

- Local conditional probability distribution, 560
- Local contrast normalization, 452
- Logistic regression, 3, 138, 139
- Logistic sigmoid, 7, 66
- Long short-term memory, 18, 24, 304, 407, 422
- Loop, 575
- Loopy belief propagation, 581
- Loss function, *see* objective function
- L^p norm, 38
- LSTM, *see* long short-term memory
- M-step, 629
- Machine learning, 2
- Machine translation, 100
- Main diagonal, 32
- Manifold, 159
- Manifold hypothesis, 160
- Manifold learning, 160
- Manifold tangent classifier, 268
- MAP approximation, 137, 501
- Marginal probability, 57
- Markov chain, 591
- Markov chain Monte Carlo, 591
- Markov network, *see* undirected model
- Markov random field, *see* undirected model
- Matrix, xi, xii, 31
- Matrix inverse, 35
- Matrix product, 33
- Max norm, 39
- Max pooling, 336
- Maximum likelihood, 130
- Maxout, 191, 422
- MCMC, *see* Markov chain Monte Carlo
- Mean field, 633, 634, 666
- Mean squared error, 107
- Measure theory, 70
- Measure zero, 70
- Memory network, 413, 415
- Method of steepest descent, *see* gradient descent
- Minibatch, 277
- Missing inputs, 99
- Mixing (Markov chain), 597
- Mixture density networks, 187
- Mixture distribution, 65
- Mixture model, 187, 506
- Mixture of experts, 446, 544
- MLP, *see* multilayer perception
- MNIST, 20, 21, 666
- Model averaging, 252
- Model compression, 444
- Model identifiability, 282
- Model parallelism, 444
- Moment matching, 696
- Moore-Penrose pseudoinverse, 44, 237
- Moralized graph, 573
- MP-DBM, *see* multi-prediction DBM
- MRF (Markov Random Field), *see* undirected model
- MSE, *see* mean squared error
- Multi-modal learning, 535
- Multi-prediction DBM, 668
- Multi-task learning, 242, 533
- Multilayer perception, 5
- Multilayer perceptron, 26
- Multinomial distribution, 61
- Multinoulli distribution, 61
- n -gram, 458
- NADE, 702
- Naive Bayes, 3
- Nat, 72
- Natural image, 555
- Natural language processing, 457
- Nearest neighbor regression, 114
- Negative definite, 88
- Negative phase, 466, 602, 604
- Neocognitron, 16, 23, 26, 364
- Nesterov momentum, 298
- Netflix Grand Prize, 255, 475
- Neural language model, 460, 472
- Neural network, 13
- Neural Turing machine, 415
- Neuroscience, 15
- Newton's method, 88, 309
- NLM, *see* neural language model
- NLP, *see* natural language processing
- No free lunch theorem, 115