

Index

- $A(n, m, p)$ architecture, 64, 69–71
 - autoencoder, 72
- activation
 - linear, 17, 18
 - polynomial, 18
 - quadratic, 18
- adaboost, *see* adaptive boosting
- adaptive boosting, 114
- adaptive PL transfer function, *see* APL
- adaptive random backpropagation, *see* ARBP
- adaptive regularization
 - linear network, 123
 - linear unit, 122
- adaptive skipped random backpropagation, *see* ASRBP
- adversarial architecture, 26
- adversarial attack, 276
- adversarial training, 128
- alternate minimization, 84
- amino acid, 258
- ANN, *see* neural network, artificial
- antihydrogen, 209, 210
- APL, 21
- approximation theorem, 65, 67, 68, 70
- ARBP, 153
- architecture
 - adversarial, 26
 - autoencoder, 26
 - ensemble, 26
 - feedforward, 24
 - layered, 24
 - locally connected, 24
 - recurrent, 24
 - Siamese, 25
- architecture depth, 111
- architecture design, 113
- ARM, 129
- artificial neural network, *see* neural network, artificial
- ASACUSA experiment, 209
- ASRBP, 153
- ATLAS detector, 207, 214
- atom-mapping, 248
- attention layer, 126
- attention mechanism, *see* gating, mechanism
- autism, 4
- auto-association, 75, 78
- autoencoder, 71
 - $A(n, m, p)$ architecture, 72
 - Boolean
 - complexity, 89
 - unrestricted, 84–87, 91, 97
 - circular, 164
 - composition, 94, 95
 - compressive, 73
 - critical point, 80
 - denoising, 91, 92
 - expansive, 73, 94
 - global minimum, 82
 - inverting, 91
 - linear, 73, 75
 - denoising, 82
 - invariance properties, 77
 - non-linear, 73, 84
 - probabilistic, 92
 - redressing, 91
 - types of, 90
 - variational, 93, 128
- autoencoder architecture, 26, 72
- autoencoder problem, 73
 - linear, 76
- autoencoder properties, 74
- autoregressive generative model, *see* ARM
- backpropagation, 105, 140, 145–147, 151, 162
 - deep targets, 148
 - efficiency, 110
 - non-locality, 141
 - probabilistic layers, 108
 - random, *see* RBP
- backpropagation error, 106, 107
- backpropagation learning equation
 - linear feedforward network, 108
- backpropagation learning rule, 12
- bagging, 114
- batch learning, 34
- Bayes' theorem, 29
- Bayesian analysis, 28

- Bayesian framework, 28
 - not-feasible, 45
- belief, degree of, 28
- belief, subjective, 28
- Bellman equation, 288
- Bellman operator, 288
- bias, 118
 - high, 119
- bias of a neuron, 18
- bias–variance decomposition, 118, 134
- biological computer, 2
- Blahut–Arimoto algorithm, 93
- Boltzmann distribution, 21, 182
- Boltzmann machine, 12, 179, 182, 188
 - conditional, 184
 - learning, 182
 - optimization, 182
 - restricted, 184
- Bolzano–Weierstrass theorem, 70
- Boolean function
 - differentiable transfer function, 62
- Boolean gate, 19, 20
- Boolean model
 - unrestricted, 64
- boosting, 114
- bootstrap aggregating, *see* bagging
- bottleneck, 77, 93, 111, 115
- brain structure, 6
- BRNN, *see* IOHMM, bidirectional
- Cambridge Structural Database, 244
- canonicalization, 112
- capacity, 26, 163
 - binary weights, 61
 - cardinal, 26, 38, 68
 - deep feedforward architecture, 103
 - recurrent network, 178
 - functional, 26, 63
 - deep linear architecture, 101, 104
 - deep unrestricted Boolean architecture, 102
 - linear threshold function, 48
 - linear threshold gate, 46
 - linear unit, 52
 - polynomial threshold function, 50
 - positive weights, 61
 - ReLU gate, 52
 - ReLU transfer function, 105
 - sets, 60
 - training, 63, 70, 105
- capacity measures, 104
- carbon-based computer, 4
- carbon-based computing, 2
- cardinal capacity
 - $A(n, m, 1)$ architecture, 68
 - $A(n, m, p)$ architecture, 69
 - recurrent network, 178
- chemical reaction
 - elementary, 248
 - global, 248
 - prediction, 249, 254
 - deep learning, 252
 - quantum-mechanical methods, 249
 - rule-based methods, 250
 - representation, 246, 247
- chromatin, 260
- chromosome, 260
- circuit
 - depth, 23
- circular architecture, 164
- classification
 - binary, 42
 - k , 43
- classification error, 128
- codon, 259
- complexity reduction, 125
- computational complexity, 36
- computational immunology, 268
- computing unit, *see* unit, computing
- conditioning
 - operant, 282
- congruent weight, 153
- conjugate transposition, 82
- connected function, 47
- connected problem, 138
- contact map, 195, 263
 - protein, 263–264
 - prediction, 265
- contrastive learning, 184
- convexity of error function, 53
- convolution, 113
- convolutional architecture, 139
- convolutional layer, 112
- convolutional network
 - recursiveness, 191
- convolutional neural network, 25
- countability, 39
- countable, 36
- credit assignment problem, 23
- critical point
 - autoencoder, 80
 - full rank, 81
- Crystallography Open Database, 244
- curse of dimensionality, 67, 298
- DAG, *see* directed acyclic graph
- dark knowledge, 125, 273
- data normalization, 55
- data processing inequality, 32, 39
- data-augmentation, 113
- decision problem, 36
 - NP , 37
 - NP -complete, 37
 - NP -hard, 37
 - P , 37

- deep feedforward architecture: top layer, 56
- deep learning
 - vs. linear regression, 5
 - antihydrogen experiments, 209
 - biomedical data, 257
 - chemical discovery, 241
 - chemistry
 - need for data, 254
 - climate physics, 233
 - complex physical systems, 237
 - cosmology, 231
 - dark matter physics, 228
 - data lean, 258
 - exotic particles, 215
 - experimental physics, 208
 - generative models, 127
 - genomics, 268
 - high-energy physics, 214
 - influence on experiments, 238
 - influence on theoretical physics, 238
 - jet particles, 218, 236
 - medical data, 274
 - medical imaging, 271
 - molecular dynamics, 250
 - molecular representation, 243
 - molecular structure prediction, 244
 - neutrino physics, 224
 - peptide identification, 268
 - proteomics, 261
 - reinforcement learning, 303
 - small molecule properties, 242
 - transcriptomics, 268
- deep learning channel, *see* DLC
- deep local learning, 142
 - limitations, 143
- deep targets learning, 147
- deep targets learning algorithm, 148
- Delta rule, 10
- depth, 23
- directed acyclic graph, 192
- distance map
 - protein, 263–264
 - prediction, 265
- distillation, *see* dark knowledge
- diversity of models, 114
- DLC, 144, 151–155
 - biological systems, 150
- DNA, 258, 259
- dropout, 119, 134, 138
 - adaptive regularization, 120
 - average, 120
 - linear network, 124
 - linear unit, 123
- dynamic programming, 283, 284, 288
 - distribution model, 295
 - exploring starts, 291
- Monte Carlo methods, 290
 - every-visit, 291
 - first-visit, 291
- policy iteration, 288
- sample model, 295
- stochastic policies, 291
 - off-policy, 291
 - on-policy, 291
- temporal difference methods, 292
 - actor–critic, 294
 - value iteration, 289
- early stopping, 118
- ensemble, 113, 134
- ensemble architecture, 26
- ensemble averaging
 - linear network, 120
 - non-linear network, 121
- ensemble methods, 113
- entropy, 31, 136
 - conditional, 31
 - differential, 97
 - relative, 31, 42, 43
- epigenetic modification, 260
- epigenome, 260
- equivalent architectures, 101
- equivariant function, 112
- error, 162
 - recirculation, 164
 - training, 122
- error autoencoder
 - full rank
 - critical point, 81
- Euclidean distance, 75
- evidence, 29
- evolutionary algorithm, 296
- exon, 259
- extreme machine, 56
- $\mathcal{F}(L, k, m, w)$ class of functions, 100
- feedforward architecture, 24, 99
- filtering, 68, 69
- fingerprint representation, 243
- FitzHugh–Nagumo model, 23
- fixed matrix, 77, 78
- flow, 130
 - normalizing, 130
- flow model, 129
- fully-connected layer, 24
- functional capacity
 - two-layer network, 63
- GAN, 128
- gate, *see* unit, computing; neuron
- gating, 125
 - mechanism, 125
 - output, 126

- gating (Cont.)
 - restrictive form, 126
 - synaptic, 125
- Gaussian process, 68
- GDB-*n*, 241
- generative adversarial network, *see* GAN
- generative model, 127, 136
- generative models of data, 127
- genetic information, 2
- genome, 259
 - as computer, 260
- geometric mean
 - normalised weighted, 121
 - weighted, 121
- GNN, *see also* RNN, 190
- gradient
 - stochastic, 56
- gradient clipping, 115
- gradient descent, 52, 116
- gradient descent equation, 106
 - stochastic, 106
- gradient descent learning
 - shallow network, 52
- gradient descent learning equation, 43, 52
- graph neural network, *see* GNN
- graphical model, 13
- growth function, 135

- halting problem, 36, 40
- Heaviside function, *see* threshold function
- Hebb's rule, 9
- Hebbian learning, 9, 11, 140, 283
 - local, 144
 - mathematical model, 9
- Hebbian rule, 180
- Hessian, 53
- Hessian methods, 110
- hetero-association, 72
- hidden Markov model, 13, *see* HMM
- hierarchical modular function, 100
- Hilbert's sixteenth problem, 162
- HMM, *see* hidden Markov model, 192
 - bidirectional, 192
 - factorial, 192
- Hodgkin–Huxley model, 6
- Hopfield model, 56, 179, 181, 186
 - NP-complete, 181
 - optimization, 181
 - shallow learning, 180
- Hopfield network, 179
- hyperparameter, 116
- hyperplane arrangements, 58
- hyperthymesia, 4

- inference
 - probabilistic, 13
- information, 146
- information bottleneck method, 32, 39
- information geometry, 32
- information plane, 134
- information storage
 - digital, 2
 - neural, 2
- information theory, 39
- inner approach, 191–197
 - representation of molecules, 245
 - vs. outer approach, 200
- intelligence, 1
- intron, 259
- invariance in network, 112
- invariance properties, 77
- invariant function, 112
- IOHMM, 192
 - bidirectional, 194

- Jensen's inequality, 31
- Johnson–Lindenstrauss Lemma, 57

- Keras, 118
- kernel method, 13
- kernel methods
 - chemistry, 243
- KL, *see* entropy, relative
- Kullback–Leibler divergence
 - seentropy, relative, 31

- LASSO, *see* regularization, L1
- layered architecture, 24
- learning, 26, 27
 - batch, 34
 - contrastive, 184
 - local, 137
 - mini-batch, 34
 - NP-completeness, 89, 93
 - on-line, 34
 - reinforcement, 35
 - self-supervised, 34
 - semi-supervised, 34
 - supervised, 33
 - transfer, 34
 - unsupervised, 34, 127
- learning algorithm, 35
 - incremental, 83
- learning algorithms, 8
- learning channel, 162
- learning rate, 55, 116
- learning rule
 - Oja, 150
- learning rules
 - functional form, 141
 - locality, 140, 141
 - stratification, 141
- learning system
 - generalization properties, 84
- likelihood, 29

- linear programming, 288
 - approximate, 303
- linear regression, 4
- local deep learning, 145, 147
- local learning, 137
 - deep, 142
- local learning principle, 139
- locality, 139, 140
 - learning rules, 140, 141
 - spatial, 166
 - spatial and temporal, 166
- logistic function, 21
- long-short term memory units, 13
- Lorentz-boosted neural network, 236
- LSTM, *see* long-short term memory units, 200
- MAP estimation, 29, 42
- Markov chain, 192
- Markov decision process, *see* MDP
- Markov decision process, partially observable, *see* POMDP
- Markov random field, 182
- maximum *a posteriori*, *see* MAP
- maximum likelihood, *see* ML
- MaxPool neuron, 127
- McCulloch–Pitts neuron, 7
- MDP, 286, 298, 299
- memory
 - associative, 180
 - learning, 7
 - long-term, 7
- mini-batch, 55
- mini-batch learning, 34
- minimal sufficient statistic, 32
- ML estimation, 29, 41
- model compression, 125
- molecular structure prediction, 244
- momentum, 116
- Monte Carlo tree search, 296
- mRNA, 259
- multiplexing, *see* filtering, 145
- multiplicative interactions, 125
- mutual information, 32
- neocognitron model, 11
- network
 - depth, 23
 - different, 178
 - equivalent, 178
 - shallow, 41
- network architecture
 - multilayer, 99
- neural machine, 2
- neural network
 - biological, 16
- neural network, 3, 16
 - artificial, 3
 - artificial vs. biological, 277
- autoencoder, 71
 - continuous time, 22
- convolutional, 11, 13, 112
 - deep learning, 14
 - invariant, 112
 - probabilistic, 45, 108
 - recurrent, 177
 - recursive, 13
 - shallow
 - local learning, 45
 - simulation, 22
 - single-compartment model, 22
 - two-layer, 65, 67
- neural network model, *see* neural network
- neuromorphic computing, 3
- neuron, 16, 17
 - spiking, 23
- noise, 56
- normalization
 - batch, 115
 - layer, 115
 - minibatch, 115
- NP-completeness, 181
- nucleosome, 260
- nucleotide, 258
- NWGM, *see* geometric mean, normalised weighted
- Oja learning rule, 150
- online-learning, 34
- Ornstein–Uhlenbeck process, 56
- outer approach, 191, 197–198
 - representation of molecules, 245
 - vs. inner approach, 200
- outerproduct formula, *see* Hebbian rule
- overfitting, 112, 117, 119, 120
- pair function, 47
- parity function, 47
- parity problem, 138
- partially connected layer, 24
- partition function, 182
- PCA, 77, 82, 91
- perceptron, 10, 53
 - multi-layer, 11
- perceptron cycling theorem, 55
- perceptron learning algorithm, 10
- perceptron learning rule, 54, 55
- physics-aware architecture, 235
- piecewise linear transfer function, *see* PL transfer function
- PL transfer function, 20
- policy gradient, 297
- polynomial threshold function
 - low-degree, 49
- posterior, 29
- postsynapsis, 106
- postsynaptic, 162
- precision reduction, 125

- presynaptic activity, 106
 principal component analysis, *see* PCA
 prior, 29, 44
 horseshoe, 44
 shrinkage, 44
 probabilistic neural network
 sampling, 45
 projection matrix, 77
 protein, 258, 260
 classification, 268
 globular, 262
 membrane, 262
 protein structure, 261
 prediction, 261, 263
 secondary prediction, 264
 PubChem, 241
 Purkinje cells, 6
- Q-learning
 Q-learning, 292
 quaternary structure, 266
- random backpropagation, *see* RBP
 random direction descent, 110
 optimized, 110
 random matrix, 152
 RBM, 184
 RBP, 152, 155, 162, 168, 170, 173
 simulating, 154
 ReactionPredictor, 253
 recirculation, 165
 biological networks, 172
 convex combination, 174
 error, 165
 multiple cycles, 174, 175
 output, 165
 plain, 168
 simulation, 167
 rectified linear transfer unit, *see* ReLU
 recurrent architecture, 24
 recurrent network
 cardinal capacity, 178
 recurrent neural network, 177
 sequential, 177
 stochastic, 178
 synchronous, 177
 recursive neural network
 application in chemistry, 245
 recursive neural network, *see* RNN
 regression, 42
 regularization, 44
 L1, 44, 58
 L2, 44
 Lp, 44, 58
 regularizer, 116
 reinforcement learning, 35
 action–value function, 285
 actions, 284
 agent, 284
 algorithms, 286
 model-based, 287
 model-free, 287
 policy space search, 287
 value function search space, 287
 complexity, 298
 deep, 303
 model-based, 305
 model-free, 304
 discount factor, 285
 dynamic programming, 284
 environment, 284
 hierarchical, 302
 inverse, 302
 multi-agent, 301
 optimal policy, 286
 planning, 295
 policy, 284
 quantum, 303
 relational, 303
 return, 285
 reward, 284
 shaping, 299
 states, 284
 value function, 285
 reinforcement learning, 282–306
 applications, 283
 relative entropy
 binary-classification, 42
 k-classification, 43
- ReLU
 capacity, 105
 ReLU, 20
 leaky, 20
 restricted Boltzmann machine,
 see RBM
 RNA, 258, 259
 RNA, messenger, *see* mRNA
 RNN, 190
 robotics, 284
- SARSA algorithm, 293
 satisfiability, 61
 Sauer–Shelah lemma, 135
 SBP, 153
 self-supervised learning, 34, 71
 semi-supervised learning, 34
 separability
 linear, 54
 separability of Boolean functions
 linear, 46, 47
 SGD, 109, 110, 133
 shallow learning
 Hopfield model, 180
 local, 46
 supervised, 43
 shallow network, 23, 41

- capacity, 46
- feedforward, 45
- shallow neural network, 56
- shattering, 105
- Siamese architecture, 25
- Siamese network, 190
- sigmoidal transfer function, 21
- Signal-P, 268
- silicon-based computer, 4
- silicon-based computing, 2
- similarity metric
 - chemistry, 243
 - Jaccard–Tanimoto, 243
- simulated annealing, 110
- simulated data
 - particle physics, 223
- singular value decomposition, *see* SVD
- skipped backpropagation, *see* SBP
- skipped random backpropagation, *see* SRBP,
 - skipped
- small molecule, 240, 258
- SMARTS, 247
- SMILES, 243, 245
- SMIRKS, 247
- softmax transfer function, 21
- sparsity, 94
- spike, 22
- spike time-dependent synaptic plasticity,
 - see* STDP
- spin glass, 12, 181
- spurious minimum, 133
- SRBP, 153, 156, 162
- SRMP, 173
- state-relevance weights, 288
- STDP, 166
- step function, *see* threshold function
- stochastic game, 283, 301
- stochastic gradient, 56
- stochastic gradient descent, *see* SGD
- stochasticity in neural network, 45
- supervised learning, 33, 162
 - shallow, 43
- support vector machine, 13, 57
- surprise, 32
- SVD, 82
- SVM, *see* support vector machine
- synapse, 6, 16
 - biological, 16
 - changes in, 8
 - storage capacity, 17
 - weight (w_{ij}), 16
- synaptic strength, *see* synapse, weight
- synaptic weight, 16, 23
 - multiplication, 125
- TD(λ) algorithm, 293
 - eligibility trace, 294
- testing set, 116
- threshold function
 - linear, 19, 46
 - polynomial, 46
 - low-degree, 49
- threshold gate
 - Boolean, 19
 - homogeneous Boolean linear, 19
 - homogeneous Boolean polynomial, 20
 - linear
 - capacity, 46
 - non-homogeneous Boolean linear, 19
 - non-homogeneous Boolean polynomial, 20
- threshold gates
 - satisfiability, 61
- training capacity, 70
- two-layer network, 63
- training set, 116
- transfer function, 18
 - APL, 21
 - arctan, 21
 - differentiable, 62
 - linear, 18
 - normalized exponential, 21
 - PL, 20
 - polynomial, 18
 - sigmoidal, 21, 133
 - softmax, 21
 - tanh, 21
- transfer learning, 34
- transitivity, 28
- traveling salesman problem, 36, 186
- two-layer neural network
 - approximating, 65, 67
- underfitting, 117
- unit
 - hidden, 23
 - max pooling, 19
 - visible, 23
- unit, computing, 17
- universal approximation, 100
- universal approximation property, 63
- unsupervised learning, 34, 127
- validation set, 116
- variable-size data, 189
- variance
 - high, 119
- variational approach to estimation, 30
- VC dimension, 105, 135
- vector exponential family, 188
- virtualization, 137
- weight decay, 44
- weight initialization, 55
- weight sharing, 25, 191
- Widrow–Hoff rule, 10
- WIMP, 229