

# Index

- $K$ -nearest neighbor, 275
- $t$ -SNE, 274
- $t$ -distributed stochastic neighbor embedding, 274
- JULIA programming language, 7
- 2-norm, 42
- 2D DFT, 59
- 2D inverse DFT, 59
- 802.11n Wi-Fi, 80
- absorbing, 305, 306, 319, 320
- activation function, 383, 387
- Adam, 361
- adaptive, 184, 277
- additive synthesis, 103
- additive white Gaussian noise, 396
- additively separable, 257, 258
- additivity, 203
- adjacency matrix, 14, 305, 306, 309
- adjoint, 20
- ADMM, 362
- affine, 11, 148, 270, 282, 385
- affine subspace, 161
- algebraic multiplicity, 70
- algebraic number, 331
- almost sure convergence, 397
- alternating direction method of multipliers, 362
- alternating projection, 369
- analysis operator, 169
- angle, 43
  - between flats, 206
  - between subspaces, 206, 233
  - between vectors, 205
  - principal, 206
- ANN, 269
- antipodal, 225
- antisymmetric, 91
- aperiodic, 307–309, 312, 313
- arctangent, 21
- arithmetic–geometric mean inequality, 217
- array, 12
  - multiplication, 368
  - slicing, 7
- artificial neural network, 191, 269, 381
- associative property, 30
- asymptotic, 316, 354
- autodifferentiation, 386
- autoencoder, 269
- AWGN, 396
- backpropagation, 270, 386
- banded, 388
- basis, 101–104, 125, 132, 179, 247, 248, 332
- basis extension theorem, 104
- basis matrix, 102
- basis vectors, 101, 193
- batch, 361
- Bayes classifier, 360
- BCCB, 20, 332
- beamforming, 80
- best fit, 187
- best Lipschitz constant, 232
- bias, 387
- bias vector, 270
- BIBO stability, 330
- biconvex, 377
- big O, 26, 354
- bilevel optimization, 162
- binary classification, 182, 356
- binary classifier design, 364
- block circulant, 20, 332
- block diagonal, 20, 285
- block lower triangular, 48
- block matrix, 19, 59
- block matrix multiplication, 33
- block matrix triangularization, 48
- block Toeplitz, 388
- block upper triangular, 48
- Boolean matrix, 369
- bounded curvature, 351
- broadcast, 37, 384
- BTTB, 388
- Burer–Monteiro factorization, 279
- canonical basis, 126
- cardinality, 64, 102, 105
- cascade, 30
- Cauchy–Schwarz inequality, 43, 203–206, 211, 391
- causal system, 16

- Cayley–Hamilton theorem, 286
- chain rule, 386
- channel estimation, 80
- characteristic equation, 50
- characteristic polynomial, 50, 53, 283–286
- Cholesky decomposition, 17, 63
- circulant matrix, 18, 67, 173, 290, 292, 293, 306, 332, 346
- circular convolution, 293
- classification, 130, 279, 386
- classifier, 3
- closed, 130, 135
- closest, 190
- clustering, 322
- CNN, 19, 381, 387
- code, 269
- codomain, 257
- coil compression, 242
- collaborative filtering, 365
- color channel, 1
- column rank, 111, 170
- column space, 109
- column vector, 10
- column-major order, 27
- commutative property, 30, 53
- commuting matrices, 293, 332, 335, 347
- compact SVD, 75, 77, 121–123, 141, 151, 153, 154, 156, 157, 160, 170, 181, 190, 193, 194, 207, 213, 236, 241, 253, 254
- companion matrix, 283, 285–288, 290, 330, 346
- compatible, 207
- complement, 106, 108
- completed the square, 152
- complex conjugate, 4, 21
- complex coordinate space, 10
- complex number, 21
- complex-conjugate pair, 50
- composite cost function, 375
- compressed sensing, 162
- concatenation, 30
- condition number, 164, 165, 192, 363
- conformable, 30
- conjugate cogradient, 150
- conjugate gradient method, 167
- conjugate transpose, 20
- connected component, 324
- consistent, 206
- constrained optimization, 134, 355
- continuous, 10
- converge, 92, 218–220, 312, 335, 353
- convergence, 229
- convergence rate, 363
- convex, 136, 257, 262, 351, 352, 368, 374, 377
- convex combination, 135
- convex function, 134, 136, 148, 150, 200, 232, 349
- convex hull, 98
- convex relaxation, 262, 371
- convex set, 130, 134, 135, 142, 200, 232
- convolution, 16, 294, 387
  - $1 \times 1$ , 388
- convolutional layer, 387
- convolutional neural network, 16, 387
- coordinate, 102
- coordinate system, 102
- correlation coefficient, 203
- cost function, 134, 335
- countably infinite, 99
- counting measure, 198
- Courant–Fischer min–max principle, 79
- covariance matrix, 51
- cross entropy, 386
- cross product, 24
- cross validation, 168, 360, 371
- cross-correlation, 387
- CWR factorization, 111
- cyclic commutative property, 55
- damping factor, 321
- data, 12
- DCT, 101, 124
- decision boundary, 356, 382
- decoder, 269
- decomposition method, 48
- deep learning, 384
- defective, 68, 69
- degree, 50, 324
- degree matrix, 324
- degrees of freedom, 91, 366
- denoising, 279, 390
- dense layer, 384
- dense matrix, 18
- derivative, 123, 148
- determinant, 41, 45, 46, 52, 54, 55, 61, 284
  - commutative property, 46
  - scaling, 47
  - transpose, 47
- DFT, 16, 67, 101, 126, 291, 331
  - matrix, 288
- diagonal, 17, 69, 87, 193, 220, 301
- diagonal majorizer, 363
- diagonal matrix, 18
- diagonalizable, 66, 68–70, 73, 83, 175, 286, 288, 296, 317, 337, 346
- diagonalize, 292
- diagonally dominant, 306, 363
- diagonally preconditioned GD, 343
- digital signal processing, 16
- dimension, 103, 104
- dimension theorem for vector spaces, 102
- dimensionality reduction, 96, 238, 244, 247, 257, 269, 271, 272, 390
- direct sum, 106
- directed edge, 303
- directed graph, 14, 305

- discrete cosine transform, 101, 124
- discrete Fourier transform, 16, 58
- discrete sine transform, 124
- discriminative, 389
- displacement vector, 225
- distance matrix, 250
- distinct, 87
- distributive property, 30, 53
- divergence, 265
- DoF, 366
- dot product, 23, 38
- downsampling, 387
- DSP, 16
- DST, 124
- dynamical system, 13
- Eckart–Young–Mirsky, 245
- economy SVD, 77, 122, 153, 154, 170
- eigendecomposition, 63–65, 69, 92
- eigenfunction, 123
- eigenvalue, 13, 14, 50, 53, 55, 61, 62, 64, 80, 91, 92, 94, 139, 283, 285, 291, 297, 331
  - algorithm, 17
- eigenvector, 14, 51, 62, 64, 65, 86, 87, 91, 92, 94, 139, 287, 291
- elementwise, 36
- EM algorithm, 378
- empirical covariance matrix, 273
- empirical measure, 183
- empirical risk minimization, 183, 356, 360
- encoder, 269
- equilibrium distribution, 315, 316, 319
- equivalent, 200, 214
- equivariant, 18, 387–389
- ERM, 183, 356, 358
- error, 161
- ETF, 172
- Euclidean norm, 43, 44, 190, 197, 198
- Euclidean space, 9
  - complex, 10
- Euler's formula, 21
- exists, 4
- factor analysis, 240, 271
- factorization, 94
- Fair potential, 352, 363
- fast Fourier transform, 292
- fast gradient method, 353
- feature map, 387
- feature vector, 13
- FEM, 124
- FFT, 292, 293
- FGM, 353, 376
- Fibonacci number, 13
- field, 8, 60, 331
- finite difference, 17, 124, 173, 332
- finite-dimensional, 104
- finite-element method, 124
- first-order, 353
- FISTA, 360, 376
- flat, 161
- floating-point number, 393
- floating-point operation, 26
- floating-point precision, 164
- FLOPs, 26
- FOIL, 222
- for all, 4
- foreground/background separation, 2, 378
- forgetting factor, 185
- Fourier series, 97, 203
- frame, 169, 170, 174, 192, 288, 331
- frame bound, 169, 331
- Frobenius inner product, 204, 205, 207, 211, 212
- Frobenius norm, 92, 93, 197, 204, 207, 216, 221, 222, 229, 233, 238, 369
- full column rank, 190, 192
- full matrix, 18
- full rank, 41, 85, 112
- full SVD, 77, 122, 154, 157, 194
- fully connected graph, 323
- fully connected layer, 384
- function composition, 131, 389
- fundamental theorem of algebra, 50
- fundamental theorem of linear algebra, 120
- Gauss–Newton matrix recovery, 377
- Gaussian elimination, 17, 41, 63, 151
- GD, 188, 218, 336
- Gelfand's formula, 217, 310
- generalized eigenvalue, 81
- generalized eigenvector, 298, 337
- generalized inverse, 221, 229, 236
- generalized PCA, 277
- generative, 389
- geodesic distance, 257
- geometric multiplicity, 70
- Geršgorin disk, 299, 310, 363
- Givens rotation, 44
- GNMR, 377
- golden ratio, 13
- GP, 355
- GPCA, 277
- Grace Wahba, 222
- grad student descent, 361
- gradient, 148–150, 233, 349
- gradient descent, 188, 191, 218, 233, 336, 349, 363
- gradient projection, 355
- Gram matrix, 17, 45, 51, 100, 151, 177, 252
- Gram–Schmidt, 64, 279
- graph
  - connected subset, 324
- graph Laplacian, 116, 322, 324, 326
- grayscale, 12

- greatest common divisor, 307
- group sparsity, 208
- Hölder's inequality, 198, 205, 212
- Haar wavelet, 173
- Hadamard product, 34, 368
- handwritten digit recognition, 3, 133, 177, 356
- hard thresholding, 259
- Heaviside step function, 259
- Helmut Wielandt, 302
- Hermitian, 21, 80, 176, 203
- Hermitian transpose, 20, 22, 150
- Hessian, 353
- Hessian matrix, 351, 359
- heteroscedastic, 188
- hidden layer, 384
- Hilbert matrix, 101
- Hilbert–Schmidt inner product, 204
- Hilbert–Schmidt norm, 207
- hinge loss, 358
- Hoffman–Wielandt inequality, 215
- homographic transformation, 82
- Householder reflection, 44
- hull, 98
- hyperbolic tangent, 383
- hyperparameter, 261
- hyperplane, 96
- idempotent, 60, 131, 132, 157, 174–176, 196, 227
- identity, 8
- identity function, 132
- identity matrix, 18, 30
- if and only if, 4
- IID, 183
- ill posed, 365
- image (range), 109
- image compression, 238
- image deblurring, 1, 294
- image registration, 222
- impulse response, 16, 291
- inclusion–exclusion principle, 107
- incremental, 277
- incremental SVD, 278
- independently and identically distributed, 183
- indicator function, 16, 198
- induced, 208, 209, 219, 310
- induced matrix norm, 208, 209, 211, 217, 218
- induced norm, 204, 232
- induction, 73, 135
- infinite dimensional, 97, 104
- infinity norm, 198
- information retrieval, 13
- inner product, 23, 31, 101, 203–205, 232, 233, 248, 250
- inner product space, 203
- input layer, 384
- integer programming, 277
- integration by substitution, 46
- intersection, 105
- invariance, 18
- invariant, 224
- inverse, 8
- inverse DFT, 59
- invertible, 40, 41, 46, 53, 68, 73, 100, 147, 288, 299
- invertible matrix, 60
- involution, 22
- irreducible, 306–309, 311, 315, 319, 320
- ISTA, 376
- ISVD, 278
- iterative shrinkage thresholding algorithm, 376
- iterative soft thresholding algorithm, 376
- Jacobian matrix, 46
- Jensen's inequality, 234
- Jordan block, 69
- Jordan normal form, 55, 63, 69, 286, 298, 318, 337
- Jordan–Wielandt matrix, 216
- Juan Gris, 2
- $K$ -nearest neighbor, 130
- kernel (nullspace), 116
- kernel PCA, 274
- Khatri–Rao, 34
- Kronecker impulse, 291
- Kronecker product, 34, 331
- Kronecker sum, 93, 289, 331
- kurtosis, 394
- Ky Fan norm, 210, 212, 213
- Lambertian surface, 250
- landmark registration, 222
- Laplace expansion, 48
- Laplacian eigenmap, 274, 326
- latent, 1, 263, 269, 366
- law of total probability, 315
- LDA, 24
- learning rate, 361
- least-squares, 187, 188, 190, 191, 232, 237, 363
- left eigenvector, 14
- left inverse, 40, 155, 170, 221
- left singular vector, 74, 82, 87, 90
- lifting, 382, 384
- line search, 189, 349
- linear, 23, 109, 388
- linear algebra, 23
- linear combination, 98, 99, 135
- linear discriminant analysis, 24
- linear least-squares, 146, 191, 224, 244, 245
- linear map, 11, 12, 108
- linear operation, 12, 14, 15, 131
- linear operator, 11, 208
- linear regression, 144
- linear separability, 382
- linear span, 98
- linear subspace, 96

- linear time-invariant, 16
- linear transform, 11, 12, 108
- linear variety, 161, 221
- linearly dependent, 81, 99, 100, 114
- linearly independent, 41, 68, 97, 99, 100, 104, 110, 125, 139, 147, 179, 247
- link graph matrix, 14
- Lipschitz constant, 232, 233, 364, 389
- Lipschitz continuous, 232, 349–353, 358, 389
- Lipschitz function, 232
- LLS, 146, 192
- Loewner order, 88, 341
- logical implication, 4, 6
- logistic function, 358, 359, 383, 386
- logistic regression, 183, 360, 364
- long-term average probability, 333
- loss function, 183, 191, 364
- low-rank, 242, 253, 365
- low-rank approximation, 165, 238, 245, 248, 270, 279, 281
- low-rank matrix completion, 365
- lower bound on the number of samples, 367
- lower Hessenberg, 17
- lower triangular, 17, 285
- lower–upper (LU) decomposition, 41
- LRMC, 365
- LTI, 16
- LU decomposition, 63
- machine learning, 3
- majorization, 335
- majorize, 342, 372, 373
- majorize–minimize, 372
- Manhattan norm, 198
- manifold, 326, 381
- mantissa, 392
- maps to, 5
- Marčenko–Pastur, 399, 404
- Markov chain, 283, 298, 301, 312, 313, 327, 328, 333, 334
- matched filter, 24
- matrix, 12–14, 22, 149, 269, 314
  - tall, 18, 75, 76, 121, 123, 129
  - wide, 18, 75, 76, 121, 122
- matrix 2-norm, 78
- matrix addition, 341
- matrix calculus, 148
- matrix completion, 262, 365, 380, 401
- matrix decomposition, 24, 30
- matrix determinant lemma, 49
- matrix Euclidian norm, 207
- matrix exponential, 36, 73
- matrix factorization, 213
- matrix inversion lemma, 41, 184
- matrix multiplication, 30, 34, 109, 368
- matrix norm, 197, 206, 217, 219, 229, 235
  - $\ell_{p,q}$ , 208
- matrix power, 53, 73
- matrix product, 113
- matrix sensing, 262, 377
- matrix square root, 210, 335, 337, 338
- matrix–matrix product, 29, 31
- matrix–vector multiplication, 15, 35, 384
- matrix–vector product, 26, 27
- max norm, 198, 207, 235
- maximum column sum matrix norm, 210
- maximum row sum matrix norm, 209
- MDL, 371
- MDS, 238, 250, 254, 257
- MIMO, 80
- min–max theorem, 81
- minimal polynomial, 69, 285–288
- minimum 2-norm solution, 190
- minimum description length, 371
- minimum norm, 160
- minimum-norm LS solution, 162, 186
- Minkowski inequality, 198
- Minkowski sum, 105
- minor, 48
- MLP, 384
- MM, 372
- model, 365
- modulo, 15
- modulus, 21
- momentum, 361
- monic polynomial, 50, 283, 285
- monomial, 50, 100
- monotone, 347, 353
- monotonicity, 215
- Moore–Penrose pseudoinverse, 154, 155, 197, 221, 229
- MSE, 264
- multi-input multi-output, 80
- multidimensional scaling, 238, 250, 254, 279, 281
- multilayer perceptron, 384
- multiple dispatch, 37
- multiplication, 55
- multiplication by a scalar, 9
- multiplicity, 53, 311, 312
- MUSIC algorithm, 116
- nearest neighbor, 275
- nearest subspace, 276
- nearest-subset classifier, 131
- necessary condition, 160
- Netflix problem, 365
- Neumann series, 220
- neural network, 145, 381
- neuron, 383
- Newton’s method, 353
- NMF, 379
- NN, 381

- NNLS, 356
- node, 303
- noncausal, 19
- nonconvex, 239, 368
- nonconvex function, 386
- nonlinear dimensionality reduction, 274, 322, 326
- nonlinear PCA, 274
- nonnegative, 308–310, 314
- nonnegative least-squares, 356
- nonnegative matrix, 301–303, 306, 309, 310
- nonnegative matrix factorization, 249, 379
- nonnegative orthant, 135
- nonsingular, 40, 140, 306
- norm, 42, 199, 204, 220, 231, 232
  - $\ell_p$ , 198
  - $\ell_\infty$ , 198
  - ordered weighted  $\ell_1$ , 229
  - sup, 198
  - uniform, 198
  - weighted, 198
- normal, 66–68, 81, 83, 84, 87, 91, 94, 115, 217, 218, 235, 332
- normal equation, 150, 158, 160, 186, 271
- normal matrix, 66, 294
- normalized root mean-squared difference, 264
- normalized root mean-squared error, 264
- not aperiodic, 307, 333
- NP hard, 368
- NRMSE, 264
- NRMSE, 264
- nuclear norm, 210, 212–214, 236, 262, 332, 371, 374
- nuclear norm regularizer, 376
- null space, 115–117, 119, 140, 141, 157, 193
- nullity, 117, 118
- numerical stability, 244
- oblique projection, 175
- ODE, 73
- OGM, 349, 354
- Oja's method, 279
- one-hot encoding, 386
- online, 184, 277
- operator norm, 78, 94, 197, 206, 208, 209, 229
- optimization, 78, 134, 335
- optimized gradient method, 349, 354
- OptShrink, 265, 282, 403
- ordinary differential equation, 73
- orthogonal, 42, 44, 52, 55, 61, 65, 70, 97, 100, 107, 125, 190
- orthogonal basis, 125–127, 332
- orthogonal complement, 107, 108, 141, 180, 227, 252
- orthogonal matrix, 43, 44, 66, 67, 74, 126, 127, 176
- orthogonal polynomials, 147
- orthogonal Procrustes problem, 222, 224, 228, 229
- orthogonal projection, 133, 142, 235
- orthogonal projection matrix, 176, 177, 181
- orthogonal projector, 157, 176, 181
- orthogonal set, 42
- orthogonal wavelet transform, 101
- orthogonality principle, 132, 161, 179, 180
- orthonormal, 42, 61, 100, 272, 291
- orthonormal  $K$ -frame, 45
- orthonormal basis, 65, 126, 132, 133, 179, 181, 248, 275, 276, 332
- orthonormal columns, 45, 155, 156
- orthonormal DFT, 291
- orthonormal rows, 155, 156
- orthonormal set, 42, 44
- orthonormal vectors, 81
- outer product, 24, 25, 33, 47, 57, 109, 110, 113, 118, 128, 208
  - matrix, 51, 58
- outlier, 201, 378, 400
- output layer, 384
- over-determined system, 154
- OWL, 229
- OWT, 101
- PageRank, 14, 283, 301, 319, 327
- parallel processing, 31
- parallelogram law, 204
- Parseval tight frame, 171–174, 177, 186, 192, 331
- Parseval's identity, 45, 202
- partial fraction expansion, 91
- PCA, 133, 249, 271, 272
- pentadiagonal, 17
- perceptron, 24, 383
- period, 307, 308, 311
- periodic, 333
- periodic end conditions, 332
- periodic functions, 97
- permutation group, 215
- permutation matrix, 46, 66
- permutation method, 242
- perpendicular, 42, 161
- Perron root, 310–312
- Perron vector, 310–312
- Perron–Frobenius eigenvalue, 310–312
- Perron–Frobenius theorem, 310–312, 315
- PGD, 335, 336, 349
- PGM, 376
- phase transition, 398
- photometric stereo, 2, 250
- pilot signal, 80
- pivoting, 63
- Plancherel theorem, 202
- POCS, 368
- POGM, 360, 376, 378
- polar decomposition, 224
- polar form, 21
- polylog, 367
- polynomial regression, 147

- pooling, 387
- poorly conditioned, 164
- positive definite, 63, 88, 89, 140, 203, 231, 301, 335, 337, 338, 341, 349, 351
- positive matrix, 301, 302, 304, 306, 308, 309, 312, 321
- positive semidefinite, 88, 89, 94, 95, 140, 193, 196, 210, 216, 224, 295, 301, 324, 336, 338, 341, 351
- power iteration, 16, 92, 296, 298, 311, 312, 333
- preconditioned gradient descent, 336, 363
- preconditioned steepest descent, 335, 349
- preconditioner, 294, 336, 338
- predict, 144
- primitive, 302, 303, 305–309, 312, 315–319
- primitive matrix, 301, 302, 305, 309, 312, 317
- principal component analysis, 13, 249, 271
- principal component regression, 166
- principal eigenvector, 16
- principal minors, 89
- principal square root, 337, 338
- Procrustes, 225, 236, 250
- Procrustes problem, 197, 222, 225, 237
- projection, 130, 131, 135, 177, 178, 196, 200, 249, 355, 369
- projection matrix, 131, 132, 174, 177, 317
- projections onto convex sets, 368
- projective transformation, 82
- proximal gradient method, 376
- proximal operator, 257, 259, 364, 376
- PSD, 335
- pseudoinverse, 157, 162, 170, 188, 195, 221
- push-through identity, 41
- QR decomposition, 64, 85, 151, 158
- quadratic, 149, 335
- query, 14
- random matrix theory, 265, 390
- random walk, 319
- range, 109, 110, 115, 139, 141, 177, 193
- range space, 119, 140
- rank, 75, 85, 93, 96, 111–114, 139, 140, 193, 249, 279
- rank constraint, 371
- rank regularizer, 371
- rank-1 approximation, 244, 245
- rank-1 matrix, 24
- rank-1 update, 41, 49, 91, 185
- rational function, 60, 278
- ray, 225
- Rayleigh quotient, 81, 273
- Rayleigh–Ritz theorem, 81
- real coordinate space, 9
- recommender system, 365
- rectangular diagonal matrix, 17, 74, 156, 267
- rectified linear unit, 383
- recurrent neural network, 217, 300
- recursive, 184, 277
- recursive least-squares, 184
- reducible, 306
- regularization parameter, 167, 279
- regularized low-rank approximation, 281
- regularized LS, 140, 191, 192, 233
- regularizer, 261
- relative complement, 5, 53
- ReLU, 383
- residual, 146, 160, 161, 187, 201
- restricted isometry property, 377
- reverse mode, 386
- reverse triangle inequality, 200, 300
- RGB, 1, 387
- ridge regression, 165, 167
- right eigenvector, 14, 90
- right inverse, 40, 155, 221
- right singular vector, 74, 82, 87
- RIP, 377
- risk, 183, 264
- RLS, 184, 185, 277
- RMT, 390
- robust PCA, 378
- robust regression, 202
- root of unity, 291
- roots, 285
- rotation
  - improper, 72
  - matrix, 67, 72, 75
- roundoff error, 207, 391, 395
- row gradient, 359
- row rank, 110, 111, 113
- row space, 109
- row vector, 10
- sample covariance matrix, 272
- sampling mask, 367
- sandwich inequality, 372
- SAT exam, 189
- scaling, 203
- scatter matrix, 51
- Schatten norm, 207, 210–212, 215, 216
- Schur complement, 40
- Schur decomposition, 64, 338
- Schur norm, 207
- Schwarz, 43, 204
- scores, 273
- scree plot, 240, 242
- second-order, 353
- seminorm, 198, 199, 230
- semiorthogonal matrix, 45
- semiunitary, 45
- sensor localization, 250
- set difference, 5, 53
- SGD, 361, 386
- Sherman–Morrison formula, 41

- Sherman–Morrison–Woodbury identity, 41
- shift invariant, 18, 387–389
- shrink, 256, 261
- sigmoid, 383
- signal-to-noise ratio, 80
- signficand, 392, 393
- similar, 62, 68, 69, 91
- similarity function, 323
- similarity graph, 323
- simplex equiangular tight frame, 172
- simultaneously diagonalizable, 69, 346, 347
- simultaneously triangularizable, 347
- single precision, 392
- singular matrix, 51, 64, 299
- singular value, 74, 75, 80, 81, 92, 94, 169, 192, 233
- singular value decomposition, 17, 24, 63, 69, 74
- singular value hard thresholding, 261, 374
- singular value soft thresholding, 262, 375
- singular vector, 74, 94, 248
- sketching, 279
- skew-symmetric, 62
- smooth function, 349
- SNR, 80
- social justice, 189
- soft thresholding, 258, 262
- softmax, 386
- source localization, 238
- span, 98, 99, 102, 105
- spark, 114
- sparse matrix, 18
- sparse subspace clustering, 277
- sparsity, 160, 258
- spectral clustering, 116, 283, 322, 323, 325, 327
- spectral norm, 78, 93, 197, 209, 214, 216, 233, 245, 281, 294, 351
- spectral normalization, 296
- spectral radius, 217, 229, 235, 298, 300, 309, 310, 363
- spectral theorem, 64–66
- specular reflection, 250
- square, 17, 75, 303, 309, 314
- square root, 336, 337
- SSC, 277
- stable rank, 249, 250
- standard basis, 101, 104, 126
- standard simplex, 386
- state, 303, 313
- stationary distribution, 315
- stationary point, 148
- statistical learning theory, 183
- steady state, 315, 316
- steepest descent, 189, 349
- Stein’s unbiased risk estimate, 168, 264
- step size, 336, 353, 363
- Stiefel manifold, 127, 212, 225, 228, 237, 276, 279
- stochastic eigenvector, 315, 316
- stochastic gradient descent, 361
- stochastic matrix, 301, 302, 309, 310, 312, 314
- streaming, 184, 277
- strictly convex, 136, 200, 235, 257, 351, 352, 359
- strong law of large numbers, 396
- strongly connected, 305, 306, 318–320
- sub-Gaussian, 397
- subadditive, 114, 197
- submultiplicative, 206–209, 213, 216, 219, 229, 235
- subset, 96, 105
- subspace, 96, 98, 103, 105, 108, 109, 116, 131, 132, 135, 139, 141, 177, 179
  - sum, 105
- subspace clustering, 277, 322
- subspace learning, 277, 279
- subspace sum, 106
- subspace union, 106
- sum of outer products, 36
- supervised learning, 168, 182, 191, 277
- supervised PCA, 276
- support-vector machine, 358
- supremum, 198
- SURE, 264
- surface normals, 250
- surrogate function, 372
- surrogate loss functions, 358
- SVD, 63, 64, 69, 74, 86, 92–94, 114, 143, 156, 157, 167, 181, 193, 221–223, 227, 229, 239, 248, 270, 273, 279, 374
- SVM, 358
- SVST, 262, 375, 376
- Sylvester’s criterion, 89
- Sylvester’s determinant theorem, 49
- Sylvester’s rank inequality, 113
- symmetric, 21, 97
- symmetric gauge, 261, 282
- synthesis operator, 169
- system of linear equations, 13, 143
- Talagrand’s concentration inequality, 395
- Taylor’s theorem, 351, 373
- tensor, 387
- term–document matrix, 13
- thin SVD, 77
- thresholding, 261
- tight, 79, 94, 209, 211, 234, 235
- tight frame, 44, 170–172, 174, 193, 331
- tight upper bound, 79
- Tikhonov regularization, 165, 167, 192
- time invariance, 18
- time-homogeneous Markov chain, 314
- Toeplitz, 18, 293–295, 387
- total variation, 199
- trace, 54, 55, 60–62
- trace norm, 210
- tracking, 277

- Tracy–Singh, 34
- training, 270
- transient, 305, 306, 313, 319
- transition matrix, 301, 309, 312, 314, 315, 317, 327, 328, 333, 334
- transition probabilities, 313
- translation, 225
- transpose, 9, 20, 22
- triangle inequality, 142, 197, 200, 206, 212, 213, 215, 230, 351
- tridiagonal, 17
- trigonometric identities, 124
- truncated SVD, 164, 165, 167, 240, 256
- tune, 167
- tuples, 304
- TV, 199
- twice differentiable, 351
- U-net, 389
- uncorrelated, 230
- uncountably infinite, 64, 74, 78, 99
- under-determined, 151, 158
- union of subspaces, 105, 139, 277, 322
- unique, 75, 243, 244, 311, 317, 370
- unit norm, 42
- unit vector, 24, 25
- unitarily invariant, 140, 202, 216, 221, 229, 245, 251, 261, 282
- unitary, 41, 44, 52, 61, 65, 66, 74, 115, 118, 140, 186, 248
- unitary DFT, 291
- unitary eigendecomposition, 66, 67, 176
- unitary invariance, 45, 82, 235
- unitary matrix, 44, 126, 172, 291
- units, 35, 47, 54, 60
- unity, 8
- universal approximation theorem, 385
- universality, 400
- unnormalized graph Laplacian, 324
- unsupervised, 269, 277
- updating, 277
- upper Hessenberg, 17
- upper triangular, 17, 103, 285
- upsampling, 387
- Vandermonde, 288, 291, 295, 331
- variational, 79
- vec operator, 34
- vec trick, 34, 58, 60
- vector, 7, 9, 10, 12, 22
- vector 2-norm, 198
- vector addition, 6, 9
- vector norm, 197, 219, 229, 232
- vector space, 8, 9, 12, 23, 35, 96, 108, 204, 206, 232
- Venn diagram, 177, 192, 196
- vertex, 303
- video, 2, 378
- volume of the parallelepiped, 46
- von Neumann’s trace inequality, 211, 212
- Wahba’s problem, 222
- weak duality, 95
- weakly differentiable, 264
- weight decay, 386
- weight matrix, 324
- weighted directed graph, 303, 314
- weighted graph, 322, 323
- weighted least-squares, 188
- weights, 387
- Weinstein–Aronszajn identity, 49
- Weyl’s inequality, 82, 215, 239, 247
- Wielandt–Hoffman theorem, 391
- Wirtinger calculus, 150
- Young’s convolution inequality, 294
- zero vector, 9