

Index

- 0/1-loss, 5
- a posteriori information, 69
- a priori information, 69
- activation function, 209
- adaptive, 88, 149
- adjoint, 285
- affine functions, 9, 41, 169
- alternating minimization, 203
- approximability set, 77, 202
- approximation error, 5
- augmented data matrix, 42
- B*-spline, 218
- backpropagation, 240
- Baire category theorem, 274, 307
- Banach space, 274
- basis pursuit, 118
- Bayes predictor, 9
- Bernstein inequality, 85, 304, 310
- Bernstein polynomials, 301, 310
- Bernstein theorem, 76, 303
- bias vectors, 209
- bias-complexity tradeoff, 8
- binary classification, 5, 16, 23
- Birkhoff theorem, 170, 183, 281
- bit-test matrix, 151
- bump function, 219
- Carathéodory theorem, 170, 277
- Carathéodory–Toeplitz theorem, 305
- center-based clustering, 49
- Chebyshev ball, 70, 181
- Chebyshev center, 70
- circulant matrix, 246
- clusters, 47
- co-observation, 69
- coherence, 119, 153
- collation channel, 232, 234
- compatibility indicator, 77, 84
- complete, 34, 274
- conic hull, 179
- convex hull, 277
- convex optimization program, 160
- convex set, 277
- convolution product, 219, 246, 302
- convolutional neural network, 246
- Courant–Fischer theorem, 58, 87, 286
- covariance matrix, 57
- covering number, 253
- cross polytope, 255
- cross-Gramian, 80
- cross-validation, 8
- curse of dimensionality, 95
- data matrix, 56
- data-consistent, 69
- depth, 208, 209
- descent direction, 161
- Dini lemma, 310
- direct theorem, 76, 302
- Dirichlet kernel, 310
- discrepancy, 103
- discrete trigonometric moment problem, 305
- disjunct, 150
- distortion, 69
- divided difference, 218
- dominated extension theorem, 278, 306
- dual cone, 184
- dual program, 177
- dual space, 274
- dual variables, 177
- dynamic programming, 197
- Eckart–Young theorem, 57, 294

316

effective sparsity, 142
 empirical covariance matrix, 58
 empirical risk, 5
 empirical risk minimization, 5, 239
 epoch, 165, 241
 equality constraints, 160
 estimation error, 5
 expressive power, 216
 extreme point, 170, 281
 extreme set, 281
 Farkas lemma, 179
 feasible point, 160
 feasible polytope, 170
 feature map, 27
 features, 4
 Fejér kernel, 302, 310
 Frobenius inner product, 291
 Frobenius norm, 291
 fundamental theorem of PAC-learning, 16
 gamma function, 250
 gauge, 280
 gaussian kernel, 28, 39
 Gelfand width, 93, 129, 131
 generalization error, 4
 global minimizer, 160
 gradient descent algorithms, 161
 Gramian, 80
 graph Laplacian, 51
 growth function, 10, 11
 Hahn–Banach extension theorem, 82, 280, 283, 293
 Hahn–Banach separation theorem, 179, 197, 280
 Halton sequence, 108
 Hammersley set, 110
 hard SVM, 26, 30
 hard thresholding operator, 139, 197
 hard thresholding pursuit, 202
 Hardy and Krause, 103
 Hausdorff distance, 174
 Helly theorem, 278
 Hilbert space, 274
 Hilbert–Schmidt integral operator, 38
 hinge loss, 27
 Hoeffding inequality, 7, 262
 Hoeffding lemma, 262, 263
 Huber loss, 246
 hypothesis class, 5
 incremental gradient, 165
 inequality constraints, 160

Index

inequality form, 169
 information complexity, 94
 information-based complexity, 94
 instances, 4
 intractable, 94
 intrinsic error, 69
 inverse theorem, 76, 303
 iterative hard thresholding, 199
 Jackson kernel, 302
 Jackson theorem, 76, 302
 Johnson–Lindenstrauss lemma, 59
 Kashin decomposition theorem, 258
 kernel, 28, 31
 kernel PCA, 58
 k -means clustering, 49
 k -median clustering, 49
 k -medoids clustering, 50
 Koksma–Hlawka inequality, 105, 111
 Kolmogorov superposition theorem, 227, 307
 Kolmogorov width, 93
 Korovkin theorem, 300
 Krein–Milman theorem, 170, 282, 284
 Ky Fan norm, 293, 296
 ℓ_0 -minimization program, 117
 ℓ_1 -minimization program, 118
 ℓ_1 -rank restricted isometry property, 134, 270
 ℓ_1 -restricted isometry property, 119, 139, 267, 273
 labels, 4
 Lagrange dual function, 177
 Lagrangian, 177
 LASSO, 43
 learning map, 4
 learning rate, 165, 241
 least-squares problem, 42
 Lidskii inequality, 287, 289
 line search algorithms, 161
 linear program, 169
 linear separability, 24
 linearly separable, 23
 Lipschitz functions, 88, 220
 Lipschitz seminorm, 88
 Lloyd algorithm, 50
 local minimizer, 160
 local worst-case error, 70, 202
 locally linear embedding, 61
 locally optimal, 69
 logistic loss, 45
 logistic regression, 45
 logistic sigmoid function, 209, 241

- Lorentz cone, 184
 loss function, 5
 Mairhuber–Curtis theorem, 33
 margin, 25
 Markov inequality, 18, 59, 167, 259, 265, 268
 Massart lemma, 22
 McDiarmid inequality, 20, 264, 266, 270
 mean-square error, 41
 Mercer theorem, 37
 method of moments, 190
 minibatch, 241
 Minkowski functional, 280
 Mirsky inequality, 132, 289, 294
 model set, 69
 model-consistent, 69
 moment generating function, 19, 260
 Monte Carlo, 102
 Moore–Aronszajn theorem, 33
 Natarajan dimension, 22
 Natarajan lemma, 22
 natural spline, 74
 nearest neighbors, 62
 Nesterov accelerated gradient methods, 164
 net, 253
 neurons, 208
 no-free-lunch theorem, 20
 noisy gradient, 165
 nonadaptive, 149
 nonnegative least-squares problem, 44
 normal, 285
 normal equations, 42
 normalized kernel, 39
 nuclear norm, 132, 291
 null error, 70
 null space property, 118, 128, 133, 256
 objective function, 160
 observation map, 69
 observations, 69
 operator norm, 291
 optimal recovery map, 69
 ordered weighted ℓ_1 -norm, 182
 overfitting, 8
 overparametrized regime, 42
 owl-norm, 182
 PAC-learnable, 6
 packing number, 253
 Pajor lemma, 13
 Paley–Zygmund inequality, 260
 PCA, 57
 Peano representation, 217, 218
 perceptron algorithm, 24, 30
 polynomial kernel, 28
 positive definite function, 32
 positive semidefinite, 184, 286
 positive semidefinite function, 32
 predictor, 4
 primal program, 177
 principal component analysis, 57
 probably approximately correct, 6
 projected gradient descent, 199
 Prony method, 130
 quadratic barrier, 119
 quadratically constrained quadratic program, 194
 quasi-Monte Carlo, 102
 quotient norm, 284
 quotient space, 127, 284
 Rademacher complexity, 22
 Rademacher variables, 18
 radical inverse function, 108
 Radon theorem, 15
 Rayleigh quotient, 58, 286
 rearrangement inequalities, 185
 recovery map, 69
 rectified linear unit, 210, 241
 regression, 5, 41
 regularizer, 43
 representer theorem, 27
 reproducing kernel Hilbert space, 31
 reproducing property, 32
 restricted isometry property, 122, 200
 ridge regression, 43
 Riesz representation theorem, 32, 274, 275, 306
 Riesz representer, 79, 95
 Riesz–Fejér theorem, 189, 304, 306
 risk, 4, 9
 robust null space property, 121, 122
 robust optimization, 181
 robustness, 123
 S-lemma, 195
 sample complexity, 6
 Sauer lemma, 13, 19
 sawtooth function, 237
 Schatten norm, 137, 292
 Schur complement, 187
 Schur product theorem, 39
 second-order cone, 184
 self-adjoint, 285
 self-adjoint dilation, 137, 289

318

Index

- self-dual, 184
- semidefinite cone, 184
- semidefinite program, 136, 184
- separable, 150
- separated, 253
- sequential Banach–Alaoglu theorem, 192, 193
- shallow network, 209
- shatter function, 10, 11, 13, 19, 222
- shattered, 10, 11
- sign product embedding property, 147
- simplex algorithm, 171, 172
- single-linkage algorithm, 48
- singular values, 288
- singular vectors, 288
- slack variables, 118, 169
- smooth, 161
- Sobolev space, 73, 98
- soft maximum, 262
- soft SVM, 27
- sort-and-split technique, 120, 124, 143, 258
- source channel, 232, 234
- sparse, 116
- spectral clustering, 51
- spectral theorem, 38, 285
- spline algorithm, 73, 186
- spline functions, 73
- square loss, 5
- stability, 123
- standard form, 169, 184
- star discrepancy, 104
- star discrepancy function, 104
- state vector, 208
- steepest descent algorithm, 162
- step function, 211
- Stirling formula, 250
- stochastic block model, 52
- stochastic gradient descent, 165, 240
- Stone–Weierstrass theorem, 217, 297, 298
- strictly subgaussian random variables, 261
- strong feasibility, 184
- strongly convex, 163, 168
- strongly selective, 150
- strongly tractable, 94
- subgaussian random variables, 261
- subgaussian standard, 261
- sublinear functional, 278
- sum-of-squares, 188
- summation by parts, 185, 236, 290, 294
- supervised learning, 47
- support vectors, 26
- supporting hyperplane theorem, 257, 278
- symmetric gauge function, 291
- synapses, 208
- tails, 259
- targets, 4
- tent function, 211, 213
- test matrix, 150
- thin plate splines, 75
- Tietze extension theorem, 83, 84, 275, 309
- Tikhonov regularization, 43, 46
- Toeplitz matrix, 190, 305
- tractable, 94
- training sample, 4
- training set, 8
- trigonometric polynomials, 76
- tuning parameter, 43
- unconstrained optimization, 160
- underfitting, 8
- underparametrized regime, 42
- uniform convergence property, 16
- unitarily invariant norm, 291
- unitary, 285
- universal approximation theorem, 216
- unsupervised learning, 47
- Urysohn lemma, 276
- validation set, 8
- van der Corput, 108
- Vapnik–Chervonenkis dimension, 10
- variation, 103
- VC-dimension, 10, 221
- Vitali, 103
- von Neumann trace inequality, 64, 290, 295
- Ward algorithm, 55
- Weierstrass theorem, 219, 226, 297
- weight locality, 245
- weight matrices, 209
- weight sharing, 245
- Weyl inequality, 54, 287
- width, 208, 209
- Wielandt minimax principle, 286
- worst-case error, 69
- Zaremba identity, 105
- Zorn lemma, 279, 283