
Index

- 7163 1-of- K representation, 344
- 7164 ℓ_1 norm, 69
- 7165 ℓ_2 norm, 70
- 7166 abduction, 243
- 7167 Abelian group, 35
- 7168 activation function, 295
- 7169 affine mapping, 60
- 7170 affine subspace, 59
- 7171 algebra, 17
- 7172 algebraic multiplicity, 106
- 7173 analytic, 140
- 7174 analytic geometry, 14
- 7175 ancestral sampling, 321, 344
- 7176 angle, 74
- 7177 Associativity, 35
- 7178 attributes, 239
- 7179 augmented matrix, 28
- 7180 auto-encoder, 324
- 7181 automatic differentiation, 158
- 7182 Automorphism, 48
- 7183 backpropagation, 156, 158
- 7184 basic variables, 30
- 7185 basis, 44
- 7186 basis vectors, 45
- 7187 Bayes factor, 267
- 7188 Bayes' theorem, 180
- 7189 Bayesian GP-LVM, 328
- 7190 Bayesian inference, 256
- 7191 Bayesian linear regression, 282
- 7192 Bayesian model selection, 266
- 7193 Bayesian networks, 259
- 7194 Bayesian PCA, 326
- 7195 Bernoulli distribution, 202
- 7196 bijective, 48
- 7197 bilinear mapping, 70
- 7198 binary classification, 349
- 7199 Binomial distribution, 203
- 7200 blind-source separation, 327
- 7201 calculus, 14
- 7202 canonical link function, 295
- 7203 canonical/standard basis, 44
- 7204 categorical variables, 175
- 7205 Cauchy-Schwarz inequality, 70
- 7206 change of variable, 192
- 7207 characteristic polynomial, 99
- 7208 Cholesky decomposition, 108
- 7209 Cholesky factor, 108
- 7210 Cholesky factorization, 108
- 7211 classes, 349
- 7212 classification, 15, 294
- 7213 Closure, 35
- 7214 code, 324
- 7215 codirected, 100
- 7216 codomain, 56
- 7217 collinear, 100
- 7218 column space, 57
- 7219 column vectors, 38
- 7220 columns, 22
- 7221 completing the squares, 287
- 7222 concave function, 223
- 7223 condition number, 217
- 7224 conditional probability, 174
- 7225 conditionally independent given z , 185
- 7226 conjugate, 205
- 7227 conjugate priors, 205
- 7228 convex conjugate, 229
- 7229 convex function, 223
- 7230 convex functions, 214, 222
- 7231 convex hull, 365
- 7232 convex optimization problem, 222, 225
- 7233 convex set, 223
- 7234 convex sets, 222
- 7235 coordinate representation, 49
- 7236 coordinate vector, 49
- 7237 coordinates, 49
- 7238 correlation, 185
- 7239 covariance, 182
- 7240 covariance matrix, 183, 195
- 7241 covariates, 239
- 7242 cross validation, 243, 248
- 7243 cross-covariance, 183
- 7244 cumulative distribution function, 173, 176
- 7245 176
- 7246 cyclic permutations, 99
- 7247 d-separation, 262
- 7248 data, 11
- 7249 data covariance matrix, 298
- 7250 data point, 239
- 7251 decoder, 324

- 7252 deep auto-encoder, 327
7253 defective matrix, 111
7254 density estimation, 15
7255 derivative, 139
7256 design matrix, 273, 275
7257 determinant, 94
7258 diagonalizable, 110
7259 Diagonalization, 111
7260 difference quotient, 138
7261 dimension, 45
7262 dimensionality reduction, 15, 297
7263 directed graphical model, 258
7264 Directed graphical models, 259, 263
7265 direction, 59
7266 direction space, 59
7267 distance, 74
7268 domain, 56
7269 dot product, 70
7270 dual SVM, 364
7271 Eckart-Young Theorem, 314
7272 eigenspace, 100
7273 eigenspectrum, 101
7274 eigenvalue, 100
7275 eigenvalue equation, 100
7276 eigenvector, 100
7277 elementary transformations, 28
7278 EM algorithm, 341
7279 embarrassingly parallel, 249
7280 empirical covariance, 183
7281 empirical mean, 183
7282 empirical risk, 245
7283 empirical risk minimization, 242, 245
7284 encoder, 324
7285 Endomorphism, 48
7286 equivalent, 54
7287 error term, 361
7288 Euclidean distance, 70, 74
7289 Euclidean norm, 70
7290 Euclidean vector space, 71
7291 events, 171
7292 evidence, 180, 266, 284
7293 example, 239
7294 expected risk, 246
7295 expected value, 181
7296 exponential family, 202, 207
7297 Extended Kalman Filter, 167
7298 factor analysis, 327
7299 Factor graphs, 263
7300 feature, 239
7301 feature map, 240
7302 feature matrix, 275
7303 feature vector, 275
7304 features, 239
7305 forward mode, 158
7306 four pillars of machine learning, 14
7307 free variables, 30
7308 full rank, 46
7309 full SVD, 125
7310 Gaussian elimination, 31
7311 Gaussian mixture model, 330
7312 Gaussian Process Latent Variable Model,
7313 327
7314 Gaussian processes, 295
7315 general linear group, 36
7316 general solution, 27, 29
7317 generalized linear models, 294
7318 generating set, 44
7319 generative process, 266
7320 generator, 325
7321 geometric multiplicity, 106
7322 Givens rotation, 89
7323 global minimum, 212
7324 gradient, 143
7325 Gram matrix, 367
7326 graphical model, 259
7327 group, 35
7328 hard margin SVM, 357
7329 Hessian, 161
7330 Hessian matrix, 162
7331 hidden variables, 257
7332 hinge loss, 360
7333 hyperparameter, 243
7334 hyperplane, 59
7335 hyperplanes, 60
7336 identity mapping, 48
7337 identity matrix, 23
7338 image, 56
7339 independent and identically distributed,
7340 245, 251
7341 Independent Component Analysis, 327
7342 inference network, 325
7343 injective, 47
7344 inner product, 71
7345 inner product space, 71
7346 intermediate variables, 159
7347 inverse, 24
7348 Inverse element:, 35
7349 invertible, 24
7350 Isomorphism, 48
7351 Jacobian, 143, 147
7352 Jacobian determinant, 149
7353 Jeffreys-Lindley paradox, 267
7354 Jensen's inequality, 225
7355 joint probability, 173
7356 Jordan Normal Form, 111
7357 Karhunen-Loève transform, 297
7358 kernel, 32, 46, 56, 240
7359 kernel matrix, 367
7360 kernel PCA, 327
7361 kernel trick, 295, 327, 367
7362 kernels, 367
7363 label, 15
7364 Lagrange multipliers, 221

- 7365 Lagrangian, 221
7366 Lagrangian dual problem, 221
7367 Laplace approximation, 167
7368 LASSO, 280, 296
7369 latent variables, 257
7370 leading coefficient, 30
7371 learning, 12
7372 least squares solution, 85
7373 least-squares loss, 151
7374 left-singular vectors, 115
7375 Legendre transform, 229
7376 Legendre-Fenchel transform, 229
7377 length, 69
7378 likelihood, 180, 250, 252, 271
7379 linear algebra, 13
7380 linear combination, 39
7381 linear manifold, 59
7382 linear mapping, 47
7383 linear subspace, 38
7384 linear transformation, 47
7385 linearly dependent, 40
7386 linearly independent, 40
7387 lines, 59, 79
7388 local minimum, 212
7389 log partition function, 208
7390 logistic regression, 295
7391 logistic sigmoid, 295
7392 loss function, 245, 359
7393 loss term, 361
7394 lower triangular matrix, 95
7395 Maclaurin series, 140
7396 Manhattan norm, 69
7397 MAP, 280
7398 MAP estimation, 256
7399 margin, 353
7400 marginal likelihood, 266, 284
7401 marginal probability, 174
7402 marginalization property, 179
7403 matrix, 18, 21
7404 matrix decomposition, 14
7405 matrix factorization, 92
7406 maximum a posteriori estimation, 253, 256
7407 256
7408 maximum a-posteriori, 280
7409 maximum likelihood, 242
7410 maximum likelihood estimate, 276
7411 maximum likelihood estimation, 250, 272
7412 272
7413 mean, 181
7414 mean function, 289
7415 mean vector, 195
7416 median, 181
7417 metric, 74
7418 minimal, 44
7419 mixture models, 330
7420 mixture weights, 330
7421 mode, 181
7422 model, 12, 237
7423 model evidence, 266
7424 model selection, 243
7425 Moore-Penrose pseudo-inverse, 34
7426 multiplication by scalars, 37
7427 multivariate Gaussian distribution, 195
7428 multivariate Taylor series, 162
7429 natural parameters, 208
7430 negative log likelihood, 250
7431 nested cross validation, 264
7432 Neutral element:, 35
7433 non-invertible, 24
7434 non-singular, 24
7435 norm, 69
7436 normal distribution, 194
7437 normal equation, 83
7438 null space, 32, 46, 56
7439 Occam's Razor, 265
7440 ONB, 76
7441 one-hot encoding, 344
7442 optimization, 14
7443 ordered basis, 49
7444 orthogonal, 75
7445 orthogonal basis, 76
7446 orthogonal complement, 309
7447 orthogonal matrix, 75
7448 orthonormal, 75
7449 orthonormal basis, 76
7450 overfitting, 247, 279
7451 PageRank, 108
7452 parameters, 59, 257
7453 parametric equation, 59
7454 partial derivative of a vector-valued
7455 function, 147
7456 partial derivatives, 143
7457 particular solution, 27, 29
7458 pivot, 30
7459 planes, 60
7460 plate, 261
7461 population mean and covariance, 183
7462 Positive definite, 69
7463 positive definite, 71, 72, 74
7464 posterior, 180, 252
7465 posterior distribution over models, 266
7466 posterior odds, 267
7467 power iteration, 315
7468 power series representations, 142
7469 PPCA likelihood, 323
7470 preconditioner, 217
7471 predictor, 240
7472 predictors, 11
7473 primal problem, 221
7474 principal component, 302
7475 principal component analysis, 297
7476 principal subspace, 306
7477 prior, 180, 252
7478 Probabilistic PCA, 320

- 7479 probability, 172
7480 probability density function, 176
7481 probability distribution, 169
7482 probability integral transform, 191
7483 probability mass function, 173, 174
7484 probability theory, 14
7485 product rule, 179
7486 projection, 79
7487 projection error, 84
7488 projection matrices, 79
7489 pseudo-inverse, 83
7490 random variable, 169, 172
7491 range, 56
7492 rank, 46
7493 rank deficient, 46
7494 rank-k approximation, 127
7495 recognition network, 325
7496 reconstruction error, 84, 307
7497 reduced hull, 366
7498 reduced row echelon form, 30
7499 reduced SVD, 125
7500 regression, 15, 269
7501 regular, 24
7502 regularization, 247, 279
7503 regularization parameter, 359
7504 regularizer, 280, 281, 359, 361
7505 representer theorem, 363
7506 responsibilities, 334
7507 reverse mode, 158
7508 right-singular vectors, 115
7509 root mean squared error (RMSE), 278
7510 rotation, 86
7511 rotation matrix, 87
7512 row echelon form, 30
7513 row vectors, 38
7514 row-echelon form (REF), 29
7515 row/column vectors, 22
7516 rows, 22
7517 sample mean, 183
7518 scalar product, 70
7519 scalars, 37
7520 similar, 54
7521 singular, 24
7522 Singular Value Decomposition, 115
7523 singular value equation, 121
7524 singular value matrix, 115
7525 singular values, 115
7526 slack variable, 358
7527 soft margin SVM, 358, 359
7528 solution, 20
7529 span, 44
7530 special solution, 27
7531 square matrices, 25
7532 standard deviation, 182
7533 standard normal distribution, 195
7534 standardization, 317
7535 state space, 171
7536 statistical learning theory, 249
7537 statistically independent, 184
7538 Stochastic gradient descent, 218
7539 strong duality, 222
7540 sufficient statistics, 206
7541 sum rule, 179
7542 supervised learning, 15
7543 support point, 59
7544 support vectors, 363
7545 supporting hyperplane, 228
7546 surjective, 47
7547 SVD, 115
7548 symmetric, 25, 71, 74
7549 symmetric, positive definite, 72
7550 symmetric, positive semi-definite, 72
7551 system of linear equations, 20
7552 Taylor polynomial, 140, 163
7553 Taylor series, 140
7554 test error, 279
7555 test set, 247, 265
7556 trace, 98
7557 training, 11
7558 training error, 279
7559 training set, 245, 272
7560 transfer function, 295
7561 transformation matrix, 50
7562 translation vector, 60
7563 transpose, 24, 38
7564 Triangle inequality, 69
7565 Triangular inequality, 74
7566 truncated SVD, 126
7567 Undirected graphical models, 263
7568 uniform distribution, 176
7569 unknowns, 20
7570 unscented transform, 167
7571 unsupervised learning, 15
7572 upper triangular matrix, 95
7573 validation set, 248, 265
7574 variable selection, 296
7575 variance, 182, 183
7576 vector addition, 37
7577 vector space, 36
7578 vector space homomorphism, 47
7579 vector space with inner product, 71
7580 vector subspace, 38
7581 vectors, 37
7582 weak duality, 222