Index

absolute continuity
  of a function, 362
  of a measure, 67, 104, 168, 216, 224, 330
absolutely continuous component of a measure, 234, 330
Active Set algorithm, 81, 102, 270, 301, 303, 311
adaptation, 13, 280, 301, 312
adaptive Lasso estimator, 91, 105
additive model, 176
adjacent vertices of a graph, 188
adjoint of a bounded linear operator, 368, 377
affine
  combination, 334
  function, 335
  hull of a set, 334
  set, 334
Aitken estimator, 24, 48
all-ones vector, 19
almost everywhere, 328
almost sure convergence, 400
analysis of variance, 48
ancestor (in a graph), 189
anisotropic Nikol'ski class, 128
ANOVA, 48
antichain, 250
  in a DAG, 300
  maximal, 250
antiderivative, 291, 305, 309, 363
antitonic, 289
  efficiency, 291
  projection, 289
  relative efficiency, 290
Approximate Message Passing, 104
Assouad's lemma, 229, 231, 232, 250, 252
asymptotic variance factor, 288
axiom of choice, 161, 182, 377

ball σ-algebra, 378
Banach space, 365
bandwidth, 107, 130, 153
basic inequality, 72, 101, 312
basis pursuit, 343
Bayes
  estimator, 246
  misclassification error, 223
  risk, 11, 246
Bayesian networks, 220
Benjamini–Hochberg procedure, 41, 47, 48
Benjamini–Yekutieli procedure, 47, 48
Bennett's inequality, 355, 409
Bernstein's inequality, 357, 409
Berry–Esseen theorem, 9, 20
Bessel's inequality, 375, 377
best subsets selection, 59
beta-min condition, 77
bi-convex/bi-concave function, 284
bias (of an estimator), 4
bias–variance trade-off, 4, 112, 118, 143
binwidth, 107
block
  coordinate descent, 81, 87, 218
  isotonic estimators, 301
blocked path, 195
blockwise matrix inverse, 35, 390
Bochner's theorem, 179, 345, 409
Bolzano–Weierstrass theorem, 379
Bonferroni
  correction, 28, 39, 48
  test, 40
bootstrap, 1, 19
Borel
  measure, 324
  set, 324
  σ-algebra, 324
Borel–Cantelli lemmas, 8, 326, 408
boundary of a subset of a topological space, 317
bounded
  linear operator/functional, 161, 367
  Lipschitz metric, 38, 401
Bretagnolle–Huber inequality, 243, 247
bump function, 249

Carathéodory's theorem, 101, 341
Cartesian product of topological spaces, 322
Cauchy
  criterion for uniform convergence, 322
  determinant, 138, 147
  matrix, 138, 147

https://doi.org/10.1017/9781009160421.011 Published online by Cambridge University Press

sequence, 159, 316
Cauchy–Schwarz inequality, 364
causal
DAG, 192
minimality, 200, 216
central limit theorem, 9, 20, 402, 411
Lindeberg–Feller, 402, 406, 411
multivariate, 407
chain (in a DAG), 300
characteristic function of a random vector, 343, 409
Chernoff bound, 352
χ^2-divergence, 225
child (in a graph), 188
classifier, 222
support vector, 165
closed
ball in a metric space, 316
set in a metric space, 316
testing procedure, 40, 48
closest point theorem, 365
closure of a subset of a topological space, 317
Cochran's theorem, 25, 45, 48
coercive function, 60, 79, 263
cofinite topology, 317
collider node (in a path), 195
compact
bounded linear operator, 367
singular value decomposition, 53, 394, 400
topological space, 318
compatibility
condition, 72, 74, 101, 104
factor, 72, 103
complementary pairs stability selection, 97, 105, 208
complete
graph, 188
metric space, 316
composite hypothesis, 29
concave function, 335
conditional
dependence, 184, 199
distributions, 382
expectation, 381, 410
tower law, 381
independence, 184, 199
graph, 190
testing, 207, 219, 220
Jensen's inequality, 234, 381
monotone convergence theorem, 215, 381
probability, 381, 410
cone invertibility condition, 104
confounder, 194
conjugate indices, 370, 391
constrained risk inequality, 253
continuity
at a point in a metric space, 316
at a point in a topological space, 321
uniform, 316
continuous
function on a metric space, 316
function on a topological space, 317
mapping theorem, 402, 410
contraction
mapping theorem, 321
property, 185, 198
convergence
almost surely, 400
in a metric space, 315
in a topological space, 321
in distribution, 401
in probability, 401
convex
combination, 334
cone, 248, 288, 301
function, 335
hull of a set, 334
least squares estimator, 301, 310
optimisation problem, 339
regression, 301, 314
set, 334
support, 266
convolution
of functions, 109, 347
of probability measures, 248
coordinate
descent, 79, 105
projection, 322
correlation, 349
counting measure, 323
covariance
between random variables, 328
matrix, 327
covariate, 21
covering number, 316
Cramér–Wold device, 37, 407, 411
cross-validation, 56, 103, 116, 128
generalised, 142, 147
leave-one-out, 58, 142
cubic spline, 139, 147, 314
curse of dimensionality, 126
CUSUM diagram, 293
cycle (in a graph), 189

d-separated sets of nodes, 195, 216
data processing inequality, 233, 235, 246, 252
Davis–Kahan theorem, 399, 410
debiased
Lasso, 93, 210
machine learning, 220
decision theory, 10
delta method, 403
dense (functions), 379

https://doi.org/10.1017/9781009160421.011 Published online by Cambridge University Press

density
  (of a distribution or random variable), 330
  quantile function, 289
descendant (in a graph), 189
design matrix, 21
desparsified Lasso, 93
dimension
  of a convex set, 334
  of an affine set, 334
Dini's theorem, 322, 376
direct product of Hilbert spaces, 378
directed
  acyclic graph (DAG), 189, 300
  cycle (in a graph), 189
  edge (of a graph), 188
  graph, 189
  path (in a graph), 189
directional separation, 195
discrete
  metric, 315
  topology, 317
disintegration, 308, 382, 388
distribution
  (generalised function), 409
  of a random vector, 324
do operator, 193
dominated
  convergence theorem, 328, 408
  measure, 330
Doob's maximal inequality, 310, 384, 386
Doob–Dynkin lemma, 333
double machine learning, 220
dual
  norm, 367
  space, 367
Dvoretzky–Kiefer–Wolfowitz–Massart–Reeve inequality, 7
Dykstra's projection algorithm, 293
Dynkin's π–λ lemma, 298, 325, 408

Earth Movers' distance, 277
Eckart–Young theorem, 395, 410
edge (of a graph), 188
effective
  domain of a convex function, 335
  kernel, 132, 133
efficiency (of a kernel), 122
eigenfunction, 166, 372
eigenspace, 166, 378
eigenvalue of a self-adjoint operator, 369, 372
eigenvector of a self-adjoint operator, 369
elastic net estimator, 90, 105
empirical
  distribution function, 7
  Kullback–Leibler loss, 281
entrywise matrix norm, 392
Epanechnikov kernel, 122, 125, 128
epigraph, 342
ϵ-covering number, 316
equal in distribution, 324
equicorrelation
  set, 65
  signs, 66
equivalent norms, 379
ergodic theory, 411
essential supremum, 369
estimator, 10
Euclidean
  metric, 315
  topology, 317
event, 324
exchangeable random variables, 33
extended real line, 332
extreme point of a convex set, 306

face of a convex set, 61
faithful distribution, 204, 216
false discovery
  proportion, 41
  rate, 41, 48
family-wise error rate, 39, 48
Fano's lemma, 234, 236, 249, 252
Fatou's lemma, 270, 328, 408
F-distribution, 30, 43, 48
f-divergence, 224, 247, 252
feasible
  point, 337
  set, 337
feature map, 150
filtration, 383
  reverse, 385
finite
  DAG, 189
  measure, 323
finite rank bounded linear operator, 379
first Borel–Cantelli lemma, 326, 387
first countable topological space, 321
first-order Sobolev kernel, 153, 155, 160, 169, 174
Fisher
  consistency, 287
  divergence, 290, 313
  information, 288
fitted values, 23
fixed design, 21
forward selection, 59, 101
Fourier transform, 347
Frisch–Waugh–Lovell theorem, 45, 48
Frobenius norm, 392
F-test, 31, 44
Fubini's theorem, 329, 408
full column rank matrix, 22
fundamental theorem
  of calculus, 261, 362, 409

https://doi.org/10.1017/9781009160421.011 Published online by Cambridge University Press

of statistics, 7
fused Lasso estimator, 89, 105
gamma function, 351
Gauss–Markov theorem, 23, 48, 50, 287
Gaussian
distribution, 348
graphical models, 207
kernel, 107, 153, 181, 183
width, 313
generalised
covariance measure, 212, 213
cross-validation, 142, 147
least squares estimator, 24
likelihood ratio test, 29
Minkowski inequality, 115, 370
Gilbert–Varshamov lemma, 236, 248, 252
Glivenko–Cantelli theorem, 7, 19, 261, 387
global Markov property, 190, 220
Gram matrix, 150
graph, 188
conditional independence, 190
directed, 189
undirected, 189, 219
graphical
Lasso, 209, 217, 220
models, 219
directed, 220
greatest convex minorant, 293, 309
Grenander
distribution projection, 260
estimator, 256, 261, 310
projection, 258, 261, 307
group Lasso estimator, 89, 105
Hadamard product, 152
Hamming distance, 236, 248
hard
intervention, 192
thresholding, 62
hashing trick, 178
Hausdorff space, 320
Heine–Borel theorem, 318, 322, 408
Hellinger
distance, 225, 252
integral, 252
Hermite polynomials, 123
Hessian matrix, 335
heteroscedasticity, 38, 130, 306
high-dimensional data, 22, 50
high-probability
Le Cam's two-point lemma, 242
minimax lower bound, 240, 253
Hilbert
matrix, 137
space, 365
projection theorem, 296, 365
Hilbert–Schmidt
norm, 368, 378
operator, 368
histogram, 107
Hochberg's procedure, 47
Hodges estimator, 5, 19
Hoeffding's
inequality, 353, 409
lemma, 353
Hölder
class, 110, 127, 134, 175, 222, 236, 262, 280
continuity, 277, 312
seminorm, 227, 249
Hölder's inequality, 370, 410
Holm's procedure, 40, 48
homeomorphism, 321
homoscedasticity, 22, 130
homothetic density, 312
homotopy method, 105
Huber loss function, 84
idempotent matrix, 22
identifiability, 50, 149, 309, 312
image measure, 233, 333
imaginary part of a complex function, 343
in-sample squared prediction error, 213
increasing
convex/concave LSE, 302
set, 47, 296
independence
conditional, 184, 199
of events, 325
of random variables, 325
of σ-algebras, 325
Independent Component Analysis, 276, 282, 312
indiscrete topology, 317
inflection point, 302
inner
product space, 364
regular measure, 331
input space, 151
integrable function, 327, 343
integral, 327
integration, 327
by parts, 330, 408
by substitution, 331, 408
interaction term, 149
intercept term, 22
interior
of a subset of a topological space, 317
point of a topological space, 317
intersection
hypothesis, 40
property, 187, 191
intervention, 192
distribution, 192
hard, 192
inverse function theorem, 308

https://doi.org/10.1017/9781009160421.011 Published online by Cambridge University Press

irrepresentable condition, 77, 105
isometric isomorphism, 379
isotonic
  LSE, 253, 292, 297, 300, 310
  projection, 297, 309
  regression, 231, 236, 245, 292, 296, 313
Jaccard similarity kernel, 154, 182, 183
jackknife, 19
Jacobian matrix, 331
James–Stein estimator, 12, 20, 221, 253
Jensen's inequality, 340, 408
  conditional version, 234, 381
Johnson–Lindenstrauss lemma, 359
Kantorovich–Rubenstein distance, 277
Karush–Kuhn–Tucker (KKT) conditions, 65, 181, 217, 337
Kendall's τ coefficient, 154
kernel, 107, 130
  (RKHS), 151
  density estimator, 107, 253
  first-order Sobolev, 153, 155, 160, 169, 174
  Gaussian, 107, 153, 181, 183
  integral operator, 166, 380
  Jaccard similarity, 154, 182, 183
  linear, 153, 160
  machine, 148, 183
  matrix, 151, 180
  of a bounded linear operator, 367
  polynomial, 153
  ridge regression, 150, 178, 183, 214
  second-order Sobolev, 154, 155, 161, 180, 181
  shift-invariant, 178
  trick, 150
  uniform, 107
knot (of a piecewise affine function), 258
Kolmogorov distance, 260
Kullback–Leibler divergence, 225, 252
ℓ_0-penalised estimator, 62
L_1-Wasserstein distance, 277
L_p space, 369, 409
ℓ_q-norm of a vector, 391
Lagrange multipliers, 337
Lagrangian, 102, 337
  dual
    function, 338
    problem, 338
  methods, 337
  sufficiency theorem, 337
λ-system, 325
Lasso
  adaptive, 91, 105
  computation, 79
  debiased/desparsified, 93, 210
  estimation error, 72, 104
  estimator, 4, 60, 104
  fused, 89, 105
  graphical, 209, 217, 220
  group, 89, 105
  prediction error, 69, 72, 104
  relaxed, 91, 105
  residual prediction tests, 82
  scaled, 105
  solution path, 63
  square-root, 87, 105, 210
  uniqueness, 64, 104
  variable selection, 76
Lava estimator, 102, 105
law of a random vector, 324
Le Cam's two-point lemma, 225, 228, 239, 249, 250, 252
  high-probability version, 242
least concave majorant, 256, 258, 293, 306
leave-one-out cross-validation, 58
Lebesgue
  decomposition theorem, 224, 233, 330, 408
  measure, 323
    translation invariance, 332
  σ-algebra, 296, 324
Lebesgue–Stieltjes
  integral, 255, 328
  measure, 328
left singular vector, 392
length of a path in a graph, 189
Lepski's method, 117, 128
leverage, 44
Lidskii's theorem, 398, 410
life tables, 1
likelihood function, 25
Lindeberg condition, 406
Lindeberg's method, 411
Lindeberg–Feller central limit theorem, 402, 406, 411
linear
  estimator, 23, 132
  kernel, 153, 160
  model, 21
    noiseless, 76
  regression, 48
  smoother, 132
loading vector, 394
local
  constant estimator, 132, 143
  linear estimator, 132, 143
  Markov property, 195
  polynomial estimator, 130, 147
    outrigger, 147
  test, 40
locally
  absolutely continuous function, 114, 126, 288, 363
  compact space, 320
  Lipschitz function, 342

https://doi.org/10.1017/9781009160421.011 Published online by Cambridge University Press

location equivariance, 51
log-concave
  density, 262
    convolution, 265
    estimation, 311
    marginal, 265
  function, 262
  ICA projection, 283
  location mixture, 313
  Marshall's lemma, 280, 312
  maximum likelihood estimator, 117, 163, 266, 279, 280, 308, 311
  projection, 266, 273, 277, 309, 311
    affine equivariance, 273
    smoothed, 275, 312
log-likelihood function, 25
logistic log-likelihood, 84
logit function, 217
log-k-affine log-concave density, 281
loss function, 10
low coherence condition, 104
lower
  minimax quantile, 241, 253
  set, 296
Lyapunov condition, 405

Mallows distance, 277
margin (between classes), 164
Markov
  blanket, 191
  equivalence class, 201
  equivalent DAGs, 201
Marshall's lemma, 260, 261, 310
  log-concave, 280, 312
martingale, 310, 383, 410, 411
  reverse, 385
matrix
  Bernstein inequality, 145, 147
  exponential, 391
  function, 391
  logarithm, 391
  norm, 391
maximal
  antichain, 250
  slope of a convex function, 224
mean
  integrated
    pth power error, 126
    squared error, 110
  squared
    error, 4, 108
    prediction error, 53, 55, 170
  vector, 327
measurable
  function, 324
  set, 323
  space, 323
measure, 323
  Borel, 324
  counting, 323
  finite, 323
  inner regular, 331
  Lebesgue, 323
  Lebesgue–Stieltjes, 328
  outer regular, 331
  probability, 324
  product, 329, 408
  Radon, 331
  σ-finite, 323
  space, 323
    separable, 378
  theory, 323
mediator, 196
Meinshausen's procedure, 46
Mercer's theorem, 167, 169, 173, 183, 371, 372, 410
M-estimator, 84
metric, 315
  bounded Lipschitz, 38, 401
  discrete, 315
  Euclidean, 315
  space, 315, 408
    closed ball, 316
    closed set, 316
    complete, 316
    convergence, 315
    open ball, 315
    open set, 315
    separable, 378
    sequentially compact, 319
    totally bounded, 316
min–max
  formula, 295, 297, 300, 309, 313
  principle, 397, 398, 410
minimax
  concave penalty, 92, 105, 208
  estimator, 11, 246
  lower bound, 221, 252
    high-probability, 240, 253
  quantile, 240
  rate, 13
  risk, 11, 222
minimum
  distance test, 222
  signal strength condition, 77
Minkowski functional, 313
Minkowski's inequality, 371
mixed primal–dual bases algorithm, 303
monotone
  cone, 231
  convergence theorem, 328, 408
    conditional version, 215, 381
Moore–Penrose pseudoinverse, 66, 156, 171, 340, 395, 410
multi-index notation, 125
multiple testing, 38

https://doi.org/10.1017/9781009160421.011 Published online by Cambridge University Press

multiplicity of an eigenvalue, 166
multivariate
  central limit theorem, 402, 407
  normal distribution, 35, 349, 409
  t-distribution, 27, 43

Nadaraya–Watson estimator, 132, 147
natural cubic
  smoothing spline, 141
  spline, 140, 145, 163
    interpolant, 140
neighbourhood
  (topology), 320
  stability condition, 105
Nikol'ski class, 114, 123
  anisotropic, 128
node of a graph, 188
nodewise regression, 207, 220
non-descendants (in a graph), 189
non-negative simple function, 327
normal
  distribution, 348
  linear model, 25
normed space, 364
nuclear norm of a matrix, 392

one-hot coding, 89
one-sided directional derivative, 79, 335
open
  ball in a metric space, 315
  box, 322
  cover, 317
  set in a metric space, 315
  set in a topological space, 317
operator norm, 367, 391
  p to q, 391
oracle inequality, 281, 300, 313
order
  notation, 402
  of a kernel, 111
  statistics, 270
ordinary least squares, 2, 22, 48, 50, 287
orthogonal
  complement (of a subspace), 366
  design, 61
  direct sum, 176, 367
  projection, 367
    matrix, 22
orthogonally invariant norm, 392
orthonormal
  basis, 368
  eigenfunctions, 166
  system, 368
outer regular measure, 331
outrigger local polynomial estimator, 147

p to q operator norm, 391
Pólya's criterion, 348
parallelogram law, 364, 377
parameter space, 10
parent (in a graph), 188
Parseval's theorem, 368, 371, 377, 409
partial correlation, 209
  regularised, 210, 220
  sample, 210
partially ordered set, 377
path (in a graph), 188
PC algorithm, 205
permutation matrix, 282
Pinsker's inequality, 226, 228, 231, 246, 247, 252, 386
π-system, 298, 324
pivot, 27
planted clique hypothesis, 254
polarisation identity, 364, 377
Polish space, 382
polynomial kernel, 153
Pool Adjacent Violators Algorithm, 256, 293
portmanteau lemma, 401, 403, 410
positive
  regression dependence on a subset, 48
  semi-definite matrix, 345
positive definite
  function, 345
  kernel, 151
positive-part James–Stein estimator, 12, 20
post-selection inference, 60
potential outcomes causal model, 220
precision matrix, 208, 220
predictor, 21
  matrix, 21
primal problem, 338
principal
  angles between subspaces, 399
  component (normalised), 394
  component (unnormalised), 179, 394
  components analysis, 53, 394
prior distribution, 11
probability
  measure, 324
  space, 324
product
  kernel, 125
  measure, 329, 408
  σ-algebra, 329
  topology, 322
projected covariance measure, 220
projection, 366
  orthogonal, 367
proper subgraph, 188
pseudo-metric space, 221, 229, 315
  separable, 378
pushforward measure, 233, 333
p-value, 3, 34

quantile function, 8

Radon measure, 331

https://doi.org/10.1017/9781009160421.011 Published online by Cambridge University Press

Radon–Nikodym
  derivative, 168, 183, 224, 233, 330, 363
  theorem, 224, 330, 408
random
  design, 21
  element, 324
  feature map, 178, 183
  variable, 324
  vector, 324
randomised test, 230
rank of a bounded linear operator, 379
real part of a complex function, 343
regression
  convex, 301, 314
  function, 130
  isotonic, 231, 236, 245, 292, 296, 313
  nodewise, 220
  spline, 143
  S-shaped, 302, 304, 314
regular
  estimator, 19
  measure, 331
regularisation parameter, 51
regularised partial correlation, 210, 220
relative interior of a convex set, 334
relaxed Lasso estimator, 91, 105
representer
  (of a bounded linear functional), 368
  theorem, 162, 170, 180, 182
reproducing
  kernel, 158
    Hilbert space, 148, 158, 182
  property, 132
residual
  prediction
    function, 33
    test, 34, 82
  sum of squares, 28
  vector, 23
response variable, 21
restricted
  eigenvalue condition, 104
  isometry condition, 104
reverse
  filtration, 385
  martingale, 385
ridge regression estimator, 51, 103, 148
Riemann integral, 328
Riemann–Lebesgue lemma, 348
Riesz representation theorem, 161, 367, 409
Riesz–Fischer theorem, 410
right singular vector, 392
risk
  Bayes, 11, 246
  function, 11
  minimax, 11, 222
  worst-case, 222, 301, 303
Rosenthal's inequality, 127, 128
sample
  partial correlation, 210
  path (of a martingale), 384
SCAD penalty, 105
scale equivariance, 55
scaled
  kernel, 107
  Lasso estimator, 105
Schatten norms, 392
Scheffé's lemma, 261, 332
Schur complement, 218, 390
s-concave density, 284, 313
score
  (principal component), 394
  function, 287
  matching objective, 288
second Borel–Cantelli lemma, 326
second-order Sobolev kernel, 154, 155, 161, 180, 181
selection probability (of a variable), 97
self-adjoint bounded linear operator, 368
semiparametric statistics, 220
separable
  measure space, 378
  pseudo-metric space, 378
separating hyperplane theorem, 342
separation (in a graph), 190
sequentially
  compact metric space, 319
  continuous function, 321
sharp oracle inequality, 281, 300, 313
shift-invariant kernel, 178
Shor's r-algorithm, 273, 311
σ-algebra, 323
  ball, 378
  Borel, 324
  generated by a class of sets, 323
  generated by a collection of functions, 324
  Lebesgue, 324
  product, 329
σ-finite
  measure, 323
  set, 334
sign function, 18
sign-monotonic function, 388
signed general position, 67
simple
  function, 327
  hypothesis, 29
Single-World Intervention Graph, 220
singular
  component of a measure, 234, 330
  measures, 224, 330
  value, 391, 392
    decomposition, 45, 53, 392, 393, 410
  vector, 392
skeleton (of a graph), 201, 216

https://doi.org/10.1017/9781009160421.011 Published online by Cambridge University Press

slack variables, 337
Slater's condition, 340
Slutsky's lemma, 402, 410
smoothed log-concave projection, 275, 312
Sobolev class, 114, 123
soft thresholding, 62
sparse alternative, 32
sparsity, 58
spectral
  norm, 391
  theorem, 369, 378
spline, 139, 147
  cubic, 139, 147, 314
  natural cubic, 140, 163
square-root Lasso estimator, 87, 105, 210
S-shaped
  function, 302
  regression, 302, 304, 314
stability selection, 105
  complementary pairs, 97, 105, 208
stacking, 58, 104
statistical dimension, 313
statistical–computational trade-off, 13
Stein's method, 411
stochastic domination, 44, 258, 259
Strassen's theorem, 275, 312
strictly convex/concave function, 335
strong
  duality, 338
  law of large numbers, 402, 411
structural causal model, 192
sub-Gamma random variable, 356, 409
sub-Gaussian
  random variable, 351, 409
  random vector, 351
sub-multiplicative norm, 392
sub-Poisson random variable, 354, 409
sub-σ-algebra, 381
subcover, 317
subdifferential, 335
subgradient, 335, 408
subgraph (of a graph), 188
subgroup selection, 313
submartingale, 384
subspace topology, 317
super-uniform, 38
superefficiency, 19
supermartingale, 384
supervised learning, 106
support
  convex, 266
  of a function, 379
  vector, 165
    classifier, 165
    machine, 165, 181, 182
supporting hyperplane theorem, 342
symmetric difference of sets, 334
Taylor's theorem
  with integral form of remainder, 115, 363
  with mean value/Lagrange form of remainder, 112, 362
  with Young/Peano form of remainder, 362
t-distribution, 27, 48
tent function, 272
thin singular value decomposition, 53, 394, 400
Tikhonov regularisation, 103
topological
  ordering (in a DAG), 189, 216
  space, 317, 408
    compact, 318
    first countable, 321
    Hausdorff, 320
    locally compact, 320
    open box, 322
    open set, 317
    Polish, 382
topology, 316
  cofinite, 317
  discrete, 317
  Euclidean, 317
  indiscrete, 317
  product, 322
  subspace, 317
total variation distance, 225
totally bounded metric space, 316
tower law
  for conditional expectation, 381
trace-class operator, 171, 369
training data, 56
translation invariance of Lebesgue measure, 332
truncated power series basis, 143
t-test, 3, 28, 44
tuning parameter, 51
Tychonoff's theorem, 322, 408
unbiased estimator, 4, 23
uncorrelated random variables, 349
undirected
  edge (of a graph), 188
  graph, 189, 219
uniform
  central limit theorem, 405
  continuity, 316
  integrability condition, 405
  kernel, 107
  Type I error control, 404
  weak law of large numbers, 405
uniformly
  Cauchy sequence of functions, 322
  most powerful test, 32
unimodal density, 306
uniqueness of extension of measures, 325

https://doi.org/10.1017/9781009160421.011 Published online by Cambridge University Press

unsupervised learning, 106
upper set, 296
Urysohn's lemma, 321, 379, 408

value function, 338
Vandermonde matrix, 131, 147, 180
variance
  (of an estimator), 4
  factor, 351
  function, 130
  proxy, 351
vertex of a graph, 188
v-structure (in a graph), 201

warm start, 81
Wasserstein distance, 277
weak
  convergence, 401
  derivative, 363
  duality, 338
  union property, 185, 191, 198, 215
Wedin's theorem, 410
weighted covariance measure, 220
Weyl's inequality, 173, 397, 410
Woodbury formula, 100, 390
worst-case risk, 11, 222, 301, 303

Young's inequality, 127, 128

Z-estimator, 287
Zorn's lemma, 368, 377

https://doi.org/10.1017/9781009160421.011 Published online by Cambridge University Press