Subject Index

alignment, 128
anisotropy, 89
AR process, see autoregressive process
ARD, see automatic relevance
determination
ARMA process, see autoregressive moving-average process
automatic relevance
determination, 106
autoregressive moving-average process, 217
noise model, 191
autoregressive process, 207
Bayes classifier, 36
Bayes' theorem, 200
Bayesian committee machine, 180
BCM, see Bayesian committee machine
bias
inductive, 165
binary classification, 33
bits, 203
blitzkrieging, see fast Gaussian processes
Bochner's theorem, 82
Brownian bridge, 213
Brownian motion, see Wiener process
canonical hyperplane, 142
Cholesky decomposition, 202
Christ, bowels of, 111
classification, 33
binary, 33
least-squares, 146
probabilistic, 148
multi-class, 33
probabilistic, 33
classifier
Gibbs, 163
predictive, 163
cokriging, 190
consistency, 155
convex
function, 206
set, 206
covariance
predictive, 18
covariance function, 13, 79, see also kernel
ANOVA, 95
compact support, 87
dot product, 89
exponential, 85
$\gamma$-exponential, 86
Gaussian, see covariance function, squared exponential
inhomogeneous polynomial, 89
Matérn, 84
neural network, 91
Ornstein-Uhlenbeck, 86
OU, see covariance function, Ornstein-Uhlenbeck
periodic, 92, 119
polynomial, 89
piecewise, 87
radial basis function, see covariance function, squared exponential
rational quadratic, 86
RBF, see covariance function, squared exponential
SE, see covariance function, squared exponential
squared exponential, 14, 83
covariance matrix, 80
Cromwell's dictum, 111
cross-validation, 111
generalized, 112
leave-one-out, 111
Subject Index

dataset
  robot, inverse dynamics, 23
  USPS, 63
decision region, 36
decision surface, 36
degenerate, see kernel, degenerate
degrees of freedom, 25
derivative observations, 191
Dirichlet process, 192
discriminative approach, 34
eigenfunction, 96
eigenvalue, 96
entropy, 203
EP, see expectation propagation
ε-insensitive error function, 145
equivalent kernel, 25, 151
error
generalization, 159
error function
  ε-insensitive, 145
  hinge, 145
error-reject curve, 36, 68
errors-in-variables regression, 192
evidence, see marginal likelihood
expectation propagation, 52
factor analysis, 89, 107
feature space, 11
Fisher information matrix, 102
Fisher kernel, 101
Fourier transform, 206
fractal, 137
gamma distribution, 87, 194
Gaussian distribution, 200
Gaussian Markov process, 207
Gaussian process, 13
Gaussian process classification, 34
Gaussian process latent variable model, 196
Gaussian process regression, 16
generalization error, 108, 159
generative approach, 34
generative topographic mapping, 196
geostatistics, 30
GMP, see Gaussian Markov process
GP, see Gaussian process
GPC, see Gaussian process classification
GPLVM, see Gaussian process latent variable model
GPR, see Gaussian process regression
Gram matrix, 80
Green’s function, 134
GTM, see generative topographic mapping
hidden Markov model, 102
hinge error function, 145
hyperparameters, 20, 106
hyperplane
  canonical, 142
index set, 13
informative vector machine, 178
integrals, evaluation of, 193
intrinsic random function, 137
invariances, 195
IRLS, see iteratively reweighted least squares
isotropy, 80
iteratively reweighted least squares, 38
IVM, see informative vector machine
jitter, 47
kernel, 80, see also covariance function
  bag-of-characters, 100
degenerate, 94, 97
equivalent, 151
  Fisher, 101
  k-spectrum, 100
  nondegenerate, 97
  positive definite, 80
  string, 100
  tangent of posterior odds, 102
kernel classifier, 167
kernel PCA, 99
kernel ridge regression, see regression, ridge, kernel
kernel smoother, 25, 167
kernel trick, 12
kriging, 30
Kullback-Leibler divergence, 54, 203
Laplace approximation, 41
latent variable model, 196
learning curve, 159
learning, supervised, 1
least-squares classification, 146
leave-one-out, 148
leave-one-out cross-validation, 111
length-scale, characteristic, 14, 83, 106
likelihood, 8, 37
  logistic, 35, 43
  multiple-logistic, 38
  non-Gaussian, 33, 191
  probit, 35, 43
linear classifier, 37
linear regression, 8
link function, 35
log odds ratio, 37
logistic function, 35
logistic regression, 35
LOO, see leave-one-out
loss
  negative log out, 23
  squared, 22
  zero-one, 36
loss function, 21
loss matrix, 36
LSC, see least-squares classification

MAP, see maximum a posteriori
margin
  functional, 142
  geometrical, 142
marginal likelihood, 18, 112
marginalization property, 13
Markov chain Monte Carlo, 41
Markov random field, 218
matrix
  covariance, 80
  Fisher information, 102
  Gram, 80
  inversion lemma, 201
  loss, 36
  partitioned inversion of, 201
  positive definite, 80
  positive semidefinite, 80
maximum a posteriori, see penalized maximum likelihood
maximum likelihood
  penalized, 10
MCMC, see Markov chain Monte Carlo
mean function, 13, 27
mean square continuity, 81
mean square differentiability, 81
mean standardized log loss, 23
mean-field approximation, 52
measure, 204
Mercer’s theorem, 96
mixture of experts, 192
ML-II, see type II maximum likelihood model
  non-parametric, 166
  parametric, 166
  semi-parametric, 166
Moore-Aronszajn theorem, 130
MS continuity, see mean square continuity
MS differentiability, see mean square differentiability
MSLL, see mean standardized log loss
multi-class classification, 33
multi-task learning, 115
multiple outputs, 190
Nadaraya-Watson estimator, 25, 155
nats, 203
neural network, 90, 166
Newton’s method, 43, 49
noise model, 8, 16
  correlated, 190
  heteroscedastic, 191
norm
  Frobenius, 202
null space, 137
Nyström approximation, 172
Nyström method, 177
Occam’s razor, 110
one-versus-rest, 147
operator, integral, 80
optimal experimental design, 159
outputs, multiple, 190
P1NN, see probabilistic one nearest neighbour
PAC-Bayesian theorem
  McAllester, 163
  Seeger, 164
penalized maximum likelihood estimate, 10
positive definite matrix, 80
positive semidefinite matrix, 80
posterior process, 18
Subject Index

PP, see projected process approximation
  prediction
    classification
      averaged, 44
      MAP, 45
  probabilistic classification, 33
  probabilistic one nearest neighbour, 69
  probability
    conditional, 199
    joint, 199
    marginal, 199
  probit regression, 35
  projected process approximation, 178
  pseudo-likelihood, 117

  quadratic form, 80
  quadratic programming, 142

  regression
    errors-in-variables, 192
    Gaussian process, 16
    linear, 8
    polynomial, 11, 28
    ridge, 11, 132
      kernel, 30, 132
    regularization, 132
    regularization network, 135
  reject option, 36
  relative entropy, see Kullback Leibler divergence
  relevance vector machine, 149
  representer theorem, 132
  reproducing kernel Hilbert space, 129
  response function, 35
  ridge regression, see regression, ridge
  risk, 22, 36
  RKHS, see reproducing kernel Hilbert space
  RVM, see relevance vector machine

  scale mixture, 87, 194
  SD, see subset of datapoints
  SDE, see stochastic differential equation
  SMSE, see standardized mean squared error
  softmax, 38
  splines, 136
  SR, see subset of regressors
  standardized mean squared error, 23
  stationarity, 79
  stochastic differential equation, 207

  Student’s t process, 194
  subset of datapoints, 177
  subset of regressors, 175
  supervised learning, 1
  support vector, 143
  support vector machine, 141
    soft margin, 143
  support vector regression, 145
  SVM, see support vector machine
  SVR, see support vector regression
  tangent of posterior odds kernel, 102
  TOP kernel, see tangent of posterior odds kernel
  transduction, 181
  type II maximum likelihood, 109

  uncertain inputs, 192
  upcrossing rate, 80
  USPS dataset, 63

  weight function, 25
  weight vector, 8
  Wiener process, 213
    integrated, 139
    tied-down, 213
  Wiener-Khintchine theorem, 82, 209

  Yule-Walker equations, 215