

# 欧文索引

## ■ 数字・記号

0-1 loss, [75](#), [198](#)

## ■ A

Absolute value rectification, [138](#)

Accuracy, [306](#)

Activation function, [123](#)

Active constraint, [69](#)

AdaGrad, [220](#)

ADALINE, → adaptive linear element

Adam, [222](#), [308](#)

Adaptive linear element, [11](#), [17](#), [18](#)

Adversarial example, [192](#)

Adversarial training, [192](#), [195](#), [388](#)

Affine, [79](#)

AIS, → annealed importance sampling

Almost everywhere, [52](#)

Almost sure convergence, [95](#)

Ancestral sampling, [423](#), [435](#)

ANN, → Artificial neural network

Annealed importance sampling, [458](#), [491](#), [527](#)

Approximate Bayesian computation, [527](#)

Approximate inference, [425](#)

Artificial intelligence, [1](#)

Artificial neural network, → Neural network

ASR, → automatic speech recognition

Asymptotically unbiased, [90](#)

Audio, [74](#), [259](#), [332](#)

Autoencoder, [3](#), [255](#), [367](#)

Automatic speech recognition, [332](#)

## ■ B

Back-propagation, [146](#)

Back-propagation through time, [277](#)

Backprop, → back-propagation

Bag of words, [342](#)

Bagging, [183](#)

Batch normalization, [192](#), [308](#)

Bayes error, [84](#)

Bayes' rule, [52](#)

Bayesian hyperparameter optimization, [316](#)

Bayesian network, → directed graphical model

Bayesian probability, [40](#)

Bayesian statistics, [98](#)

Belief network, → directed graphical model

Bernoulli distribution, [45](#)

BFGS, [227](#)

Bias, [90](#), [163](#)

bias parameter, [79](#)

Biased importance sampling, [434](#)

Bigram, [335](#)

Binary relation, [350](#)

Block Gibbs sampling, [438](#)

Boltzmann distribution, [416](#)

Boltzmann machine, [416](#), [481](#)

BPTT, → back-propagation through time

Broadcasting, [25](#)

Burn-in, [437](#)

## ■ C

CAE, → contractive autoencoder

Calculus of variations, [129](#)

Categorical distribution, → multinoulli distribution

CD, → contrastive divergence

Centering trick (DBM), [495](#)

Central limit theorem, [47](#)

Chain rule (calculus), [148](#)

Chain rule of probability, [43](#)

Chess, [1](#)

Chord, [420](#)

Chordal graph, [421](#)

Class-based language models, [336](#)

Classical dynamical system, [270](#)

Classification, [72](#)

Clique potential, → factor (graphical model)

CNN, → convolutional neural network

Collaborative Filtering, [348](#)

Collider, → explaining away

Color images, [259](#)

Complex cell, 262  
 Computational graph, 146  
 Computer vision, 328  
 Concept drift, 393  
 Condition number, 201  
 Conditional computation, → dynamic structure  
 Conditional independence, xvi, 44  
 Conditional probability, 43  
 Conditional RBM, 503  
 Connectionism, 13, 321  
 Connectionist temporal classification, 334  
 Consistency, 94, 374  
 Constrained optimization, 67, 169  
 Content-based addressing, 301  
 Content-based recommender systems, 349  
 Context-specific independence, 419  
 Contextual bandits, 349  
 Continuation methods, 235  
 Contractive autoencoder, 380  
 Contrast, 329  
 Contrastive divergence, 208, 446, 494  
 Convex optimization, 102  
 Convolution, 237, 502  
 Convolutional network, 12  
 Convolutional neural network, 181, 237, 308, 334  
 Coordinate descent, 231, 492  
 Correlation, 45  
 Cost function, → objective function  
 Covariance, xvi, 45  
 Covariance matrix, 45  
 Coverage, 307  
 Critical temperature, 441  
 Cross-correlation, 239  
 Cross-entropy, 55, 96  
 Cross-validation, 88  
 CTC, → connectionist temporal classification  
 Curriculum learning, 236  
 Curse of dimensionality, 112  
 Cyc, 2

## ■ D

D-separation, 418  
 DAE, → denoising autoencoder  
 Data generating distribution, 80, 95  
 Data generating process, 80  
 Data parallelism, 324  
 Dataset, 75  
 Dataset augmentation, 194, 332  
 DBM, → deep Boltzmann machine

DCGAN, 403, 516  
 Decision tree, 104, 400  
 Decoder, 3  
 Deep belief network, 18, 388, 463, 483, 485, 503, 509  
 Deep Blue, 1  
 Deep Boltzmann machine, 17, 18, 388, 463, 478, 483, 487, 494, 503  
 Deep feedforward network, 121, 308  
 Deep learning, 1, 4  
 Degree of belief, 40  
 Denoising autoencoder, 373, 506  
 Denoising score matching, 454  
 Density estimation, 74  
 Derivative, xvi, 61  
 Design matrix, 77  
 Detector layer, 244  
 Determinant, xv  
 Diagonal matrix, 30  
 Differential entropy, 55, 474  
 Dirac delta function, 48  
 Directed graphical model, 57, 370, 411, 509  
 Directional derivative, 62  
 Discriminative fine-tuning, → supervised fine-tuning  
 Discriminative RBM, 504  
 Distributed representation, 13, 109, 399  
 Domain adaptation, 392  
 Dot product, 25, 102  
 Double backprop, 195  
 Doubly block circulant matrix, 239  
 Dream sleep, 446, 478  
 DropConnect, 191  
 Dropout, 185, 308, 312, 494, 506  
 Dynamic structure, 325

## ■ E

E-step, 465  
 Early stopping, 176–179, 308  
 EBM, → energy-based model  
 Echo state network, 17, 18, 291  
 Effective capacity, 82  
 Eigendecomposition, 31  
 Eigenvalue, 31  
 Eigenvector, 31  
 ELBO, → evidence lower bound  
 Element-wise product, → Hadamard product  
 EM, → expectation maximization  
 Embedding, 378

Empirical distribution, 48  
 Empirical risk, 198  
 Empirical risk minimization, 198  
 Encoder, 3  
 Energy function, 415  
 Energy-based model, 415, 435, 481, 487  
 Ensemble methods, 183  
 Epoch, 176  
 Equality constraint, 68  
 Equivariance, 242  
 Error function, → objective function  
 ESN, → echo state network  
 Euclidean norm, 29  
 Euler-Lagrange equation, 474  
 Evidence lower bound, 463, 486  
 Example, 72  
 Expectation, 44  
 Expectation maximization, 465  
 Expected value, → expectation  
 Explaining away, 418, 463, 472  
 Exploitation, 349  
 Exploration, 349  
 Exponential distribution, 48

## ■ F

F-score, 307  
 Factor (graphical model), 413  
 Factor analysis, 357  
 Factor graph, 422  
 factor of variation, 3  
 Feature, 72  
 Feature selection, 169  
 Feedforward neural network, 121  
 Fine-tuning, 232  
 Finite differences, 318  
 Forget gate, 219  
 Forward propagation, 146  
 Fourier transform, 259  
 Fovea, 263  
 FPCD, 450  
 Free energy, 417, 499  
**Freebase**, 351  
 Frequentist probability, 40  
 Frequentist statistics, 98  
 Frobenius norm, 34  
 Fully-visible Bayes network, 519  
 Functional derivatives, 473  
 FVBN, → fully-visible Bayes network

## ■ G

Gabor function, 264  
 GANs, → generative adversarial networks  
 Gated recurrent unit, 308  
 Gaussian distribution, → normal distribution  
 Gaussian kernel, 103  
 Gaussian mixture, 49, 135  
 GCN, → Global contrast normalization  
**GeneOntology**, 351  
 Generalization, 80  
 Generalized Lagrange function, → generalized Lagrangian  
 Generalized Lagrangian, 68  
 Generative adversarial networks, 506, 515  
 Generative moment matching networks, 517  
 Generator network, 510  
 Gibbs distribution, 414  
 Gibbs sampling, 423, 438  
 Global contrast normalization, 330  
 GPU, → graphics processing unit  
 Gradient, 62  
 Gradient clipping, 207, 298  
 Gradient descent, 61, 63  
 Graph, xv  
 Graphical model, → structured probabilistic model  
 Graphics processing unit, 322  
 Greedy algorithm, 232  
 Greedy layer-wise unsupervised pretraining, 386  
 Greedy supervised pretraining, 232  
 Grid search, 313

## ■ H

Hadamard product, xv, 25  
 Hard tanh, 141  
 Harmonium, → restricted Boltzmann machine  
 Harmony theory, 416  
 Helmholtz free energy, → evidence lower bound  
 Hessian, 160  
 Hessian matrix, xvi, 64  
 Heteroscedastic, 135  
 Hidden layer, 5, 122  
 Hill climbing, 63  
 Hyperparameter optimization, 313  
 Hyperparameters, 87, 312  
 Hypothesis space, 81, 85

## ■ I

i.i.d. assumptions, 80, 89, 192  
 Identity matrix, 26

ILSVRC, → ImageNet Large Scale Visual Recognition Challenge  
 ImageNet Large Scale Visual Recognition Challenge, 19  
 Immorality, 420  
 Importance sampling, 433, 457, 513  
 Importance weighted autoencoder, 513  
 Independence, xvi, 44  
 Independent and identically distributed, → i.i.d. assumptions  
 Independent component analysis, 359  
 Independent subspace analysis, 360  
 Inequality constraint, 68  
 Inference, 410, 425, 463, 466, 468, 475, 477  
 Information retrieval, 383  
 Initialization, 215  
 Integral, xvi  
 Invariance, 244  
 Isotropic, 48

## ■ J

Jacobian matrix, xvi, 53, 63  
 Joint probability, 41

## ■ K

$k$ -means, 261, 400  
 $k$ -nearest neighbors, 103, 400  
 Karush-Kuhn-Tucker, 68  
 Karush-Kuhn-Tucker conditions, 69, 169  
 Kernel machine, 401  
 Kernel trick, 102  
 Kernel (convolution) , 238  
 KKT, → Karush-Kuhn-Tucker  
 KKT conditions, → Karush-Kuhn-Tucker conditions  
 KL divergence, → Kullback-Leibler divergence  
 Knowledge base, 2, 351  
 Krylov methods, 160  
 Kullback-Leibler divergence, xvi, 55

## ■ L

Label smoothing, 173  
 Lagrange multipliers, 68, 474  
 Lagrangian, → generalized Lagrangian  
 LAPGAN, 516  
 Laplace distribution, 48, 362  
 Latent variable, 49  
 Layer (neural network), 121  
 LCN, → local contrast normalization

Leaky ReLU, 138  
 Leaky units, 293  
 Learning rate, 63  
 Line search, 63, 68  
 Linear combination, 28  
 Linear dependence, 28  
 Linear factor models, 357  
 Linear regression, 77, 79, 101  
 Link prediction, 352  
 Lipschitz constant, 67  
 Lipschitz continuous, 67  
 Liquid state machine, 291  
 Local conditional probability distribution, 411  
 Local contrast normalization, 331  
 Logistic regression, 2, 102  
 Logistic sigmoid, 6, 49  
 Long short-term memory, 13, 19, 219, 295, 308  
 Loop, 420  
 Loopy belief propagation, 427  
 Loss function, → objective function  
 $L^p$  norm, 29  
 LSTM, → long short-term memory

## ■ M

M-step, 465  
 Machine learning, 2  
 Machine translation, 73  
 Main diagonal, 24  
 Manifold, 115  
 Manifold hypothesis, 116  
 Manifold learning, 115  
 Manifold tangent classifier, 195  
 MAP approximation, 100, 369  
 Marginal probability, 43  
 Markov chain, 435  
 Markov chain Monte Carlo, 435  
 Markov network, → undirected model  
 Markov random field, → undirected model  
 Matrix, xiv, 24  
 Matrix inverse, 27  
 Matrix product, 25  
 Max norm, 29  
 Max pooling, 244  
 Maximum likelihood, 95  
 Maxout, 138, 308  
 MCMC, → Markov chain Monte Carlo  
 Mean field, 468, 469, 494  
 Mean squared error, 78  
 Measure theory, 52

Measure zero, 52  
 Memory network, 300, 301  
 Method of steepest descent, → gradient descent  
 Minibatch, 200  
 Missing inputs, 72  
 Mixing (Markov chain), 439  
 Mixture density networks, 135  
 Mixture distribution, 49  
 Mixture model, 135, 372  
 Mixture of experts, 326, 400  
 MLP, → multilayer perception  
 MNIST, 15, 16, 494  
 Model averaging, 183  
 Model compression, 324  
 Model identifiability, 204  
 Model parallelism, 324  
 Moment matching, 517  
 Moore-Penrose pseudoinverse, 33, 171  
 Moralized graph, 420  
 MP-DBM, → multi-prediction DBM  
 MRF (Markov Random Field), → undirected model  
 MSE, → mean squared error  
 Multi-modal learning, 394  
 Multi-prediction DBM, 495  
 Multi-task learning, 174, 393  
 Multilayer perception, 4  
 Multilayer perceptron, 18  
 Multinomial distribution, 46  
 Multinoulli distribution, 46

## ■ N

*n*-gram, 335  
 NADE, 521  
 Naive Bayes, 2  
 Nat, 54  
 Natural image, 408  
 Natural language processing, 334  
 Nearest neighbor regression, 83  
 Negative definite, 65  
 Negative phase, 341, 443, 445  
 Neocognitron, 12, 17, 18, 263  
 Nesterov momentum, 215  
 Netflix Grand Prize, 184, 348  
 Neural language model, 336, 346  
 Neural network, 10  
 Neural Turing machine, 301  
 Neuroscience, 12  
 Newton's method, 66, 223

NLM, → neural language model  
 NLP, → natural language processing  
 No free lunch theorem, 85  
 Noise-contrastive estimation, 454  
 Non-parametric model, 83  
 Norm, xvii, 29  
 Normal distribution, 47, 91  
 Normal equations, 79, 81, 166  
 Normalized initialization, 217  
 Numerical differentiation, → finite differences

## ■ O

Object detection, 328  
 Object recognition, 328  
 Objective function, 61  
 OMP-*k*, → orthogonal matching pursuit  
 One-shot learning, 393  
 Operation, 146  
 Optimization, 59, 60  
 Orthodox statistics, → frequentist statistics  
 Orthogonal matching pursuit, 18, 183  
 Orthogonal matrix, 31  
 Orthogonality, 30  
 Output layer, 121

## ■ P

Parallel distributed processing, 13  
 Parameter initialization, 215, 293  
 Parameter sharing, 181, 241, 269, 270, 280  
 Parameter tying, → Parameter sharing  
 Parametric model, 83  
 Parametric ReLU, 138  
 Partial derivative, 62  
 Partition function, 414, 443, 492  
 PCA, → principal components analysis  
 PCD, → stochastic maximum likelihood  
 Perceptron, 11, 18  
 Persistent contrastive divergence, → stochastic maximum likelihood  
 Perturbation analysis, → reparametrization trick  
 Point estimator, 89  
 Policy, 349  
 Pooling, 237, 502  
 Positive definite, 65  
 Positive phase, 341, 443, 445, 482, 491  
 Precision, 307  
 Precision (of a normal distribution), 46, 47  
 Predictive sparse decomposition, 382  
 Preprocessing, 328

Pretraining, 232, 386  
 Primary visual cortex, 261  
 Principal components analysis, 35, 106, 107, 357, 463  
 Prior probability distribution, 98  
 Probabilistic PCA, 357, 358, 464  
 Probability density function, 42  
 Probability distribution, 41  
 Probability mass function, 41  
 Probability mass function estimation, 74  
 Product of experts, 416  
 Product rule of probability, → chain rule of probability  
 PSD, → predictive sparse decomposition  
 Pseudolikelihood, 450

## ■ Q

Quadrature pair, 265  
 Quasi-Newton methods, 227

## ■ R

Radial basis function, 141  
 Random search, 315  
 Random variable, 41  
 Ratio matching, 453  
 RBF, 141  
 RBM, → restricted Boltzmann machine  
 Recall, 307  
 Receptive field, 242  
 Recommender Systems, 347  
 Rectified linear unit, 125, 138, 308, 370  
 Recurrent network, 18  
 Recurrent neural network, 272  
 Regression, 73  
 Regularization, 87, 128, 163, 312  
 Regularizer, 86  
 REINFORCE, 507  
 Reinforcement learning, 19, 76, 349, 506  
 Relational database, 351  
 Relations, 350  
 ReLU, → Rectified linear unit  
 Reparametrization trick, 506  
 Representation learning, 2  
 Representational capacity, 81  
 Restricted Boltzmann machine, 255, 333, 348, 427, 463, 482, 483, 494, 497, 499, 500, 502  
 Ridge regression, → weight decay  
 Risk, 198  
 RNN-RBM, 504

## ■ S

Saddle points, 205  
 Sample mean, 91  
 Scalar, xiv, 23  
 Score matching, 374, 452  
 Second derivative, 63  
 Second derivative test, 65  
 Self-information, 54  
 Semantic hashing, 383  
 Semi-supervised learning, 174  
 Separable convolution, 259  
 Separation (probabilistic modeling), 417  
 Set, xv  
 SGD, → Stochastic gradient descent  
 Shannon entropy, xvi, 54  
 Shortlist, 338  
 Sigmoid, xvii, → logistic sigmoid  
 Sigmoid belief network, 18  
 Simple cell, 262  
 Singular value, → singular value decomposition  
 Singular value decomposition, 33, 107, 348  
 Singular vector, → singular value decomposition  
 Slow feature analysis, 360  
 SML, → stochastic maximum likelihood  
 Softmax, 132, 301, 326  
 Softplus, xvii, 50, 141  
 Spam detection, 2  
 Sparse coding, 231, 255, 362, 463, 509  
 Sparse initialization, 218, 293  
 Sparse representation, 106, 162, 182, 369, 406  
 Spearmint, 316  
 Spectral radius, 291  
 Speech recognition, → automatic speech recognition  
 Sphering, → whitening  
 Spike and slab restricted Boltzmann machine, 500  
 SPN, → sum-product network  
 Square matrix, 28  
 ssRBM, → spike and slab restricted Boltzmann machine  
 Standard deviation, 45  
 Standard error, 92  
 Standard error of the mean, 93, 200  
 Statistic, 89  
 Statistical learning theory, 80  
 Steepest descent, → gradient descent  
 Stochastic back-propagation, → reparametrization trick

Stochastic gradient descent, 11, 109, 200, **211**, 494  
 Stochastic maximum likelihood, 448, 494  
 Stochastic pooling, 191  
 Structure learning, 424  
 Structured output, 73, 503  
 Structured probabilistic model, 56, 407  
 Sum rule of probability, 43  
 Sum-product network, 404  
 Supervised fine-tuning, 387, 486  
 Supervised learning, **76**  
 Support vector machine, 102  
 Surrogate loss function, 198  
 SVD, → singular value decomposition  
 Symmetric matrix, 30, 32

## ■ T

Tangent distance, 194  
 Tangent plane, 377  
 Tangent prop, 194  
 TDNN, → time-delay neural network  
 Teacher forcing, 276  
 Tempering, 441  
 Template matching, 102  
 Tensor, xiv, 24  
 Test set, 80  
 Tikhonov regularization, → weight decay  
 Tiled convolution, 252  
 Time-delay neural network, 263, 269  
 Toeplitz matrix, 239  
 Topographic ICA, 360  
 Trace operator, 34  
 Training error, 80  
 Transcription, 73  
 Transfer learning, 392  
 Transpose, xv, 24  
 Triangle inequality, 29  
 Triangulated graph, → chordal graph  
 Trigram, 335

## ■ U

Unbiased, 90  
 Undirected graphical model, 57, 370  
 Undirected model, 412  
 Uniform distribution, 42  
 Unigram, 335  
 Unit norm, 30  
 Unit vector, 30  
 Universal approximation theorem, 142  
 Universal approximator, 404

Unnormalized probability distribution, 413  
 Unsupervised learning, **76**, 104  
 Unsupervised pretraining, 333, 386

## ■ V

V-structure, → explaining away  
 V1, 261  
 VAE, → variational autoencoder  
 Vapnik-Chervonenkis dimension, 82  
 Variance, xvi, 44, 163  
 Variational autoencoder, 506, **512**  
 Variational derivatives, → functional derivatives  
 Variational free energy, → evidence lower bound  
 VC dimension, → Vapnik-Chervonenkis dimension  
 Vector, xiv, 23  
 Virtual adversarial examples, 193  
 Visible layer, 5  
 Volumetric data, 259

## ■ W

Wake-sleep, 477, 486  
 Weight decay, **86**, 128, 165, 312  
 Weight space symmetry, 204  
 Weights, 11, 78  
 Whitening, 330  
 Wikibase, 351  
 Word embedding, 337  
 Word-sense disambiguation, 352  
 WordNet, 351

## ■ Z

Zero-data learning, → zero-shot learning  
 Zero-shot learning, 393

# 日本語索引

## ■ 数字・記号

0/1 損失, [75](#), [198](#)

1 次視覚皮質, [261](#)

## ■ A

ADALINE, → 適応の線形要素

AIS, → 焼きなまし重点サンプリング

ANN, → 人工ニューラルネットワーク

ASR, → 自動音声認識

## ■ B

Backprop, → 誤差逆伝播法

BPTT, → 通時的誤差逆伝播法

## ■ C

CAE, → 縮小自己符号化器

CD, → コントラストティブ・ダイバージェンス

CNN, → 畳み込みニューラルネットワーク

## ■ D

DAE, → 雑音除去自己符号化器

DBM, → 深層ボルツマンマシン

d 分離, [418](#)

## ■ E

EBM, → エネルギーベースモデル

ELBO, → エビデンス下界

EM, → 期待値最大化

ESN, → エコーステートネットワーク

## ■ F

F 値, [307](#)

FVBN, → すべての変数が可視変数のベ이지アンネットワーク

## ■ G

GANs, → 敵対的生成ネットワーク

GCN, → 大域コントラスト正規化

GPU, → グラフィックスプロセッシングユニット

## ■ I

i.i.d. 仮定, [80](#), [89](#), [192](#)

ILSVRC, → イメージネット大規模視覚認識チャレンジ

## ■ K

$k$  近傍法, [103](#), [400](#)

$k$  平均法, [261](#), [400](#)

Karush-Kuhn-Tucker 条件, [69](#), [169](#)

KKT, → カルーシュ・クーン・タッカー

KKT 条件, → Karush-Kuhn-Tucker 条件

KL ダイバージェンス, → カルバック・ライブラーダイバージェンス

## ■ L

LCN, → 局所コントラスト正規化

Leaky ユニット, [293](#)

$L^p$  ノルム, [29](#)

LSTM, → 長期短期記憶

## ■ M

MAP 推定, [100](#), [369](#)

MCMC, → マルコフ連鎖モンテカルロ法

MLP, → 多層パーセプトロン

MP-DBM, → 多予測 DBM

MRF (マルコフ確率場), → 無向モデル

MSE, → 平均二乗誤差

## ■ N

NLM, → ニューラル言語モデル

NLP, → 自然言語処理

## ■ O

OMP- $k$ , → 直交マッチング追跡

## ■ P

PCA, → 主成分分析

PCD, → 確率的最尤法

PDF, → 確率密度関数

PMF, → 確率質量関数

PreLU, → パラメトリック ReLU

PSD, → 予測スパース分解



## ■ R

RBF, → 動径基底関数

RBM, → 制限付きボルツマンマシン

ReLU, → 正規化線形関数

## ■ S

SGD, → 確率的勾配降下法

SML, → 確率的最尤法

SPN, → 積和ネットワーク

ssRBM, → スパイクスラブ型制限付きボルツマンマシン

SVD, → 特異値分解

## ■ T

TDNN, → 時間遅れニューラルネットワーク

## ■ V

VAE, → 変分自己符号化器

Vapnik-Chervonenkis 次元, 82

VC 次元, → Vapnik-Chervonenkis 次元

V 構造, → 弁明

## ■ あ

アダマール積, xv, 25

アフィン, 79

アンサンブル手法, 183

鞍点, 205

## ■ い

一様分布, 42

一様性, 94, 374

一般化ラグランジアン, 68

一般化ラグランジュ関数, → 一般化ラグランジアン

イメージネット大規模視覚認識チャレンジ, 19

因子 (グラフィカルモデル), 413

因子グラフ, 422

因子分析, 357

## ■ う

埋め込み, 378

## ■ え

エキスパートの積, 416

エコーステートネットワーク, 17, 18, 291

エネルギー関数, 415

エネルギーベースモデル, 415, 435, 481, 487

エビデンス下界, 463, 486

エポック, 176

演算, 146

## ■ お

オイラー・ラグランジュ方程式, 474

重み, 11, 78

重み空間の対称性, 204

重み減衰, 86, 128, 165, 312

音声, 74, 259, 332

音声認識, → 自動音声認識

温度遷移, 441

## ■ か

回帰, 73

回帰結合型ニューラルネットワーク, 272

回帰結合型ネットワーク, 18

概収束, 94

ガウスカーネル, 103

ガウス分布, → 正規分布

学習率, 63

確率質量関数, 41

確率質量関数推定, 74

確率的 PCA, 357, 358, 464

確率的勾配降下法, 11, 109, 200, 211, 494

確率的誤差逆伝播法, → 再パラメータ化トリック

確率的最大プーリング, 502

確率的最尤法, 448, 494

確率的プーリング, 191

確率の加法定理, 43

確率の乗法定理, → 確率の連鎖律

確率の連鎖律, 44

確率分布, 41

確率変数, 41

確率密度関数, 42

隠れ層, 5, 122

可視層, 5

仮説空間, 81, 85

仮想的敵対事例, 193

活性化関数, 123

活性制約, 69

活用, 349

カテゴリ分布, → マルチヌーイ分布

カーネル (畳み込み), 238

カーネルトリック, 102

カーネルマシン, 401

可分畳み込み, 259

ガボール関数, 264

カラー画像, 259

カリキュラム学習, 236

カルーシュ・クーン・タッカー, 68

カルバック・ライブラーダイバージェンス, xvi, 55

関係, 350

関係データベース, 351

## ■ き

機械学習, 2  
 機械翻訳, 73  
 擬似尤度, 450  
 期待値, 44  
 期待値最大化, 465  
 期待値ステップ, 465  
 ギブスサンプリング, 423, 438  
 ギブス分布, 414  
 逆行列, 27  
 球状化, → 白色化  
 強化学習, 20, 76, 349, 506  
 教師あり学習アルゴリズム, 76  
 教師あり再学習, 387, 486  
 教師強制, 276  
 教師なし学習, 104  
 教師なし学習アルゴリズム, 76  
 教師なし事前学習, 333, 386  
 協調フィルタリング, 348  
 共分散, xvi, 45  
 共分散行列, 45  
 行列, xiv, 24  
 行列式, xv  
 行列の積, 25  
 局所コントラスト正規化, 331  
 局所条件付き確率分布, 411  
 近似推論, 425  
 近似ベイズ計算, 527

## ■ く

クラスベース言語モデル, 336  
 グラフ, xv  
 グラフィカルモデル, → 構造化確率モデル  
 グラフィックスプロセッシングユニット, 322  
 クリークポテンシャル, → 因子 (グラフィカルモデル)  
 グリッドサーチ, 313  
 クリロフ法, 160  
 訓練誤差, 80

## ■ け

計画行列, 77  
 経験損失, 198  
 経験損失最小化, 198  
 経験分布, 48  
 計算グラフ, 146  
 継続法, 235  
 欠損値のある入力, 72

決定木, 104, 400  
 ゲートありの回帰ユニット, 308  
 弦, 420  
 弦グラフ, 421  
 検出器の段階, 244

## ■ こ

交差エントロピー, 55, 96  
 交差検証, 88  
 構造化確率モデル, 56, 407  
 構造学習, 424  
 構造出力, 73, 503  
 勾配, 62  
 勾配クリッピング, 207, 298  
 勾配降下法, 61, 63  
 語義曖昧性解消, 352  
 誤差関数, 61  
 誤差逆伝播法, 146  
 コスト関数, 61  
 古典的な動的システム, 270  
 コネクションニズム, 13, 321  
 固有値, 31  
 固有値分解, 31  
 固有ベクトル, 31  
 混合 (マルコフ連鎖), 439  
 混合エキスパート, 326, 400  
 混合ガウス, 49, 135  
 混合分布, 49  
 混合密度ネットワーク, 135  
 混合モデル, 135, 372  
 コントラスティブ・ダイバージェンス, 208, 446, 494  
 コンセプトドリフト, 393  
 コンテンツベースアドレッシング, 301  
 コンテンツベース推薦システム, 349  
 コントラスト, 329  
 コンピュータビジョン, 328

## ■ さ

再学習, 232  
 最急降下法, → 勾配降下法  
 最近傍回帰, 83  
 再現率, 307  
 最大化ステップ, 465  
 最大値ノルム, 29  
 最大プーリング, 244  
 最適化, 59, 60  
 再パラメータ化トリック, 506  
 最尤法, 95  
 雑音除去自己符号化器, 373, 506

雑音除去スコアマッチング, 454

雑音対照推定, 454

座標降下法, 230, 492

サポートベクトルマシン, 102

三角化グラフ, → 弦グラフ

三角不等式, 29

サンプル平均, 91

## ■ し

時間遅れニューラルネットワーク, 263, 269

識別的 RBM, 504

識別的な再学習, → 教師あり再学習

シグモイド, xvii, → ロジスティックシグモイド

シグモイド信念ネットワーク, 18

次元の呪い, 112

自己情報量, 54

自己符号化器, 3, 255, 367

指数分布, 48

事前学習, 232, 386

事前確率分布, 98

自然画像, 408

自然言語処理, 334

持続的コントラストティブ・ダイバージェンス, → 確率的最尤法

自動音声認識, 332

シャノンエントロピー, xvi, 54

自由エネルギー, 417, 499

集合, xv

重点サンプリング, 433, 457, 513

周辺確率分布, 43

重要度重み付き自己符号化器, 513

縮小自己符号化器, 380

主成分分析, 35, 106, 107, 357, 463

主対角線, 24

出力層, 121

受容野, 242

順伝播, 146

順伝播型ニューラルネットワーク, 121

準ニュートン法, 227

条件数, 201

条件付き RBM, 503

条件付き確率, 43

条件付き計算, → 動的構造

条件付き独立, xvi, 44

衝突, → 弁明

情報検索, 383

初期化, 215

ショートリスト, 338

事例, 72

神経科学, 12

人工知能, 1

人工ニューラルネットワーク, → ニューラルネットワーク

深層学習, 1, 4

深層順伝播型ネットワーク, 121, 308

深層信念ネットワーク, 18, 388, 463, 483, 485, 503, 509

深層ボルツマンマシン, 17, 18, 388, 463, 478, 483, 487, 494, 503

信念ネットワーク, → 有向グラフィカルモデル

信念の度合い, 40

## ■ す

推薦システム, 347

推定, 466, 468

推論, 410, 425, 463, 475, 477

数値微分, → 有限差分法

スカラー, xiv, 23

スコアマッチング, 374, 452

素性選択, 169

スパイクスラブ型制限付きボルツマンマシン, 500

スパースコーディング, 231, 255, 509

スパース初期化, 218, 293

スパース表現, 106, 162, 182, 369, 406

スパース符号化, 362, 463

スパム検出, 2

スペクトル半径, 291

すべての変数が可視変数のベイジアンネットワーク, 519

## ■ せ

正規化された初期化, 217

正規化線形関数, 125, 138, 308, 370

正規分布, 46, 47, 91

正規方程式, 79, 81, 166

制限付きボルツマンマシン, 255, 333, 348, 427, 463, 482, 483, 494, 497, 499, 500, 502

正項, 341

生成器ネットワーク, 510

正則化, 87, 128, 163, 312

正則化項, 86

正段階, 443, 445, 482, 491

正定値, 65

精度, 306

精度（正規分布の）, 46, 47

正方行列, 28

制約付き最適化, 67, 169

積分, xvi

積和ネットワーク, 404

接距離, 194

接線伝播, 194

摂動解析, → 再パラメータ化トリック

接平面, 377

セマンティックハッシング, 383

ゼロショット学習, 393

ゼロデータ学習, → ゼロショット学習

漸近不偏, 90

線形因子モデル, 357

線形回帰, 77, 79, 101

線形結合, 28

線形従属, 28

潜在変数, 49

## ■ そ

層 (ニューラルネットワーク), 121

相関, 45

早期終了, 176–179, 308

相互相関, 239

層ごとの貪欲教師なし事前学習, 386

測度零, 52

測度論, 52

ソフトプラス, xvii, 50, 141

ソフトマックス, 132, 301, 326

損失, 198

損失関数, 61

## ■ た

大域コントラスト正規化, 330

対角行列, 30

対称行列, 30, 32

体積データ, 259

代理損失関数, 198

タイル型畳み込み, 252

多項分布, 46

多層パーセプトロン, 4, 18

畳み込み, 237, 502

畳み込みニューラルネットワーク, 12, 181, 237, 308, 334

多様体, 115

多様体学習, 115

多様体仮説, 116

多様体接分類器, 195

多予測 DBM, 495

単位行列, 26

単位ノルム, 30

単位ベクトル, 30

単語埋め込み, 337

探索, 349

単純細胞, 262

## ■ ち

チェス, 1

知識ベース, 2, 351

中心窩, 263

中心化トリック (DBM), 495

中心極限定理, 47

長期短期記憶, 13, 19, 219, 295

直線探索, 63, 68

直角位相ベア, 265

直交, 30

直交行列, 31

直交マッチング追跡, 18, 183

## ■ つ

通時的誤差逆伝播法, 277

## ■ て

ティホノフ正則化, → 重み減衰

ディラックのデルタ関数, 48

適応的線形要素, 11, 17, 18

適合率, 307

敵対的学習, 192, 195, 388

敵対的事例, 192

敵対的生成ネットワーク, 506, 515

テスト集合, 80

データ集合, 75

データ集合拡張, 194, 332

データ生成過程, 80

データ生成分布, 80, 95

データ並列処理, 324

テプリッツ行列, 239

転移学習, 392

転写, 73

伝承サンプリング, 423, 435

点推定量, 89

テンソル, xiv, 24

転置, xv, 24

伝統的な統計, → 頻度論統計

テンバリング, 441

テンプレートマッチング, 102

## ■ と

等価性, 242

動径基底関数, 141

統計的学習理論, 80

統計量, 89

同時確率分布, 42  
 等式制約, 68  
 動的構造, 325  
 等方性, 48  
 特異値, → 特異値分解  
 特異値分解, 33, 107, 348  
 特異ベクトル, → 特異値分解  
 特徴量, 72  
 特徴量選択, 169  
 独立, xvi, 44  
 独立成分分析, 359  
 独立同一分布, → i.i.d. 仮定  
 独立部分空間分析, 360  
 凸最適化, 102  
 ドット積, 25, 102  
 トポグラフィック ICA, 360  
 ドメイン適応, 392  
 トライグラム, 335  
 トレース演算子, 34  
 ドロップアウト, 185, 308, 312, 494, 506  
 ドロップコネクト, 191  
 貪欲教師あり事前学習, 232  
 貪欲法, 232

## ■ な

ナイーブベイズ, 2

## ■ に

二階微分, 63  
 二階微分による極値判定法, 65  
 二項関係, 350  
 二重逆伝播法, 195  
 二重循環行列, 239  
 ニュートン法, 66, 223  
 ニューラル言語モデル, 336, 346  
 ニューラル自己回帰密度推定器, 521  
 ニューラルチューリングマシン, 301  
 ニューラルネットワーク, 10

## ■ ね

ネオコグニトロン, 12, 17, 18, 263  
 ネステロフのモメンタム, 215

## ■ の

ノーフリーランチ定理, 85  
 ノルム, xvii, 29  
 ノンパラメトリックモデル, 83

## ■ は

バイアス, 90, 163

バイアス重点サンプリング, 434  
 バイアスパラメータ, 79  
 バイグラム, 335  
 ハイパーパラメータ, 87, 312  
 ハイパーパラメータ最適化, 313  
 バギング, 183  
 白色化, 330  
 パーセプトロン, 11, 18  
 バッチ正規化, 192, 308  
 ハーモニウム, → 制限付きボルツマンマシン  
 ハーモニー理論, 416  
 パラメータ共有, 181, 241, 269, 270, 280  
 パラメータ拘束, → パラメータ共有  
 パラメータの初期化, 215, 293  
 パラメトリック ReLU, 138  
 パラメトリックモデル, 83  
 バリアンス, xvi, 163  
 バーンイン, 437  
 汎化, 80  
 汎関数微分, 473  
 半教師あり学習, 174  
 万能近似器, 404  
 万能近似定理, 142

## ■ ひ

非正規化確率分布, 413  
 微分, xvi, 61  
 微分エントロピー, 55, 474  
 非モラル, 420  
 表現学習, 2  
 表現容量, 81  
 標準誤差, 92  
 標準偏差, 45  
 頻度確率, 40  
 頻度論統計, 98

## ■ ふ

復号化器, 3  
 複雑細胞, 262  
 負項, 341  
 符号化器, 3  
 負段階, 443, 445  
 物体検出, 328  
 物体認識, 328  
 負定値, 65  
 不等式制約, 68  
 不等分散, 135  
 不偏, 90  
 不変性, 244

フーリエ変換, 259  
 プーリング, 237, 502  
 ブロックギブスサンプリング, 438  
 ブロードキャストリング, 25  
 フロベニウスノルム, 34

分散, 44  
 分散表現, 13, 109, 399  
 分配関数, 414, 443, 492  
 文脈付きバンディット, 349  
 文脈特有の独立性, 419  
 分離 (確率モデリング), 417  
 分類, 72

## ■ ヘ

平均二乗誤差, 78  
 平均の標準誤差, 93, 200  
 平均場近似, 468, 469, 494  
 ベイジアンネットワーク, → 有向グラフィカルモデル  
 ベイズ確率, 40  
 ベイズ誤差, 84  
 ベイズ則, 52  
 ベイズ的ハイパーパラメータ最適化, 316  
 ベイズ統計, 98  
 並列分散処理, 13  
 ベクトル, xiv, 23  
 ヘッセ行列, xvi, 64, 160  
 ベルヌーイ分布, 45  
 変動の要因, 4  
 偏微分, 62  
 変分自己符号化器, 506, 512  
 変分自由エネルギー, → エビデンス下界  
 変分導関数, → 汎関数微分  
 変分法, 129  
 弁明, 418, 463, 472

## ■ ほ

忘却ゲート, 219  
 方向微分, 62  
 方策, 349  
 ほとんど至るところで, 52  
 ボルツマン分布, 416  
 ボルツマンマシン, 416, 481

## ■ ま

前処理, 328  
 マックスアウト, 138, 308  
 マルコフ確率場, 412  
 マルコフネットワーク, → 無向モデル  
 マルコフ連鎖, 435

マルコフ連鎖モンテカルロ法, 435  
 マルチタスク学習, 174, 393  
 マルチヌーイ分布, 46  
 マルチモーダル学習, 394

## ■ み

密度推定, 74  
 ミニバッチ, 200

## ■ む

ムーア・ペンローズ擬似逆行列, 33, 171  
 無向グラフィカルモデル, 57, 370  
 無向モデル, 412

## ■ め

メモリネットワーク, 300, 301

## ■ も

網羅率, 307  
 目的関数, 61  
 モデル圧縮, 324  
 モラル化グラフ, 420  
 モデル同定可能性, 204  
 モデル平均化, 183  
 モデル並列処理, 324  
 モーメントマッチング, 517  
 モーメントマッチング生成ネットワーク, 517

## ■ や

焼きなまし重点サンプリング, 458, 491, 527  
 ヤコビ行列, xvi, 53, 63  
 山登り法, 63

## ■ ゆ

有限差分法, 318  
 有向グラフィカルモデル, 57, 370, 411, 509  
 有効容量, 82  
 ユークリッドノルム, 29  
 ユニグラム, 335

## ■ よ

要素ごとの積, → アダマール積  
 予測スパース分解, 382

## ■ ら

ラグランジアン, → 一般化ラグランジアン  
 ラグランジュの未定乗数法, 68, 474  
 ラプラス分布, 48, 362  
 ラベル平滑化, 173  
 ランダムサーチ, 315

## ■ り

リキッドステートマシン, 291

リッジ回帰, → 重み減衰

リブシット定数, 67

リブシット連続, 67

リンク予測, 352

## ■ る

ループ, 420

ループあり確率伝播法, 427

## ■ れ

レシオマッチング, 453

レム睡眠, 446, 478

連鎖律（微積分）, 147

## ■ ろ

ロジスティック回帰, 2, 101, 102

ロジスティックシグモイド, 6, 49

## ■ わ

ワンショット学習, 393