CAMBRIDGE

Cambridge University Press
978-1-108-47004-9 — Mathematics for Machine Learning
Marc Peter Deisenroth , A. Aldo Faisal , Cheng Soon Ong
Index
More Information

Index

1-of-K representation, 328
ℓ_1 norm, 58
ℓ_2 norm, 59

abduction, 231
Abel–Ruffini theorem, 302
Abelian group, 25
absolutely homogeneous, 58
activation function, 284
affine mapping, 50
affine subspace, 48
Akaike information criterion, 259
algebra, 8
algebraic multiplicity, 89
analytic, 123
ancestral sampling, 307, 329
angle, 63
associativity, 14, 16, 25
attribute, 227
augmented matrix, 19
auto-encoder, 310
automatic differentiation, 140
automorphism, 37

backpropagation, 139
basic variable, 20
basis, 33
basis vector, 34
Bayes factor, 258
Bayes' law, 164
Bayes' rule, 164
Bayes' theorem, 164
Bayesian GP-LVM, 313
Bayesian inference, 245
Bayesian information criterion, 259
Bayesian linear regression, 273
Bayesian model selection, 257
Bayesian network, 250, 254
Bayesian PCA, 312
Bernoulli distribution, 183
Beta distribution, 184
bilinear mapping, 59
bijective, 37
binary classification, 335
Binomial distribution, 183, 184
blind-source separation, 312
Borel σ-algebra, 160

canonical basis, 34
canonical feature map, 352
canonical link function, 284
categorical variable, 159
Cauchy–Schwarz inequality, 61
change-of-variable technique, 195
characteristic polynomial, 88
Cholesky decomposition, 97
Cholesky factor, 97
Cholesky factorization, 97
class, 335
classification, 283
closure, 25
code, 310
codirected, 89
codomain, 46, 120
collinear, 89
column, 13
column space, 47
column vector, 13, 27
completing the squares, 276
concave function, 211
condition number, 206
conditional probability, 158
conditionally independent, 173
conjugate, 185
conjugate prior, 185
convex conjugate, 217
convex function, 211
convex hull, 350
convex optimization problem, 211, 214
convex set, 211
coordinate, 38
coordinate representation, 38
coordinate vector, 38
correlation, 170
covariance, 168, 169
covariance matrix, 169, 175
covariate, 227
CP decomposition, 117
cross-covariance, 169
cross-validation, 231, 236
cumulative distribution function, 158, 161

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-108-47004-9 — Mathematics for Machine Learning
Marc Peter Deisenroth , A. Aldo Faisal , Cheng Soon Ong
Index
More Information

d-separation, 252
data covariance matrix, 287
data point, 227
data-fit term, 272
decoder, 310
deep auto-encoder, 313
defective, 94
denominator layout, 130
derivative, 122
design matrix, 264, 266
determinant, 83
diagonal matrix, 98
diagonalizable, 98
diagonalization, 99
difference quotient, 122
dimension, 34
dimensionality reduction, 286
directed graphical model, 249, 250, 254
direction, 48
direction space, 48
distance, 62
distribution, 157
distributivity, 14, 16
domain, 46, 120
dot product, 59
dual SVM, 349

Eckart–Young theorem, 113, 301
eigendecomposition, 99
eigenspace, 89
eigenspectrum, 89
eigenvalue, 88
eigenvalue equation, 88
eigenvector, 88
elementary transformations, 18
EM algorithm, 325
embarrassingly parallel, 237
empirical covariance, 171
empirical mean, 170
empirical risk, 233
empirical risk minimization, 231, 233
encoder, 310
endomorphism, 37
epigraph, 212
equivalent, 44
error function, 264
error term, 346
Euclidean distance, 59, 62
Euclidean norm, 58
Euclidean vector space, 60
event space, 155
evidence, 165, 256, 275
example, 227
expected risk, 234
expected value, 166
exponential family, 183, 188
extended Kalman filter, 149

factor analysis, 312
factor graph, 254
feature, 227
feature map, 228
feature matrix, 266
feature vector, 266
Fisher discriminant analysis, 117
Fisher–Neyman theorem, 187
forward mode, 141
free variable, 20
full rank, 36
full SVD, 110
fundamental theorem of linear mappings, 48

Gaussian elimination, 21
Gaussian mixture model, 315
Gaussian process, 284
Gaussian process latent-variable model, 313
general linear group, 26
general solution, 18, 20
generalized linear model, 244, 284
generating set, 33
generative process, 244, 257
generator, 311
geometric multiplicity, 91
Givens rotation, 79
global minimum, 201
GP-LVM, 313
gradient, 127
Gram matrix, 352
Gram–Schmidt orthogonalization, 74
graphical model, 250
group, 25

Hadamard product, 13
hard margin SVM, 342
Hessian, 144
Hessian eigenmaps, 117
Hessian matrix, 144
hinge loss, 345
histogram, 333
hyperparameter, 231
hyperplane, 48, 49
hyperprior, 252

i.i.d., 173
identity automorphism, 37
identity mapping, 37
identity matrix, 14
image, 46, 120
independent and identically distributed, 173, 233, 239
independent component analysis, 312
inference network, 311
injective, 37
inner product, 60
inner product space, 60
intermediate variables, 141
inverse, 14

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE
Cambridge University Press
978-1-108-47004-9 — Mathematics for Machine Learning
Marc Peter Deisenroth , A. Aldo Faisal , Cheng Soon Ong
Index
More Information

inverse element, 25
invertible, 14
Isomap, 117
isomorphism, 37

Jacobian, 127, 130
Jacobian determinant, 132
Jeffreys–Lindley paradox, 258
Jensen's inequality, 213
joint probability, 158

Karhunen–Loève transform, 286
kernel, 22, 36, 46, 228, 352
kernel density estimation, 333
kernel matrix, 352
kernel PCA, 313
kernel trick, 284, 313, 352

label, 228
Lagrange multiplier, 209
Lagrangian, 209
Lagrangian dual problem, 209
Laplace approximation, 149
Laplace expansion, 86
Laplacian eigenmaps, 117
LASSO, 272, 285
latent variable, 247
law, 157, 160
leading coefficient, 20
least-squares loss, 134
least-squares problem, 234
least-squares solution, 73
left-singular vectors, 101
Legendre transform, 217
Legendre–Fenchel transform, 217
length, 58
likelihood, 165, 238, 241, 262
line, 49, 68
linear combination, 29
linear manifold, 48
linear mapping, 37
linear program, 214
linear subspace, 28
linear transformation, 37
linearly dependent, 29
linearly independent, 29
link function, 244
loading, 291
local minimum, 201
log-partition function, 189
logistic regression, 284
logistic sigmoid, 284
loss function, 233, 345
loss term, 346
lower-triangular matrix, 84

Maclaurin series, 123
Manhattan norm, 58
MAP, 270
MAP estimation, 241
margin, 339
marginal, 169
marginal likelihood, 165, 257, 275
marginal probability, 158
marginalization property, 163
Markov random field, 254
matrix, 12
matrix factorization, 82
maximum a posteriori, 270
maximum a posteriori estimation, 241
maximum likelihood, 231
maximum likelihood estimate, 267
maximum likelihood estimation, 238, 263
mean, 167
mean function, 278
mean vector, 175
measure, 160
median, 167
metric, 62
minimal, 33
minimax inequality, 210
misfit term, 272
mixture model, 315
mixture weight, 315
mode, 167
model, 225
model evidence, 257
model selection, 231
Moore–Penrose pseudo-inverse, 24
multidimensional scaling, 117
multiplication by scalars, 26
multivariate, 158
multivariate Gaussian distribution, 175
multivariate Taylor series, 145

natural parameters, 189
negative log-likelihood, 238
nested cross-validation, 231, 255
neutral element, 25
noninvertible, 14
nonsingular, 14
norm, 58
normal distribution, 175
normal equation, 72
normal vector, 66
null space, 22, 36, 46
numerator layout, 130

Occam's razor, 256
ONB, 65
one-hot encoding, 328
ordered basis, 38
orthogonal, 63
orthogonal basis, 65
orthogonal complement, 65
orthogonal matrix, 64
orthonormal, 63
orthonormal basis, 65

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-108-47004-9 — Mathematics for Machine Learning
Marc Peter Deisenroth , A. Aldo Faisal , Cheng Soon Ong
Index
More Information

outer product, 27
overfitting, 235, 243, 269

PageRank, 96
parameters, 49
parametric equation, 49
partial derivative, 126
particular solution, 17, 19
PCA, 286
pdf, 160
penalty term, 236
pivot, 20
plane, 49
plate, 252
population mean and covariance, 170
positive definite, 58, 59, 60, 62
posterior, 165, 241
posterior odds, 257
power iteration, 302
power series representation, 125
PPCA, 307
preconditioner, 206
predictor, 3, 228
primal problem, 209
principal component, 291
principal component analysis, 117, 286
principal subspace, 294
prior, 164, 241
prior odds, 257
probabilistic programming, 249
probabilistic inverse, 165
probabilistic PCA, 307
probability, 155
probability density function, 160
probability distribution, 152
probability integral transform, 193
probability mass function, 157, 158
product rule, 164
projection, 68
projection error, 73
projection matrix, 68
pseudo-inverse, 72

random variable, 152, 155
range, 46
rank, 35
rank deficient, 36
rank-k approximation, 112
rank-nullity theorem, 48
raw-score formula for variance, 171
recognition network, 311
reconstruction error, 73, 295
reduced hull, 351
reduced row-echelon form, 21
reduced SVD, 110
REF, 19
regression, 260
regular, 14
regularization, 235, 272, 346
regularization parameter, 236, 272, 344
regularized least squares, 272
regularizer, 236, 272, 344, 346
representer theorem, 348
responsibility, 318
reverse mode, 141
right-singular vectors, 101
RMSE, 268
root mean square error, 268
rotation, 76
rotation matrix, 77
row, 13
row vector, 13, 27
row-echelon form, 19, 20

sample mean, 170
sample space, 154
scalar, 26
scalar product, 59
sigmoid, 190
similar, 44
singular, 14
singular value decomposition, 101
singular value equation, 106
singular value matrix, 101
singular values, 101
slack variable, 344
soft margin SVM, 344
solution, 11
span, 33
special solution, 17
spectral clustering, 117
spectral norm, 112
spectral theorem, 94
spectrum, 89
square matrix, 15
standard basis, 34
standard deviation, 169
standard normal distribution, 176
standardization, 304
statistical independence, 172
statistical learning theory, 237
stochastic gradient descent, 206
strong duality, 211
sufficient statistics, 187
sum rule, 163
support point, 48
support vector, 348
supporting hyperplane, 217
surjective, 37
SVD, 101
SVD theorem, 101
symmetric, 59, 62
symmetric matrix, 15
symmetric, positive definite, 60
symmetric, positive semidefinite, 60
system of linear equations, 11

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-108-47004-9 — Mathematics for Machine Learning
Marc Peter Deisenroth , A. Aldo Faisal , Cheng Soon Ong
Index
More Information

target space, 155
Taylor polynomial, 123, 145
Taylor series, 123
test error, 270
test set, 235, 255
Tikhonov regularization, 237
trace, 87
training, 4
training error, 270
training set, 233, 263
transfer function, 284
transformation matrix, 40
translation vector, 50
transpose, 15, 27
triangle inequality, 58, 62
truncated SVD, 110
Tucker decomposition, 117

underfitting, 243
undirected graphical model, 254
uniform distribution, 161
univariate, 158
unscented transform, 149
upper-triangular matrix, 84

validation set, 236, 255
variable selection, 285
variance, 169
vector, 26
vector addition, 26
vector space, 26
vector space homomorphism, 37
vector space with inner product, 60
vector subspace, 28

weak duality, 210

zero-one loss, 345

© in this web service Cambridge University Press www.cambridge.org