
ALPHABETICAL INDEX

- F -distribution, 636
 χ^2 -distribution, 632
 σ field generation, 500
 σ -algebra, 118, 500
 σ -field, 118, 500
 t -distribution, 634
Adaboost, 1426, 1430
AdaDelta, 1532
AdaGrad, 1532
Adam, 1535
adapted process, 1051
advantage function, 1700
affine combination, 428
affine independence, 431
affine set, 428
Akaike's Information Criterion(AIC), 866
AlexNet, 1562
algebraic multiplicity, 189
almost everywhere convergence, 125
alternating series, 32
analytic function, 94
approximation, 96
AR process, 1176
AR(1) process, 1176
arc length, 70
Armijo linesearch, 333
autoregressive conditional
 heteroscedastic model, 1224, 1225,
 1236
Autoregressive process, 1176
autoregressive process, 1176
backpropagation through time, 1582
Bagging, 1420
bagging, 1421
Barycentric coordinate system, 432
basic antithetic variates, 938
basis, 156
Bayes' theorem, 511
Bayesian Information Criterion(BIC), 867
Bayesian statistical model, 695
Bernoulli distribution, 604
Bessel's inequality, 293
best linear unbiased estimator, 828
Beta distribution, 627
bilinear form, 245
binomial distribution, 608
Binomial NBC, 1366
BLUE, 828
Bootstrap method, 698
Bootstrap sample, 698
Borel algebra, 120, 121, 503
Borel measurable function, 504
Borel set, 120, 503
breakdown point, 670
canonical correlation analysis, 744
Cauchy integral Remainder Theorem, 98
Cauchy point, 340
Cauchy sequence, 17
Cauchy-Schwarz inequality, 280, 563
CBOW, 1549
central limit theorem, 573
chain rule, 803
change of basis, 166
Chapman-Kolmogorov Equation, 1117
characteristic equation, 954

- Chebychev's inequality, [560](#)
- Chi-square goodness-of-fit test, [706](#)
- Cholesky decomposition, [222](#)
- closed set, [49](#)
- closure, [43](#)
- closure point, [43](#)
- Cochran's theorem, [654](#)
- coefficient of determination, [850](#)
- coefficient of multiple correlation, [850](#)
- coercive function, [68](#)
- collinearity, [870](#)
- compact, [48](#)
- compact set, [48](#)
- complementary subspace, [173](#), [283](#)
- completeness axiom, [11](#)
- completing square, [233](#)
- complex eigenvector, [192](#)
- concave function, [442](#)
- condition number, [151](#), [207](#)
- conditional entropy, [596](#), [1394](#)
- conditional expectation, [544](#)
- conditional independence, [524](#)
- conditional variance identity, [535](#)
- confidence interval, [715](#)
- congruence, [201](#)
- conjugate directions, [345](#)
- conjugate prior, [697](#)
- conjugate transpose, [200](#)
- consecutive coin toss, [1138](#)
- consistence, [144](#)
- constant sum game, [475](#)
- constraint qualification, [383](#), [391](#)
- continuous bags of words, [1549](#)
- continuous function, [61](#)
- continuously differentiable function, [82](#)
- contraction mapping, [274](#)
- control variate, [941](#)
- controllability, [1629](#)
- convergence almost surely, [566](#)
- convergence in distribution, [569](#)
- convergence in probability, [567](#)
- convex function, [442](#)
- convex set, [434](#)
- convolution formula, [529](#)
- convolution theorem, [307](#)
- Cook's distance, [888](#)
- correlation vs. causation, [809](#)
- cosine similarity kernel, [1287](#)
- Courant-Fisher theorem, [199](#)
- covariance, [533](#)
- covariance matrix, [1214](#)
- Cramer-Rao lower bound, [685](#), [687](#)
- cross-correlation matrix, [1214](#)
- cumulative distribution function, [519](#)
- curvature, [71](#)
- curvature of a curve, [71](#)
- curve, [68](#)
- curve length, [70](#)

- DDPG, [1706](#)
- Deep deterministic policy gradient
 algorithm, [1706](#)
- deep Q learning, [1697](#)
- Deep Recurrent Q network, [1701](#)
- DeepAR, [1595](#)
- delta method, [535](#)
- DeMorgan's Law, [5](#)
- dense set, [17](#)
- dense subset, [10](#)
- descent direction, [324](#)
- detailed balance, [1152](#)
- diagonalizable matrix, [190](#)
- Dickey-Fuller ρ test, [1195](#)
- Dickey-Fuller t test, [1195](#)
- difference equation, [950](#)
- differential function, [82](#)
- diffusion map, [1490](#)
- dimensionality, [156](#)
- direct sum, [154](#), [155](#)
- discrete Fourier transform matrix, [308](#)
- discrete-time white noise, [1040](#)
- double Q learning, [1700](#)

- DQN, [1697](#)
- DRQN, [1701](#)
- dual function, [450](#)
- dual problem, [1301](#)
- duality theory, [450](#)
- Durbin-Watson test, [881](#)
- dynamic programming, [1639](#)
- dynamic programming operator, [1642](#)
- efficient estimator, [688](#)
- Elastic regression, [1304](#)
- elementary matrix, [1750](#)
- empirical conditional entropy, [1396](#)
- empirical entropy, [1394](#), [1396](#)
- entropy, [590](#), [1393](#)
- envelope theorem, [398](#)
- epigraph, [444](#)
- equality constrained linear programming, [404](#)
- equality constrained nonlinear optimization, [381](#)
- equilibrium point, [1004](#), [1005](#)
- equivalent norm, [271](#)
- equivalent relation, [6](#)
- ergodic chain, [1131](#)
- ergodic state, [1131](#)
- error propagation, [1017](#)
- event, [506](#)
- expectation, [532](#)
- Expectation Maximization algorithm, [1503](#)
- exponential distribution, [623](#)
- exponential martingale, [1054](#)
- exponentially weighted moving averages, [1171](#)
- extended Kalman filter, [1020](#)
- factor model, [788](#)
- factorization theorem, [692](#), [805](#)
- Farkas' lemma, [370](#), [392](#), [440](#), [441](#)
- feasible path, [381](#), [389](#)
- feature, [1285](#)
- feature map, [1285](#)
- feature space, [1285](#)
- filtration, [502](#)
- First Fundamental Theorem of Calculus, [109](#)
- Fisher information, [680](#)
- Fisher linear discriminant, [1338](#)
- fixed point, [274](#)
- fixed point theorem, [275](#)
- Fourier transform, [304](#)
- Frisch-Waugh-Lovell theorem, [835](#)
- Frobenius norm, [208](#)
- Frobenius-Perron, [1140](#)
- Frobnius norm, [206](#), [1450](#)
- function, [8](#)
- Fundamental counting principle, [576](#)
- Fundamental lemma of recursive least squares, [1014](#)
- fundamental theorem of algebra, [263](#)
- fundamental theorem of linear algebra, [171](#)
- fundamental theorem of linear maps, [162](#)
- fundamental theorem of ranks, [171](#)
- Gambler's ruin, [1156](#)
- Gamma distribution, [624](#)
- GAN, [1609](#)
- GARCH, [1236](#)
- gated recurrent unit, [1587](#)
- Gauss-Markov theorem, [828](#)
- Gauss-Newton method, [351](#)
- Gaussian kernel, [1288](#)
- Gaussian mixture model, [1505](#)
- Gaussian NBC, [1371](#)
- Gaussian process, [1031](#)
- generalized eigenspace, [215](#)
- generalized eigenvector, [212](#), [214](#)
- generalized least square, [856](#)
- Generative adversarial network, [1609](#)
- generic additive model algorithm, [1432](#)
- geometric Brownian motion, [1047](#)

- geometric distribution, [606](#)
- geometric multiplicity, [189](#)
- geometric SDE, [1086](#), [1089](#)
- Gerschgorin Taussky theorem, [237](#)
- Gerschgorin theorem, [236](#)
- global minimum, [316](#)
- GMM, [1505](#)
- goodness-of-fit test, [706](#)
- Gram matrix, [1286](#)
- Gram-Schmidt process, [292](#)
- Gramian matrix, [232](#)
- GRU, [1587](#)
- halfspace, [1348](#)
- Hamilton-Jacobi-Bellman (HJB) equation, [1637](#)
- Heavy-tailed distribution, [638](#)
- Hermitian matrix, [199](#)
- heteroskedasticity, [875](#)
- high leverage point, [885](#)
- Hilbert space, [280](#)
- Holder's inequality, [562](#)
- hypergeometric distribution, [626](#)
- hyperplane, [1348](#)
- hypothesis, [1249](#)
- hypothesis space, [1249](#)
- idempotent, [174](#)
- importance sampling, [923](#)
- improper integral, [115](#)
- independence, [516](#)
- induced norm, [281](#)
- inequality constrained nonlinear programming, [389](#)
- inertia, [201](#)
- inertia matrix, [201](#)
- infinite horizon optimal control, [1628](#)
- influential point, [885](#)
- initial value problem, [278](#)
- inner product, [280](#)
- interior, [44](#)
- interior point, [44](#)
- Intermediate value theorem, [63](#)
- interval estimation, [715](#)
- invariant distribution, [1144](#)
- inverse function, [8](#)
- inverse image, [8](#)
- irrational number, [10](#)
- irreducible matrix, [237](#)
- isolated local minimizer, [317](#)
- isometry, [1067](#)
- iterated conditioning, [550](#)
- Ito integral, [1065](#)
- Jacobi algorithm, [257](#)
- Jarque-Bera test, [878](#)
- Jensen's inequality, [561](#)
- Jordan basis, [217](#)
- Jordan form, [218](#)
- Jordan form decomposition, [218](#)
- K-means, [1495](#)
- Kalman filter, [1017](#)
- Kendall's tau, [771](#)
- kernel, [1285](#)
- kernel PCA, [1483](#)
- KKT condition, [392](#)
- KL divergence, [595](#)
- Kolmogorov forward equation, [1117](#)
- Kolmogorov-Smirnov goodness-of-fit test, [709](#)
- kurtosis, [643](#)
- lag operator, [1173](#)
- lag polynomial, [1173](#)
- Lagrange multiplier, [385](#), [393](#)
- Lagrange polynomial, [95](#)
- Lagrange Remainder Theorem, [97](#)
- Laplace distribution, [612](#)
- Laplacian eigenmap, [1487](#)
- Laplacian matrix, [1485](#)
- Lasso regression, [1302](#)
- Law of total expectation, [549](#)
- Law of total probability, [510](#)

- least square, [820](#), [825](#), [1011](#)
- Lebesgue integral, [127](#)
- Lebesgue measure, [119–121](#)
- left eigenvector, [188](#)
- LeNet, [1562](#)
- leverage, [885](#)
- likelihood function, [673](#)
- limit point, [24](#)
- linear correlation, [770](#)
- linear difference equation with constant coefficients, [954](#)
- linear discriminant model, [1329](#)
- linear equation, [143](#)
- linear Gaussian quadratic control, [1638](#)
- linear independence, [156](#)
- linear independence constraint qualification, [383](#), [391](#)
- linear kernel, [1287](#)
- linear maps, [160](#)
- linear regression, [816](#)
- linear regression residual, [832](#)
- linearity of expectation, [533](#)
- linearly deterministic process, [1212](#)
- linearly feasible direction, [390](#)
- Liouville's formula, [986](#)
- Lipschitz continuous, [77](#), [80](#)
- Lipschitz continuous function, [462](#)
- local minimum, [316](#)
- logistic regression, [1321](#)
- lognormal distribution, [619](#)
- long short term memory, [1584](#)
- low rank approximation, [208](#), [1451](#)
- LSTM, [1584](#)
- LU decomposition, [222](#)

- MA Representation Theorem, [1186](#)
- Markov chain, [1115](#)
- Markov chain Monte Carlo, [926](#)
- Markov decision process, [1649](#)
- Markov's inequality, [561](#)
- martingale, [1053](#)
- martingale transform, [1056](#)
- matrix 2-norm, [197](#), [207](#), [210](#), [1451](#), [1452](#)
- matrix exponential, [220](#)
- matrix norm, [234](#)
- matrix power, [220](#)
- matrix square root, [230](#)
- matrix trace, [1749](#)
- maximum a posterior estimator, [696](#)
- maximum likelihood estimator, [673](#)
- MDP, [1649](#)
- MDS, [1477](#)
- mean vector, [1214](#)
- mean-value theorem, [85](#)
- measurable function, [121](#), [504](#)
- measurable set, [503](#)
- measurable space, [119](#), [503](#)
- measure, [119](#), [503](#)
- measure space, [119](#)
- median absolute deviation, [670](#)
- measurable map, [514](#)
- Mercer's condition, [1286](#)
- Mercer's theorem, [1286](#)
- metric, [38](#)
- metric space, [38](#)
- Metropolis-Hasting algorithm, [927](#)
- minimax principle, [199](#)
- Minkowski's inequality, [562](#)
- mixed strategy, [479](#)
- MMLR, [900](#)
- moment generating function, [537](#)
- monotone sequence, [19](#)
- Monte Carlo, [922](#)
- Monte Carlo integration, [922](#)
- multi-collinearity, [870](#)
- multidimensional scaling, [1477](#), [1478](#)
- multilinear form, [246](#)
- Multinomial NBC, [1368](#)
- multivariate Gaussian/normal distribution, [613](#), [723](#)
- multivariate multiple linear regression, [900](#)

- multivariate time series, [1214](#)
- mutual information, [596](#), [1395](#)
- Naive Bayes classifier, [1366](#)
- Nash equilibrium, [477](#), [479](#)
- natural filtration, [1051](#)
- NBC, [1366](#)
- nested interval property, [13](#)
- Neumann series, [277](#)
- Newton direction, [326](#)
- Newton method, [354](#)
- No free lunch theorem, [1257](#)
- non-negative matrix, [1140](#)
- non-negative matrix factorization, [1465](#)
- noncentral F distribution, [637](#)
- normal distribution, [609](#)
- normal equation, [288](#)
- normal-form game, [475](#)
- null recurrent, [1130](#)
- null space, [160](#), [169](#)
- nullity, [169](#)
- observability, [1629](#)
- off-policy Q learning, [1668](#)
- on-policy Q learning, [1667](#)
- open ball, [43](#)
- open cover, [48](#)
- open set, [53](#)
- open-loop control, [1628](#)
- optional stopping, [1059](#)
- order statistics, [586](#)
- Ornstein-Uhlenbeck process, [1097](#), [1101](#),
[1106](#), [1107](#)
- orthogonal complement, [175](#)
- orthogonal iteration, [260](#)
- orthogonal projection, [229](#), [819](#)
- orthogonal projector, [177](#)
- orthonormal expansion, [294](#)
- orthonormal matrix, [202](#)
- orthonormal set, [292](#)
- orthonormal system, [292](#)
- outlier, [885](#)
- p value, [702](#)
- PageRank, [1150](#)
- parallel affine sets, [429](#)
- Pareto optimal, [481](#)
- Parseval equality, [306](#)
- partition, [6](#)
- persistent random walk, [1154](#)
- Person correlation, [770](#)
- Picard's existence and uniqueness
theorem, [278](#)
- Pointcare inequality, [198](#)
- Poisson distribution, [604](#)
- Poisson process, [1048](#)
- policy iteration, [1651](#)
- polynomial, [95](#), [261](#)
- polynomial degree, [261](#)
- polynomial factor, [262](#)
- polynomial factorization, [263](#)
- polynomial kernel, [1287](#)
- positive definite symmetric kernel, [1286](#)
- positive matrix, [1140](#)
- positive recurrent, [1130](#)
- power method, [258](#), [260](#)
- power series, [92](#)
- predictable process, [1056](#)
- predictor, [816](#)
- previsible process, [1056](#)
- primal function, [450](#)
- Principal component analysis, [1453](#)
- principle of inclusion exclusion, [5](#)
- Principle of Optimality, [1639](#)
- probability density function, [519](#)
- probability inequalities, [560](#)
- probability mass function, [518](#)
- probability space, [506](#)
- projection along subspace, [173](#)
- projection onto a convex set, [436](#)
- projector, [174](#)
- pseudoinverse, [238](#), [240](#)
- purely deterministic process, [1212](#)
- Pythagorean formula, [292](#)

- Q learning, 1659
- QR decomposition, 182, 221
- quadratic discriminant model, 1333
- quadratic forms, 224
- quadratic variation, 1038
- random forest, 1423
- random number generation, 911
- random variable, 514
- random walk, 1154
- range, 161
- range-null decomposition, 172
- rank, 169
- Rank-nullity theorem, 162
- rank-nullity theorem, 170
- rational number, 10
- Rayleigh quotient, 195
- real symmetric matrix, 193
- recommender system, 1470
- reduced Hessian, 387
- reduced rank regression, 903
- regressor, 816
- relation, 6
- ResNet, 1564
- ridge regression, 873
- Riemann integral, 105
- Riemann-Lebesgue Lemma, 294
- right eigenvector, 188
- RMSProp, 1532
- RNN, 1581
- roots fro nonlinear equation, 353
- rotation matrix, 203
- SAC, 1716
- sample covariance, 634
- sample mean, 662
- sample path, 1026
- sample space, 506
- sample variance, 655, 662
- sample variance distribution, 655
- separating hyperplane theorem, 437
- sequence, 13
- series, 13
- set, 4
- similarity transform, 167
- simple function, 126
- simple process, 1065
- singular value decomposition, 204, 1448
- singularity, 235
- skewness, 642
- Skip-gram, 1549
- Soft Actor-Critic, 1716
- soft margin SVM, 1354
- soft thresholding operator, 1303
- span, 156
- sparsity regularization, 470
- Spearman's rho, 773
- spectral radius, 234
- square root of matrices, 230
- stability, 1006
- state transition matrix, 988
- state-action value function, 1661
- stationary distribution, 1144
- statistic, 662
- statistical power, 705
- steepest-descent direction, 324, 325
- stochastic differential equations, 1078
- stochastic process, 1026
- stopping time, 1057
- strict local minimizer, 317
- strictly stationary process, 1028
- Strong Law of Large Numbers, 572
- Student's theorem, 653
- subdifferential, 449
- subgradient, 449
- subsequence, 23
- subspace, 54, 153
- subspace optimization, 329, 346, 347
- sum, 154
- supporting hyperplane theorem, 439
- survival copula, 758
- SVD, 204, 1448
- Sylvester inertia theorem, 202

- tangent cone, [381](#), [389](#)
- Taylor series, [99](#)
- Taylor's theorem, [97](#)
- TD(o) learning, [1666](#)
- The Wronskian, [985](#)
- topology space, [53](#)
- transpose, [200](#)
- trimmed mean, [670](#)
- Twin-delayed deep deterministic policy
 gradient, [1708](#)
- two sample test, [714](#)
- unbiased estimator, [663](#)
- unconstrained nonlinear optimization,
 [316](#)
- uniform continuity, [74](#)
- union bound, [509](#)
- unitary matrix, [202](#)
- universal value function approximator,
 [1703](#)
- value iteration, [1655](#)
- value iteration theorem, [1655](#)
- Vandermonde matrix, [255](#)
- VAR operator, [1219](#)
- variance, [533](#)
- variance bias decomposition, [665](#)
- variance decomposition, [830](#)
- variance inflation factor, [870](#)
- variance of an estimator, [664](#)
- variance reduction, [938](#)
- vector space, [152](#)
- VGG, [1563](#)
- Wald's equation, [1058](#)
- Weak Law of Large Numbers, [571](#)
- weakly stationary, [1214](#)
- weakly stationary process, [1029](#)
- Weierstrass Approximation Theorem, [96](#)
- Weierstrass theorem, [67](#)
- weighted least square, [877](#)
- white noise, [1040](#)
- Wiener integral, [1072](#)
- Wishrt distribution, [633](#)
- Wold decomposition theorem, [1212](#), [1213](#)
- Wold Representation theorem, [1212](#)
- Wronskian, [961](#)
- zero measure set, [107](#)
- zero-sum game, [475](#)
- zero-sum matrix game, [482](#)

