
Index

A

Accidental properties, 108
Accuracy, 57
Activation function, 67, 81
Adaptable learning rate, 97
Analogical reasoning, 14
Artificial intelligence, 51
Autoencoder, 156, 177

B

BAbI, 181
Backpropagation, 17, 79, 83, 87, 93
Backpropagation through time, 150
Bag, 18
Bag of words, 18, 75, 141, 148, 166
Bayes theorem, 59
Benchmark, viii
Bernoulli distribution, 35
Bias, 36
Bias absorption, 84
Binary threshold neuron, 84
Binning, 33
Boltzmann machine, 176

C

Categorical features, 55
CBOW, 166
Centroid, 70
Chain rule, 24
Classification, 51
Clustering, 70
Cognitive science, 51
Committee, 69
Confusion matrix, 58
Connectionism, 10, 12, 14, 15, 186
Continuous function, 20

Contrastive divergence, 177
Convergence, 20
Convolutional layer, 122, 128
Convolutional neural network, 69, 79
Corpus, 75
Correlation, 72
Cosine similarity, 166
Covariance, 153
Covariance matrix, 153
Cross-entropy error function, 151
Curriculum learning, 117

D

Datapoint, 52
Dataset, 54
1D convolutional layer, 122
2D convolutional layer, 123
Deep belief networks, 177
Delta rule, 88
Distributed representations, 70, 72, 165
DMB, 10
Dot product, 26
Dropout, 115
Dunn coefficient, 71

E

E, 20
Early stopping, 114
Eigendecomposition, 155
Eigenvalue, 154
Eigenvectors, 154
Elman networks, 10, 141
Epoch, 65, 102, 117
Error function, 64, 89
Estimator, 36
Euclidean distance, 25

Euclidean norm, 109

Expected value, 35

F

False negative, 57

False positive, 57

Feature engineering, 55

Feature maps, 125

Features, 33, 52

Feed-forward neural network, 95

Finite difference approximation, 95

Forward pass, 83

Fragment, 75

Fully-connected layer, 124, 131

Function minimization, 24

G

Gaussian cloud, 38

Gaussian distribution, 37

General weight update rule, 65, 97, 110, 115

GOFAI, *v*, 12, 15, 171, 186

Gradient, 30, 31, 99

Gradient descent, 17, 24, 31, 93

H

Hadamard product, 179

Hopfield networks, 175

Hyperbolic tangent, 67, 144

Hyperparameter, 89, 114

Hyperplane, 31, 53

I

Iterable, 44

Iteration, 102

J

Jordan networks, 10, 141

K

K-means, 70

Kullback-Liebler divergence, 176

L

L1 regularization, 110

L2 norm, 26

L2 pooling, 162

L2 regularization, 109

Label, 52

Latent variable, 72, 157

Learning rate, 31, 89, 94, 97, 111

Limit, 21

Linear combination, 26

Linear constraint, 88

Linearly separable, 55

Linear neuron, 89

List, 18

Local minima, 115

Local receptive field, 122

Local representations, 72

Logistic function, 62, 67, 81, 140, 143

Logistic neuron, 90

Logistic regression, 61, 79, 84, 108

Logit, 62, 66, 67, 81, 97

Low faculty reasoning, 171

LSTM, 10, 142

M

Markov assumption, 137, 150

Matrix transposition, 27

Max-pooling, 125, 128

Mean squared error, 89

Median, 33

Memory networks, 179, 181

MNIST, 10, 127, 135, 159

Mode, 33

Momentum, 114

Momentum rate, 115

Monotone function, 20

MSE, 102

Multiclass classification, 52

Multilayered perceptron, 87

Mutability, 18

N

Naive Bayes classifier, 59

Necessary property, 107

Neural language models, 165

Neural turing-machines, 178

Neuron, 80

Noise, 74

Nonlinearity, 67

Normalized vector, 26

Numpy, 159

O

One-hot encoding, 19, 56, 64, 128, 148, 166

Online learning, 102

Ordinal feature, 55

Orthogonal matrix, 30

Orthogonal vectors, 26

Orthonormal, 26

Overfitting, 107

P

Padding, 123
Parameters, 63
Parity, 86
Partial derivative, 30
PCA, 70, 72, 172
Perceptron, 6, 84, 87
Positive-definite matrix, 154
Precision, 58
Prior, 35
Probability distribution, 35
Python indentation, 44

Q

Qualia, 69

R

Recall, 58
Receptive field, 121
Recurrent neural networks, 136
Regularization, 108
Regularization rate, 109
Reinforcement learning, 51, 117
ReLU, 20, 124
Restricted Boltzmann machine, 177
Ridge regression, 109
Row vector, 27

S

Scalar multiplication, 28
Sentiment analysis, 130
Shallow neural networks, 79
Sigmoid function, 62, 81, 90
Simple recurrent networks, 141
Skip-gram, 166
Softmax, 129, 140, 146, 167
Sparse encoding, 76
Square matrix, 28
Standard basis, 26
Standard deviation, 36
Step function, 20
Stochastic gradient descent, 129
Stride, 123
Supervised learning, 51

Support vector machine, 10
Symmetric matrix, 28

T

Target, 52
Tensor, 30
Test set, 58
Tikhonov regularization, 109
Train-test split, 59
True negative, 57
True positive, 57
Tuple, 18

U

Underfitting, 107
Uniform distribution, 35
Unit matrix, 30
Unsupervised learning, 51
Urelements, 17

V

Validation set, 112
Vanishing gradient, 118
Variance, 36
Vector, 18
Vector component, 18
Vector dimensionality, 18
Vector space, 25
Voting function, 20

W

Weight, 80
Weight decay, 109
Winograd sentences, 181
Word embedding, 165
Word2vec, 14, 166, 168, 180
Workhorse neuron, 62, 66, 121

X

XOR, 86

Z

Zero matrix, 30