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
convex hull, 179
convex optimization program, 160
convex set, 277
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
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
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
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