

Index

Page numbers in bold type indicate a main explanatory entry.

- Acoustic model, 3, 55, **59**
 Acoustic unit discovery, 372
 Agglomerative clustering, 240
 Akaike information criterion, 217
 Approximate inference, 47, 137, **320**, 385
 Automatic relevance determination, 192, 193
- Backoff smoothing, 107, **108**, 208, 380
 Backward algorithm, 66, **73**
 Backward variable, **72**, 85, 90, 159, 266
 Bag of words, 114, **318**, 324
 Basis representation, 192
 Baum–Welch algorithm, 90, 160, 254, 257, 284, 285
 Bayes decision, 8, **54**, 97, 218, 274
 Bayes factor, **214**, 233, 238, 240, 282
 Bayes risk, 55, 56
 Bayesian compressive sensing, 193
 Bayesian information criterion, 211, **214**, 216, 273, 361
 Bayesian interpretation, 189, 379
 Bayesian model comparison, 185
 Bayesian network, 41
 Bayesian neural network, 225
 Bayesian nonparametrics, 337, **345**
 Bayesian predictive classification, 58, **218**, 274, 282
 Belief propagation, 46
 Bernoulli distribution, 131, 226
 Beta distribution, 348, 386, 393, 394
- Canonical form, 17, 24, 27, 34
 Chinese restaurant franchise, 355
 Chinese restaurant process, 349, 380, 383
 Cholesky decomposition, 94, 289
 Co-occurrence, 113
 Collapsed Gibbs sampling, **344**, 363
 Complete data, **61**, 77
 Complete data likelihood, **61**, 65
 Complete square, 30, 34, 152, **391**
 Concave function, 244, 389
 Concentration parameter, 348, 385
- Conditional independence, 38, 40, 119, 132, 244, 341
 Conditional probability, 14
 Conjugate distribution, 17, **24**, 138, 292
 Conjugate prior, **25**, 177, 178, 191, 206, 223, 344
 Context-dependent phoneme, 59
 Continuous density hidden Markov model, **63**, 143, 254, 373
 Convex function, 78, 320
 Corrective training, 177
 Correlation matrix, 114
 Cosine similarity, **114**, 118, 173
 Cross entropy error function, 228
 Cross validation, 189
- De Finetti's theorem, **346**, 348
 Decision tree clustering, **230**, 282
 Decoding algorithm, 51, 204
 Deep neural network, 4, **224**
 Di-gamma function, 209, 262, 265, 321, **389**
 Dirac delta function, 16, 33, 140, 213, 246, **388**
 Dirichlet class language model, 326
 Dirichlet distribution, 38, 144, 190, 256, **394**
 Dirichlet process, 348
 Dirichlet process mixture model, **351**, 373
 Discrete uniform distribution, 102, **392**
 Discriminative approach, 130
 Discriminative training, **166**, 200
 Distribution estimation, **47**, 184, 187, 193, 228, 269, 379
 Document model, 176, 318
 Dynamic Bayesian network, 41
 Dynamic programming, 70, 75
- Eigenvoice, 164
 EM algorithm, 66, **76**, 113, 120, 123, 141, 178, 227, 251
 Emission probability, 62
 Entropy, 116
 Evidence framework, 185, 190, 208
 Evidence function, 17, 184, 185, 188, 196, 207, 208

- Exchangeability, **346**, 357
 Expectation propagation, 7
 Exponential family, 17, 26, 188
 Extended Baum–Welch algorithm, 167
- Factor analysis, 7, 174, 197, 307, 317
 Factor graph, 45
 Factor loading, 197
 Feature extraction, 4, 9, 336
 Feature-space MLLR, 204, 289
 Fisher information, 216
 Forward algorithm, 66, **71**
 Forward variable, **70**, 75, 85, 90, 158, 266
 Forward–backward algorithm, 46, **70**
- Gamma distribution, 30, 38, 256, 308, **400**
 Gamma function, 30, 278, 389
 Gaussian distribution, 18–20, 38, **395**
 Gaussian mixture model, 4, **66**, 67, 172, 361
 Gaussian supervector, **173**, 306
 Gaussian–gamma distribution, 31, 38, **402**
 Gaussian–Wishart distribution, 35, 38, 190, 256, **403**
 GEM distribution, 349
 Generalization error, 191
 Generative approach, 130
 Generative model, **43**, 53
 Gibbs sampling, **343**, 347, 373, 375, 386
 Graphical model, 13, 40, 67, 367
- Heavy-tailed distribution, 383
 Hessian matrix, **187**, 212, 215, 221, 229
 Hidden Markov model, 3, **59**, 68, 188
 Hidden variable, 77, 309, 375
 Hierarchical Dirichlet language model, **205**, 379, 383
 Hierarchical Dirichlet process, 337, **352**
 Hierarchical Pitman–Yor language model, **378**, 384
 Hierarchical Pitman–Yor process, 383
 Hierarchical prior, **193**, 206, 207, 291, 356, 384
 Hyperparameter, **25**, 43, 180, 184, 187, 208, 219, 229
- i*-smoothing factor, 169
 Ill-posed problem, 188, 218
 Implicit solution, 198, **200**, 210
 Importance sampling, 338
 Importance weights, 339
 Incomplete data, 77, 195, 227
 Incremental hyperparameter, 180
 Incremental learning, 177, **179**
 Independently, identically distributed (iid), **216**, 349
 Infinitely exchangeable, 346
 Information retrieval, 3, **10**, 115
 Interpolation smoothing, **102**, 103, 104, 176, 208
 Inverse gamma distribution, **401**
- Jacobian, 338
 Jelinek–Mercer smoothing, 102
- Jensen’s inequality, 76, 244, 290, 320, **389**
 Joint probability, 14
- Katz smoothing, 106
 Kernel parameter, 191
 Kronecker delta function, 16, 85, 246, 324, **388**
 Kullback–Leibler divergence, **79**, 201, 243, 288, 290, 316, 320
- l^0 regularization, 217
 l^1 regularization, 139
 l^2 regularization, 139
 Lagrange multiplier, **86**, 100, 123, 178, 246, 253
 Language model, 3, 55, **97**, 97, 116, 125, 205, 208, 324, 378, 383
 Laplace approximation, 187, 211, 220, 229
 Laplace distribution, 139, **399**
 Latent Dirichlet allocation, 360
 Latent model, 61
 Latent semantic analysis, **113**, 177
 Latent topic, **119**, 319
 Latent topic model, 7, 43, 53, **119**, 123, 138, 176, 242, 318, 337, 360
 Least-squares regression, 188
 Level-1 inference, 188
 Level-2 inference, 188
 Levenshtein distance, 56
 Lexical model, 60
 Linear regression, 188
 Logistic regression, 46
 Logistic sigmoid, 58, 226
 Loopy belief propagation, 47
 Loss function, 55
- Machine learning, 3
 Machine translation, 3, 55, 97
 MAP adaptation, 165, 172
 MAP decision rule, **55**, 58, 129, 130
 MAP–EM algorithm, 143
 Marginal likelihood, **17**, 51, 65, 70, 184, 196, 208, 228, 293, 319, 361, 362, 368
 Marginalization, 50
 Markov chain, 340
 Markov chain Monte Carlo, **337**, 347, 356, 358, 385
 Markov random field, 40, 44
 Matrix variate Gaussian distribution, 292, **399**
 Max sum algorithm, 46
 Maximum a-posteriori, 49, **137**, 177, 228
 Maximum evidence, 186
 Maximum likelihood, **76**, 113, 120, 227
 Maximum likelihood linear regression, **91**, 163, 164, 174, 200, 287
 Maximum mutual information, **167**, 201
 Message passing, 46
 Metropolis–Hastings algorithm, 341
 Minimum Bayes risk, 56
 Minimum classification error, 167

- Minimum description length, 230, 273
 Minimum discrimination information, 127
 Minimum mean square error, 139
 Minimum phone error, 167
 Model selection, 8, **49**, 186
 Model variable, 131
 Modified Kneser–Ney smoothing, **110**, 204, 387
 Multilayer perceptron, **224**, 226, 326
 Multinomial distribution, 22, 62, 99, **392**
 Multivariate Gaussian distribution, 38, **396**

n-gram, 3, **97**, 99, 174
 Naive Bayes classifier, **39**, 41
 Natural parameter, 17
 Nested Chinese restaurant process, 360
 Nested Dirichlet process, 360
 Neural network, 188, **224**, 326
 Noisy channel model, 8, **55**
 Noisy speech recognition, 188, 224
 Non-negative matrix factorization, 124

 Occam factor, 186
 Over-fitting problem, **5**, 188

 Pólya urn model, **346**, 348
 Partition function, 45
 Phoneme, 59
 Pitman–Yor process, 379
 Plug-in MAP, 140
 Point estimation, **47**, 184, 187, 218
 Positive definite, 197, 396
 Posterior, 6, **15**, 133, 185
 Power-law, **110**, 379, 380
 Precision matrix, 34
 Predictive distribution, **219**, 275
 Prior, **15**, 48, 138, 185
 Probabilistic latent semantic analysis, **119**, 120, 176, 318
 Product rule, **14**, 132
 Proposal distribution, 339

 Quasi-Bayes, 179

 Regression tree, **92**, 287
 Regularization, 6, **49**, 138, 141, 167, 188
 Regularization parameter, 139, **188**, 191
 Regularization theory, 188
 Regularized least-squares, 188
 Relevance vector machine, 192
 Reproducible distribution pair, 181
 Robust decision rule, 58, 218
 Robustness, 50, 203

 Segmental K-means, 90
 Singular value decomposition, 114, 177
 Slice sampling, 344

 Sparse Bayesian learning, 184, **194**
 Sparse prior distribution, **194**
 Speaker adaptation, 91, **163**, 200
 Speaker adaptive training, 91, 97, 163
 Speaker clustering, 10, **239**, 360
 Speaker dependent, 163
 Speaker diarization, 240
 Speaker independent, 163
 Speaker segmentation, 10
 Speaker verification, 10, **171**, 306
 Speech recognition, 3, 7, 8, **9**, 51, 53, 54, 59, 91, 97, 128, 180, 185, 191, 225, 254
 Spherical Gaussian distribution, 292, **398**
 Stick-breaking construction, **348**, 353
 Stirling’s approximation, 285, **389**
 Student’s *t*-distribution, 193, 281, **404**
 Sufficient statistics, **17**, 89, 95, 159, 180, 259, 363, 368, 377
 Sum product algorithm, 46
 Sum rule, 14, 132
 Sum-of-squares error function, 189
 Support vector machine, 173, 188

 tf-idf, 113
 Tied-state HMM, 230
 Triphone, 230
 Type-2 maximum likelihood, **187**, 319, 328

 Uncertainty, 5
 Uncertainty techniques, 51
 Undirected graph, 44
 Unigram model, 99
 Unigram rescaling, **127**, 325
 Universal background model, **172**, 306, 367

 Variational Bayes, **242**, 243, 318, 329
 Variational Bayes (VB) auxiliary function, 258
 Variational Bayes (VB) Viterbi algorithm, 269
 Variational Bayes Baum–Welch algorithm, 269
 Variational Bayes expectation and maximization (VB–EM) algorithm, 252
 Variational distribution, 320
 Variational lower bound, 242, **243**, 269, 285, 288–290, 316
 Variational method, 242, **246**, 252
 Variational parameter, 320
 Vector quantization, 62
 Vector space model, 10, **114**
 Viterbi algorithm, 46, 66, **74**, 275
 Viterbi training, 90

 Weak-sense auxiliary function, 201
 Weighted finite state transducer, 60
 Wishart distribution, 38, 145, **403**
 Woodbury matrix inversion, 196, 324, **391**
 Word-document matrix, 113