

# Index

- $F_1$  score, 302, 311
- $K$ -means clustering, 17, 242, 245, 351, 355
- $K$ -nearest neighbor graph, 97
- $K$ -nearest neighbors (KNN), 42, 90
- $L^1$  norm, 37, 99, 256, 340, 344
- $L^2$  norm, 36, 37, 54–56, 59, 67, 95, 99, 256, 266, 324, 340, 344, 356
- $L^p$  norm for vectors, 340
- $\chi^2$  test, 304
- $\delta$ -tolerance closed frequent pattern mining, 228, 235
- $t$ -test, 94
  
- accuracy, 300
- AdaBoost, 81
- adjacency matrix, 20, 96, 109, 238, 248, 260, 344
- adjusted rand index (ARI), 314
- agglomerative hierarchical clustering, 27, 29, 184
- all subsequence kernel, 170, 209
- all subtree kernel, 208, 211
- alternating least squares, 347
- alternating least squares (ALS) algorithm, 36, 55, 56, 249, 250
- Apriori algorithm, 122
- area under the precision-recall (AUPR), 311
- area under the ROC curve (AUC), 46, 94, 308
- artificial neural network, *see* layered neural network
  
- back-propagation (BP), 67
- bag of words (BOW), 119, 127, 152
- Baum-Welch algorithm, 159, 166
  
- Bayes estimation, 353
- Bayes theorem, 25, 132, 353
- Bayesian belief network (BBN), 32
- Bayesian belief network classifier, 53
- Bayesian learning, 127
- Bernoulli distribution, 327
- Bernoulli trial, 326
- beta distribution, 328
- bi-factorization, 35
- bicluster, 32
- biclustering, 32, 110
- binary classification, 42, 117
- binary tree, 27
- Binomial distribution, 327
- boosting, 83
- bootstrap aggregating (bagging), 74
- bootstrapping, 73, 296
- breadth-first search (BFS), 191
  
- candidate generation, 122, 146
- canonical correlation analysis (CCA), 265, 351
- canonical form, 209
- case-based reasoning (CBR), 43, 90
- Cholesky decomposition, 336
- classification, 42, 115
- closed frequent pattern mining, 235
- closed frequent pattern mining, 228
- cluster assignment matrix, 18, 40, 101, 240, 242, 248
- clustering, 4, 16, 40
- collaborative matrix factorization, 288
- computational linguistics, 8
- conditional random field (CRF), 168
- conjugate prior distribution, 133, 354
- connected subgraph, 11

- constrained  $K$ -means clustering, 23, 98, 102, 244, 246, 285
- context free grammar (CFG), 164
- contingency table, 298
- convex optimization, 346
- correlation coefficient, 94, 330
- covariance, 329, 363
- covariance matrix, 96, 104, 269
- cross-validation, 109, 296
  
- data fusion, 13
- data sparsity, 5, 8, 239
- data transformation, 139
- decision stump, 44, 92, 98
- decision tree, 47, 63, 74, 75, 98, 116
- deep learning, 70, 72
- dendrogram, 27, 28, 110
- depth-first search (DFS), 149, 150, 191, 221
- dimensionality reduction, 92
- directed acyclic graph, 33
- Dirichlet distribution, 130, 131, 133, 328
- discriminative model, 26, 62, 169
- divisive hierarchical clustering, 27
- downward closure property, 121, 146, 220, 276
- drug-target interaction, 111
- dynamic programming, 158, 162, 182, 184, 185, 197, 202, 277
  
- E-commerce (electric commerce), 5, 8, 15, 121, 261, 263
- eigenvalue decomposition, *see* eigenvalue problem
- eigenvalue problem, 22, 23, 34, 40, 95, 97, 98, 108, 242, 260, 267, 286, 333
- elastic net, 100, 256, 270
- ensemble learning, 63, 64, 72
- entropy, 45, 325, 331
- enumeration tree, 225, 231
- error rate, 301
- exclusive OR (XOR), 62
- Expectation and Maximization (EM) algorithm, 25, 128, 155
- Expectation-Maximization (EM) algorithm, 159, 352
  
- expected reciprocal rank (ERR), 321
- feature learning, 91
- feature selection, 91
- feature vector, 7
- feed-forward neural network, *see* layered neural network
- finite mixture model (FMM), 24, 153
- Fisher discriminant analysis, 351
- Fisher's exact test, 304
- FP-growth algorithm, 123, 147, 151
- frequent itemset mining, 120, 144, 147
- frequent pair mining, 275
- frequent pattern mining, 120
- frequent subgraph mining, 220, 226, 231
- frequent subsequence mining, 145, 146, 178
- frequent subtree mining, 210, 216
- Frobenius norm, 341, 342
- full rank, 34
  
- G-protein coupled receptor (GPCR), 114
- gene expression, 110, 115, 119, 279
- gene regulatory network, 12, 237
- gene set enrichment analysis (GSEA), 113
- generalization, 42, 48
- generalized Rayleigh quotient, 98
- generalized CCA, 268
- generalized eigenvalue problem, 97, 102, 244, 267, 268, 272, 334, 351
- generalized linear model, 56
- generalized Rayleigh quotient, 23
- generalized sequential patterns (GSP) algorithm, 146, 147
- generative model, 26, 32, 62, 127, 131, 169
- Gibbs sampling, 132, 354
- Gini index, 309
- glycan, 10, 212
- gradient descent, 58, 59, 67, 68, 169
- graph, 6, 9, 10, 12, 96, 109, 219
- graph cut, 242
- graph kernel, 229, 231
- graph Laplacian, 97, 98, 241, 248, 287, 288, 344, 346

- graph learning, 237, 239
- graph partitioning, 238, 240, 260
- graphical model, 33, 52, 127, 131, 156, 192
- greedy algorithm, 29, 31
- gSpan algorithm, 221
- hard margin, 85
- hidden Markov model (HMM), 154
- hidden Markov model (HMM), 156, 184, 186, 188, 193
- hidden tree Markov model, 213
- hidden tree Markov model (HTMM), 193, 206
- hidden variable, *see* latent variable
- hierarchical clustering, 16, 27, 110
- high-throughput data, 5, 115, 237, 279
- Hilbert-Schmidt norm, *see* Frobenius norm
- hyperplane, 84
- information, 45
- information gain, 45
- inner product, 89, 325, 329
- inner product space, 336
- invertible matrix, *see* nonsingular matrix
- Jaccard index (JI), 303
- Karush-Kuhn-Tucker (KKT) conditions, 256, 349
- kernel  $K$ -means clustering, 101, 103, 246, 360
- kernel canonical correlation analysis (KCCA), 269
- kernel function, 89, 100, 103, 109, 269, 337
- kernel learning, 89, 100, 109, 135, 145, 169, 211, 229, 234, 239, 274
- kernel principal component analysis (kernel PCA), 107
- kernel ridge regression, 103, 360
- kernel trick, 100
- Kullback-Leibler (KL) divergence, 331
- label propagation, 250, 258, 260, 281, 287, 290
- Laplacian eigenmaps, 92, 96, 247, 282
- latent Dirichlet allocation (LDA), 127
- latent variable, 4, 24, 127, 131, 132, 135, 153, 154, 169
- layered neural network, 64, 67, 69, 72, 358
- least absolute shrinkage and selection operator (LASSO), 100, 256
- least squares, 347
- likelihood, 58, 128
- linear regression, 54, 99
- linkage, 29
- logistic function, *see* sigmoid function
- logistic regression, 58, 60, 62, 64, 66, 67, 69, 72, 99, 169, 357
- low-rank approximation, *see* matrix factorization
- majority voting, 42
- margin, 84
- market basket, 8, 119, 121, 145
- Markov blanket, 93
- Markov model, 153
- Markov property, 152
- match kernel, 136
- matrix completion, 41
- matrix factorization, 34, 35, 41, 55, 56, 247, 249, 286, 290, 356
- Matthew's correlation coefficient (MCC), 301
- maximal frequent pattern mining, 228, 235
- maximum likelihood estimation, 25, 59, 60, 128, 155, 157, 159, 351
- mean, 18, 60, 324
- mean average precision (MAP), 319
- mean reciprocal rank (MRR), 319
- metabolic network, 179
- method of Lagrange multipliers, 22, 37, 86, 88, 95, 102, 104, 242, 249, 266, 271, 348
- minimum cut, 243
- minimum description length (MDL), 46
- minimum DFS code, 222, 226, 231
- mixture Markov model, 153, 154, 179
- mixture pair model, 153

- molecular graph, 6, 11, 219, 235, 272
- most likely path, 162, 185, 196
- mostlikely path, 204
- multinomial distribution, 131
- multinomial distribution, 130, 327
- multiple kernel learning, 274, 290
- multiple kernel learning (MKL), 264
- multiple sequence alignment (MSA), 180, 183, 184, 186
- multiplicative update, 38, 39
- multivariate analysis, 58
- multiview learning, 13, 262
- mutual information, 330
  
- naive Bayes classifier, 51, 60, 72, 169
- natural language processing, 8, 12, 119, 126, 152, 156, 163
- Needleman-Wunsch algorithm, 182
- network sciences, 237
- node degree, 11
- nonparametric distribution, 138
- nonsingular matrix, 36, 37, 54, 56, 97, 332
- normal distribution, 26, 60
- normalized cut, 244
- normalized discounted cumulative gain (nDCG), 321
- normalized graph Laplacian, 244, 252, 260
- normalized mutual information (NMI), 315
  
- observable variable, 3
- ordered tree Markov model, 213
- ordered tree Markov model (OTMM), 198, 207
- outlier, 5, 21
- overfitting, 42, 55, 72, 99, 129, 130
  
- pair model, 153
- pairwise sequence alignment (PSA), 180, 184
- partitional clustering, 16, 31
- pattern growth, 122, 146
- perceptron, 57
- positive semidefinite matrix, 141
- positive definite matrix, 37
  
- positive semidefinite
  - matrix, 56, 109, 241, 245, 334, 335
- posterior distribution, 25, 131
- precision-recall (PR) curve, 309
- PrefixSpan algorithm, 147, 151
- principal component analysis (PCA), 92, 94, 107, 110, 246, 265, 270, 351
- prior distribution, 25
- prior knowledge, 21, 64, 109, 130, 153, 184, 186, 231, 274
- probabilistic latent semantic analysis (pLSA), 127
- probabilistic model, 23, 33
- probability density function (PDF), 323, 327
- probability mass function (PMF), 323, 326
- profile hidden Markov model (profile HMM), 184, 185
- pruning, 48
- pyramid match kernel, 138
  
- rand index (RI), 313
- random forest, 75
- ratio cut, 244
- Rayleigh quotient, 96
- Rayleigh quotient, 23, 41, 103, 108, 245, 253, 286, 350
- Receiver Operator Characteristic (ROC) curve, 305
- recommendation, 41
- recursive partitioning, 47, 74
- regression, 42
- regular matrix, *see* nonsingular matrix
- regularization, 42, 54
- regularization coefficient, 5, 37
- regularization term, *see* regularizer
- regularizer, 4, 36, 55, 130
- relative risk, 302
- reverse search, 225, 226
- ridge regression, 54, 56, 103, 105
- root mean square deviation (RMSD), 312
  
- sampling with replacement, 73, 75, 323

- sampling without replacement, 323
- scalability, 109
- semi-structured data, 12
- semi-supervised classification, 280
- semi-supervised clustering, 280, 285
- semi-supervised learning, 238, 279
- sequence, 6, 8, 12, 115, 145
- sequence classification, 118
- set, 6, 7, 12
- sigmoid function, 57, 65, 69, 325
- single nucleotide polymorphisms (SNPs), 116
- singular matrix, 332
- singular value decomposition (SVD), 36, 96, 334
- social network, 11
- social network services (SNS), 237, 261, 279
- soft margin, 87
- sparse learning, 37, 92, 98, 109, 239
- spectral clustering, 103, 245, 248, 286, 287, 290, 351
- spectral embedding, *see* Laplacian eigenmaps
- spectrum kernel, 170
- speech recognition, 8, 12, 156, 163
- squared error (loss), *see*  $L^2$  norm
- standard deviation, 324
- state transition diagram, 153, 155, 156, 193
- steepest descent, 347
- step function, 57
- string, 8
- string kernel, 169, 190
- subgraph, 11, 112, 220, 231
- subsequence, 8, 112, 146
- substring, 8
- subtree, 10, 211, 213, 231
- suffix tree, 172
- supervised learning, 4, 26, 42
- support vector machine, 359
- support vector machine (SVM), 84, 100, 103, 291
- topic model, 127
- traceback, 182, 185
- tree, 6, 9, 12, 191
- tree grammar, 166, 187
- tri-factorization, 35
- unsupervised learning, 4, 16, 26
- user-item matrix, 41
- variance, 60, 94, 324, 361
- vector, 5, 7, 15
- visualization, 32
- Viterbi algorithm, 163
- weak learner, 84
- world wide web (WWW), 237
- Z-test, 304

