Index

1-of-K representation, 364
ℓ_2 norm, 72
ℓ_1 norm, 71

abduction, 258
Abel-Ruffini theorem, 334
Abelian group, 36
absolutely homogeneous, 71
activation function, 315
affine mapping, 63
affine subspace, 61
Akaike information criterion, 288
algebra, 17
algebraic multiplicity, 106
analytic, 143
ancestral sampling, 340, 364
angle, 76
associativity, 23, 25, 36
attribute, 253
augmented matrix, 29
auto-encoder, 343
automatic differentiation, 161
automorphism, 49

backpropagation, 159
basic variables, 30
basis, 44
basis vector, 45
Bayes factor, 286
Bayes' law, 185
Bayes' rule, 185
Bayes' theorem, 185
Bayesian GP-LVM, 347
Bayesian inference, 273
Bayesian information criterion, 288
Bayesian linear regression, 303
Bayesian model selection, 285
Bayesian network, 278, 282
Bayesian PCA, 346
Bernoulli distribution, 205
Beta distribution, 206
bijective, 48
bilinear mapping, 72
binary classification, 370
Binomial distribution, 206
blind-source separation, 346
Borel σ-algebra, 180

canonical basis, 45
canonical feature map, 389
canonical link function, 315
categorical variable, 180
Cauchy-Schwarz inequality, 75
change of variable, 218
characteristic polynomial, 104
Cholesky decomposition, 114
Cholesky factor, 114
Cholesky factorization, 114
class, 370
classification, 315
closure, 36
code, 343
codirected, 105
codomain, 58, 139
collinear, 105
column, 22
column space, 59
column vector, 22, 38
completing the squares, 307
concave function, 237
condition number, 230
conditional probability, 179
conditionally independent, 195
conjugate, 208
conjugate prior, 208
convex conjugate, 242
convex function, 237
convex hull, 386
convex optimization problem, 236, 239
convex set, 236
coordinate, 50
coordinate representation, 50
coordinate vector, 50
correlation, 191
covariance, 190
covariance matrix, 190, 198
covariate, 253
CP decomposition, 136
cross validation, 257, 263
cross-covariance, 191
cumulative distribution function, 178, 181

d-separation, 281

Draft (April 8, 2019) of "Mathematics for Machine Learning" ©2019 by Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong. To be published by Cambridge University Press. Please do not post or distribute this file, please link to `https://mml-book.com`.

data covariance matrix, 318
data point, 253
data-fit term, 302
decoder, 343
deep auto-encoder, 347
defective, 111
denominator layout, 151
derivative, 141
design matrix, 294, 296
determinant, 99
diagonal matrix, 115
diagonalizable, 116
diagonalization, 116
difference quotient, 141
dimension, 45
dimensionality reduction, 317
directed graphical model, 277, 278, 282
direction, 61
direction space, 61
distance, 75
distribution, 177
distributivity, 24, 26
domain, 58, 139
dot product, 72
dual SVM, 385

Eckart-Young theorem, 131, 334
eigendecomposition, 116
eigenspace, 106
eigenspectrum, 106
eigenvalue, 105
eigenvalue equation, 105
eigenvector, 105
elementary transformations, 28
EM algorithm, 360
embarrassingly parallel, 264
empirical covariance, 192
empirical mean, 192
empirical risk, 260
empirical risk minimization, 257, 260
encoder, 343
endomorphism, 49
epigraph, 237
equivalent, 56
error function, 294
error term, 382
Euclidean distance, 72, 75
Euclidean norm, 72
Euclidean vector space, 73
event space, 175
evidence, 186, 285, 306
example, 253
expected risk, 261
expected value, 187
exponential family, 205, 211
extended Kalman filter, 170

factor analysis, 346
factor graph, 282
feature, 253
feature map, 254
feature matrix, 296
feature vector, 295
Fisher discriminant analysis, 136
Fisher-Neyman theorem, 210
forward mode, 161
free variables, 30
full rank, 47
full SVD, 128
fundamental theorem of linear mappings, 60

Gaussian elimination, 31
Gaussian mixture model, 349
Gaussian process, 316
Gaussian process latent variable model, 347
general linear group, 37
general solution, 28, 30
generalized linear models, 315
generating set, 44
generative process, 272, 285
generator, 344
geometric multiplicity, 108
Givens rotation, 94
global minimum, 225
GP-LVM, 347
gradient, 146
Gram matrix, 389
Gram-Schmidt orthogonalization, 89
graphical model, 278
group, 36

Hadamard product, 23
hard margin SVM, 377
Hessian, 164
Hessian eigenmaps, 136
Hessian matrix, 165
hinge loss, 381
histogram, 369
hyper-prior, 280
hyperparameter, 258
hyperplane, 61, 62

i.i.d., 194
ICA, 346
identity automorphism, 49
identity mapping, 49
identity matrix, 23
image, 58, 139
independent and identically distributed, 194, 260, 266
independent component analysis, 346
inference network, 344
injective, 48
inner product, 73
inner product space, 73
intermediate variables, 162
inverse, 24

inverse element, 36
invertible, 24
Isomap, 136
isomorphism, 49

Jacobian, 146, 150
Jacobian determinant, 152
Jeffreys-Lindley paradox, 286
Jensen's inequality, 238
joint probability, 178

Karhunen-Loève transform, 318
kernel, 33, 47, 58, 254, 388
kernel density estimation, 369
kernel matrix, 389
kernel PCA, 347
kernel trick, 316, 347, 389

label, 253
Lagrange multipliers, 234
Lagrangian, 234
Lagrangian dual problem, 234
Laplace approximation, 170
Laplace expansion, 102
Laplacian eigenmaps, 136
LASSO, 303, 316
latent variables, 275
law, 177, 181
leading coefficient, 30
least-squares loss, 154
least-squares problem, 261
least-squares solution, 88
left-singular vectors, 119
Legendre transform, 242
Legendre-Fenchel transform, 242
length, 71
likelihood, 185, 265, 268, 291
line, 61, 82
linear combination, 40
linear manifold, 61
linear mapping, 48
linear program, 239
linear subspace, 39
linear transformation, 48
linearly dependent, 40
linearly independent, 40
loading, 322
local minimum, 225
log-partition function, 211
logistic regression, 315
logistic sigmoid, 315
loss function, 260, 381
loss term, 382
lower-triangular matrix, 101

Maclaurin series, 143
Manhattan norm, 71
MAP, 300
MAP estimation, 269
margin, 374
marginal, 190
marginal likelihood, 186, 286, 306
marginal probability, 179
marginalization property, 184
Markov random field, 282
matrix, 22
matrix factorization, 98
maximum a posteriori, 300
maximum a posteriori estimation, 269
maximum likelihood, 257
maximum likelihood estimate, 296
maximum likelihood estimation, 265, 293
mean, 187
mean function, 309
mean vector, 197
measure, 180
median, 188
metric, 76
minimal, 44
minimax inequality, 234
misfit term, 302
mixture models, 349
mixture weights, 349
mode, 188
model, 251
model evidence, 286
model selection, 258
Moore-Penrose pseudo-inverse, 35
multidimensional scaling, 136
multiplication by scalars, 37
multivariate, 178
multivariate Gaussian distribution, 197
multivariate Taylor series, 166

natural parameters, 211
negative log-likelihood, 265
nested cross validation, 258, 284
neutral element, 36
non-invertible, 24
non-singular, 24
norm, 71
normal distribution, 197
normal equation, 86
normal vector, 80
null space, 33, 47, 58
numerator layout, 150

Occam's razor, 284
ONB, 79
one-hot encoding, 364
ordered basis, 50
orthogonal, 77
orthogonal basis, 79
orthogonal complement, 79
orthogonal matrix, 78
orthonormal, 77
orthonormal basis, 79
outer product, 38
overfitting, 262, 270, 299

©2019 M. P. Deisenroth, A. A. Faisal, C. S. Ong. To be published by Cambridge University Press.

PageRank, 114
parameters, 61
parametric equation, 61
partial derivative, 146
particular solution, 27, 30
PCA, 317
pdf, 181
penalty term, 262
pivot, 30
plane, 62
plate, 280
population mean and covariance, 191
positive definite, 71, 73, 74, 76
posterior, 185, 268
posterior odds, 286
power iteration, 334
power series representation, 145
PPCA, 340
preconditioner, 230
predictor, 255
predictors, 12
primal problem, 234
principal component, 322
principal component analysis, 136, 317
principal subspace, 326
prior, 185, 268
prior odds, 286
probabilistic inverse, 186
probabilistic PCA, 340
probabilistic programming, 277
probability, 175
probability density function, 181
probability distribution, 172
probability integral transform, 216
probability mass function, 178
product rule, 184
projection, 82
projection error, 88
projection matrix, 82
pseudo-inverse, 86

random variable, 172, 175
range, 58
rank, 47
rank deficient, 47
rank-k approximation, 130
rank-nullity theorem, 60
raw-score formula for variance, 192
recognition network, 344
reconstruction error, 88, 327
reduced hull, 388
reduced row echelon form, 31
reduced SVD, 129
REF, 30
regression, 289
regular, 24
regularization, 262, 302, 382
regularization parameter, 262, 302, 380
regularized least squares, 302
regularizer, 262, 302, 380, 382
representer theorem, 384
responsibility, 352
reverse mode, 161
right-singular vectors, 119
RMSE, 298
root mean squared error, 298
rotation, 91
rotation matrix, 92
row, 22
row echelon form, 30
row vector, 22, 38
row-echelon form, 30

sample mean, 192
sample space, 175
scalar, 37
scalar product, 72
sigmoid, 213
similar, 56
singular, 24
singular value decomposition, 119
singular value equation, 124
singular value matrix, 119
singular values, 119
slack variable, 379
soft margin SVM, 379, 380
solution, 20
span, 44
special solution, 27
spectral clustering, 136
spectral norm, 131
spectral theorem, 111
spectrum, 106
square matrix, 25
standard basis, 45
standard deviation, 190
standard normal distribution, 198
standardization, 336
statistical independence, 194
statistical learning theory, 264
stochastic gradient descent, 231
strong duality, 236
sufficient statistics, 210
sum rule, 184
support point, 61
support vectors, 384
supporting hyperplane, 242
surjective, 48
SVD, 119
SVD theorem, 119
symmetric, 73, 76
symmetric matrix, 25
symmetric, positive definite, 74
symmetric, positive semi-definite, 74
system of linear equations, 20

target space, 175

Taylor polynomial, 142, 166
Taylor series, 142
test error, 300
test set, 261, 284
Tikhonov regularization, 264
trace, 103
training, 12
training error, 300
training set, 260, 292
transfer function, 315
transformation matrix, 51
translation vector, 63
transpose, 25, 38
triangle inequality, 71, 76
truncated SVD, 129
Tucker decomposition, 136

underfitting, 271
undirected graphical model, 282
uniform distribution, 182
univariate, 178
unscented transform, 170
upper-triangular matrix, 101

validation set, 263, 284
variable selection, 316
variance, 190
vector, 37
vector addition, 37
vector space, 37
vector space homomorphism, 48
vector space with inner product, 73
vector subspace, 39

weak duality, 235

zero-one loss, 381

©2019 M. P. Deisenroth, A. A. Faisal, C. S. Ong. To be published by Cambridge University Press.