 535
    - numerical stability, 529
    - Riccati equations, 535
    - smoothing, 532
    - subspace method, 538
    - switching, 547
    - symmetrising updates, 532
  - linear Gaussian state space model, 527
  - linear model, 367
    - Bayesian, 392
    - classification, 375
    - factor analysis, 465
    - latent, 462
    - regression, 368
  - linear parameter model, 368
    - Bayesian, 393
  - linear perceptron, 378
  - linear separability, 377
  - linear transformation, 658
  - linearly separable, 377
  - Linsker’s as-if-Gaussian approximation, 634
  - localisation, 501
  - logic
    - Aristotle, 11
  - logistic regression, 376
  - logistic sigmoid, 376
  - logit, 375

- loop, 23
- loop cut set, 97
- loopy, 24
- loss function, 308, 309
  - squared, 310
  - zero-one, 309
- loss matrix, 309
- Luenberger expanding subspace theorem, 669
  
- Mahalanobis distance, 322, 357
- manifold
  - linear, 329
  - low dimensional, 329
- MAP, *see* most probable a posteriori
- MAR, *see* missing at random
- margin, 383, 384
  - soft, 385
- marginal, 4
  - generalised, 138
- marginal likelihood, 84, 118, 182, 395, 418
  - approximate, 396, 402, 404, 427
- marginalisation, 4
- Markov
  - chain, 281, **489**
    - first-order, 521
    - stationary distribution, 491
  - equivalent, 47
  - global, 61
  - hyper, 238
  - local, 61
  - model, 489
  - pairwise, 61
  - random field, 233
    - approximation, 621
- Markov blanket, *see* directed acyclic graph, 594
- Markov chain, 77, 147, 598
  - absorbing state, 91
  - PageRank, 491
- Markov chain Monte Carlo, **598**
  - auxiliary variable, 601
  - hybrid Monte Carlo, 602
  - slice sampling, 606
  - Swendson–Wang, 603
  - Gibbs sampling, 595
  - Metropolis–Hastings, 599
  - proposal distribution, 599
  - structured Gibbs sampling, 596
- Markov decision process, 141, 158
  - Bellman’s equation, 143
  - discount factor, 143
  - partially observable, 153
  - planning, 146
  - policy iteration, 144
  - reinforcement learning, 154
  - temporally unbounded, 143
  - value iteration, 144
- Markov equivalence, 46
- Markov network, 59
  - continuous-state temporal, 520
  - discrete-state temporal, 489
  - Gibbs distribution, 59
  - Hammersley Clifford theorem, 62
  - pairwise, 60
  - potential, 59
- Markov random field, 62, 652, 653
  - alpha-expansion, 648
  - attractive binary, 645
  - graph cut, 646
  - map, 643
  - Potts model, 648
- matrix, **657**
  - adjacency, 25
  - block inverse, 660
  - Cholesky, 362
  - clique, 26
  - Gramm, 353
  - Hankel, 539
  - incidence, 26
  - inversion, 659
  - inversion lemma, 660
  - orthogonal, 659
  - positive definite, **662**
  - pseudo inverse, 660
  - rank, 659
- matrix factorisation, 349
- max-product, 86
  - $N$  most probable states, 89
- max-sum, 91
- maximum cardinality checking, 115
- maximum likelihood, 15, 182, 183, 185, 378
  - belief network, 204
  - Chow–Liu tree, 223
  - counting, 205
  - empirical distribution, 185
  - factor analysis, 465
  - Gaussian, 186
  - gradient optimisation, 279
  - Hopfield network, 572
  - Markov network, 224
  - ML-II, 203
  - naive Bayes, 244
  - properties, 183
- MCMC, *see* Markov chain Monte Carlo
- mean-field theory, 625
  - asynchronous updating, 626
- Mercer kernel, 422
- message
  - passing, 106
  - schedule, 106
- message passing, 77
- Message schedule, 83
- Metropolis–Hastings acceptance function, 599
- Metropolis–Hastings sampling, 599
- minimum clique cover, 456
- missing at random, 257
  - completely, 258
- missing data, 256
- mixed inference, 94
- mixed membership model, 451
- mixing matrix, 475
- mixture
  - Gaussian, 195
- mixture model, 432
  - Bernoulli product, 436
  - Dirichlet process mixture, 450
  - expectation maximisation, 434
  - factor analysis, 469
  - Gaussian, 439
  - indicator approach, 449
  - Markov chain, 492
  - PCA, 469
- mixture of experts, 448
- MN, *see* Markov network
- mode, 167
- model
  - auto-regressive, 522
  - change-point, 560
  - deterministic latent variable, 577
  - faces, 469
  - leaky integrate and fire, 581
  - linear, 367
  - mixed membership, 451
  - mixture, 432
  - Rasch, 479
- model selection, **284**
  - approximate, 290
- moment representation, 530
- momentum, 667
- monadic data, 455
- money
  - financial prediction, 299
  - loads, 299
- moralisation, 65, 110
- most probable a posteriori, 15, 182
- most probable path
  - multiple-source multiple-sink, 93
- most probable state
  - $N$  most probable, 89
- MRF, *see* Markov random field

- multiply connected, 24
- multiply connected-distributions, 111
- mutual information, 171, 223, 632
  - approximation, 632
  - conditional, 171
  - maximisation, 633
- naive Bayes, 243, 436n1
  - Bayesian, 249
  - tree augmented, 251
- naive mean field, 625
- naive mean field theory, 623
- nearest neighbour, 322
  - probabilistic, 325
- network flow, 652
- network modelling, 351
- neural computation, 568
- neural network, 390, 579
  - depression, 581
  - dynamic synapses, 580
  - leaky integrate and fire, 581
- Newton update, 381
- Newton's method, 671
- no forgetting principle, 133
- node
  - extremal, 88
  - simplicial, 88
- non-negative matrix factorisation, 349
- normal distribution, 174
- normal equations, 369
- normalised importance weights, 607
- observed linear dynamical system, 520
- Occam's razor, 287
- One of  $m$  encoding, 165
- online learning, 307
- optimisation, 207, 666
  - Broyden–Fletcher–Goldfarb–Shanno, 672
  - conjugate gradients algorithm, 670
  - conjugate vectors algorithm, 669
  - constrained optimisation, 673
  - critical point, 666
  - gradient descent, 666
  - Luenberger expanding subspace theorem, 669
  - Newton's method, 671
  - quasi Newton method, 672
- option pricing, 148
- ordinary least squares, 367
- orthogonal, 655
- orthogonal least squares, 368
- orthonormal, 655
- outlier, 387
- over-complete representation, 345
  - probabilistic, 472
- overcounting, 103
- overfitting, 235, 288, 417
- PageRank, 491
- pairwise comparison models, 482
- pairwise Markov network, 60
- parents, *see* directed acyclic graph
- part-of-speech tagging, 513
- Partial least squares, 475
- partial order, 132
- partially observable MDP, 153
- particle filter, 609
- partition function, 59, 651
- partitioned matrix
  - inversion, 191
- Parzen estimator, 325, 445
- path
  - blocked, 43
- PC algorithm, 216
- PCA, *see* principal components analysis
- perceptron, 378
  - logistic regression, 378
- perfect elimination order, 114
- perfect map, *see* independence
- perfect sampling, 593
- planning, 146
- plant monitoring, 307
- plate, 200
- Poisson distribution, 172, 195
- policy, 144
  - iteration, 144
- Polya distribution, 450
- POMDP, *see* partially observable MDP
- positive definite
  - matrix, 414
  - parameterisation, 458
- posterior, 182
  - Dirichlet, 213
- potential, 59
- Potts model, 648
- precision, 176, 188, 393
- prediction
  - auto-regression, 522
  - financial, 299
  - non-parametric, 412
  - parameteric, 412
- predictive variance, 394
- predictor-corrector, 497
- Principal components analysis, **329**, 336
  - algorithm, 333
  - high-dimensional data, 336
  - kernel, 352
  - latent semantic analysis, 338
  - missing data, 341
- principal directions, 332
- printer nightmare, 239
  - missing data, 281
- prior, 182
- probabilistic latent semantic analysis, 345
  - conditional, 348
  - EM algorithm, 346
- probabilistic PCA, 472
- probability
  - conditional, 4
  - function, 206
  - density, 4
  - frequentist, 5
  - posterior, 182
  - potential, 137
  - prior, 182
  - subjective, 5
- probit, 375
- probit regression, 376
- projection, 656
- proposal distribution, 599
- pseudo inverse, 570
  - Hopfield network, 571
- pseudo likelihood, 237
- quadratic form, 662
- quadratic programming, 384
- query learning, 308
- questionnaire analysis, 479
- radial basis functions, 371
- Raleigh quotient, 362
- Random Boolean networks, 576
- Rasch model, 479
  - Bayesian, 481
- Rauch–Tung–Striebel, 498
- reabsorption, 122
- region graphs, 638
- regression
  - linear parameter model, 368
- regression, 306, 413
  - logistic, 376
- regularisation, 288, 311, 371, 381
- reinforcement learning, 154, 308
- rejection sampling, 590
- relevance vector machine, 400, 408
- reparameterisation, 102
- representation
  - dual, 373
  - over-complete, 345
  - sparse, 345
  - under-complete, 344

- resampling, 609
- reset model, 560
- residuals, 368
- resolution, 11
- responsibility, 441
- restricted Boltzmann machine, 74
- Riccati equation, 535
- ridge regression, 371
- risk, 309
- robot arm, 627
- robust classification, 387
- Rose–Tarjan–Lueker elimination, 114
- running intersection property, 107, 109
- sample
  - mean, 169
  - variance, 169
- sampling, 450
  - ancestral, 592
  - forward-filtering-backward-sampling, 499, 533
  - Gibbs, 594
  - importance, 607
  - multi variate, 591
  - particle filter, 610
  - rejection, 590
  - univariate, 588
- sampling importance resampling, 609
- scalar product, 655
- scaled mixture, 175
- search engine, 491
- self localisation and mapping, 503
- semi-supervised learning, 308
  - lower dimensional representations, 317
- separator, 103
- sequential importance sampling, 609
- sequential minimal optimisation, 386
- set chain, 122
- Shafer–Shenoy propagation, 120
- shortest path, 91
- shortest weighted path, 92
- sigmoid
  - logistic, 376
- sigmoid belief network, 282
- sigmoid function
  - approximate average, 405
- simple path, 92
- simplicial nodes, 114
- Simpson’s paradox, 48
- singly connected, 24
- singular value decomposition, 337, **662**
  - thin, 363, 662
- skeleton, 46, 216
- skewness, 168
- slice sampling, 606
- smoothing, 497
- softmax function, 381, 448
- spam filtering, 249
- spanning tree, 24
- sparse representation, 345
- spectrogram, 525
- speech recognition, 512
- spike response model, 579
- squared Euclidean distance, 322
- squared loss, 310
- standard deviation, 167
- standard normal distribution, 174
- stationary, 599
  - distribution, 80
- stationary Markov chain, 490
- stop words, 453
- strong junction tree, 137
- strong triangulation, 138
- structure learning, 215
  - Bayesian, 220
  - network scoring, 220
  - PC algorithm, 216
  - undirected, 238
- structured expectation propagation, 643
- subsampling, 597
- subspace method, 538
- sum-product, 634
- sum-product algorithm, 83
- supervised learning, **305**
  - semi, 308
    - classification, 306
    - regression, 306
- support vector machine, 383
  - chunking, 386
  - training, 387
- support vectors, 385
- SVD, *see* singular value decomposition
- SVM, *see* support vector machine
- Swendson–Wang sampling, 603
- switching AR model, 540
- switching kalman filter, *see* switching linear dynamical system
- switching linear dynamical system, 547
  - changepoint model, 560
  - collapsing Gaussians, 553
  - expectation correction, 554
  - filtering, 549
  - Gaussian sum smoothing, 554
  - generalised pseudo Bayes, 557
  - inference
    - computational complexity, 548
  - likelihood, 553
  - smoothing, 555
- symmetry breaking, 443
- tagging, 513
- tall matrix, 344
- term-document matrix, 338
- test set, 305
- text analysis, 350, 451
  - latent semantic analysis, 338
  - latent topic, 339
  - probabilistic latent semantic analysis, 345
- TF-IDF, 338
- time-invariant LDS, 535
- Tower of Hanoi, 145
- trace-log formula, 661
- train set, 305
- training
  - batch, 379
  - discriminative, 316
  - generative, 315
  - generative-discriminative, 316
  - HMM, 504
  - linear dynamical system, 536
  - online, 380
- transfer matrix, 79
- transition distribution, 495
- transition matrix, 520, 527
- tree, 24, 78
  - Chow–Liu, 222
- tree augmented network, 252
- tree width, 117
- triangulation, 112, 113
  - check, 116
  - greedy elimination, 114
  - maximum cardinality, 115
  - strong, 138
  - variable elimination, 114
- TrueSkill, 483
- uncertainty, 170
- under-complete representation, 344
- undirected graph, 22
- undirected model
  - learning
    - hidden variable, 279
    - latent variable, 279
- uniform distribution, 172
- unit vector, 655
- unlabelled data, 307
- unsupervised learning, 306, 462
- utility, 127, 308
  - matrix, 309
  - money, 127

## Index

697

- potential, 137
- zero-one loss, 309
- validation, 311
  - cross, 311
- value, 142, 143
- value iteration, 144
- variable
  - hidden, 260
  - missing, 260
  - visible, 260
- variable elimination, 77
- variance, 167
- variational approximation
  - factorised, 623
  - structured, 625, 627
- variational Bayes, **273**
  - expectation maximisation, 275
- varimax, 464
- vector algebra, 655
- vector quantisation, 447
- visualisation, 356
- Viterbi, 89, 271
- Viterbi algorithm, 500
- Viterbi alignment, 496
- Viterbi training, 271
- volatility, 525
- Voronoi tessellation, 322
- web
  - modelling, 351
- website, 351
  - analysis, 491
- whitening, 179, 195
- Woodbury formula, 660
- XOR function, 377
- zero-one loss, 309, 387