

# Index

## Symbols

$\beta$ -VAE, 297  
 $\sigma$ -algebra, 9  
 $c$ -transforms, 279  
 $f$ -GAN, 285  
 $f$ -divergence, 272, 285

## A

Absolutely continuous, 270  
Absolutely continuous measure, 10  
Action potential, 82  
Activation function, 93  
Adaptive instance normalization (AdaIN), 161  
Adjacency matrix, 138  
Adjoint, 5  
Affine function, 18  
AlexNet, 114  
Algorithmic robustness, 261  
Ambient space, 61  
Atlas, 221  
Attention, 164  
Attentional GAN (AttnGAN), 172  
Autoencoder, 220  
Auxiliary classifier, 307  
Average pooling, 120  
Axon, 80  
Axon hillock, 80, 82

## B

Backpropagation, 102  
Backward-propagated estimation error, 106  
Bag-of-words (BOW) kernel, 64  
Banach space, 6

Basis, 8  
Basis pursuit, 200  
Basis vectors, 8  
Batch norm (BN), 158  
Batch normalization, 157  
Benign optimization landscape, 233  
Bias–variance trade-off, 56  
Bidirectional Encoder Representations from Transformers (BERT), 178  
Biological neural network, 79  
Break point, 249

## C

Calculus of variations, 106  
Cauchy–Schwarz inequality, 6  
Cauchy sequence, 4  
Channel attention, 168  
Chart, 221  
Chemical synapses, 81  
Classifier, 29  
Collaborative GAN (CollaGAN), 307  
Colored graph, 139  
Community detection, 135  
Complete, 4  
Compressed sensing, 200  
Concave, 20  
Concentration inequalities, 245  
Content image, 161  
Continuous bag-of-words (CBOW), 141  
Convex, 19  
Convex conjugate, 21  
Convolution, 119  
Convolutional neural network (CNN), 113  
Convolution framelet, 205

Corpus vocabulary, 143  
 Counting measure, 10  
 Covariate shift, 158  
 Cross-domain attention, 172  
 Cross entropy, 274  
 Cycle-consistency, 305  
 CycleGAN, 303

**D**

Data augmentation, 127  
 Deep convolutional framelets, 208  
 DeepWalks, 147  
 Dendrites, 80  
 Denominator layout, 15  
 Dense Convolutional Network (DenseNet),  
 117  
 Dependent variable, 46  
 Discriminator, 284  
 Disentanglement, 297  
 Divergences, 272  
 Domain, 17  
 Double descent, 256  
 Dropout, 128  
 Dual frame, 199  
 Duality gap, 24

**E**

Earth-mover distance, 276  
 Edges, 135, 138  
 Eigen-decomposition, 12  
 Electric synapses, 80  
 Empirical risk minimization (ERM), 244  
 Encoder–decoder CNN, 209  
 Entropy regularization, 281  
 Evidence lower bound (ELBO), 292  
 Excitatory postsynaptic potentials (EPSPs),  
 81  
 Expressivity, 196, 212

**F**

Feature engineering, 42  
 Feature space, 61  
 Feedforward neural network, 95  
 First-order necessary conditions (FONC),  
 34  
 Fixed points, 17  
 Forward-propagated input, 106  
 Fréchet differentiable, 21  
 Frame, 9, 198  
 Frame condition, 199  
 Framelets, 200

**G**

General linear model (GLM), 47  
 Generative adversarial network (GAN), 283  
 Generative models, 267  
 Generative pre-trained transformer (GPT), 182  
 Generator, 284  
 GLOW, 301  
 GoogLeNet, 115  
 Gradient descent method, 98  
 Gram matrix, 54  
 Graph, 135  
 Graph attention network (GAT), 172  
 Graph coloring, 139  
 Graph embedding, 139  
 Graph isomorphism, 138  
 Graph neural networks (GNNs), 148  
 Growth function, 248

**H**

Haar wavelet, 200  
 Hankel matrix, 204  
 Hilbert space, 6  
 Hinge loss, 38  
 Hoeffding's inequality, 245  
 Hubel and Wiesel model, 86

**I**

ImageNet, 114  
 Image style transfer, 161  
 Implicit bias, 234, 261  
 Inception module, 115  
 Independent variable, 46  
 Indicator function, 18  
 Induced norm, 6  
 Inductive bias, 260  
 Inhibitory postsynaptic potentials (IPSPs), 81  
 Inner product, 5  
 Instance normalization, 160  
 Interpolation threshold, 256  
 Invexity, 233  
 Ionotropic receptors, 164

**J**

Jennifer Aniston cell, 88  
 Jensen–Shannon (JS) divergence, 274

**K**

Kantorovich formulation, 278  
 Karush–Kuhn–Tucker (KKT) conditions, 35  
 Kernel, 63

Kernel SVM, 41  
 Kernel trick, 40  
 Key, 166  
 Kronecker product, 13  
 Kullback–Leibler (KL) divergence, 273

**L**

Lagrangian dual problem, 26  
 Lagrangian multipliers, 26  
 Latent space, 61  
 Layer normalization, 160  
 Least squares (LS) regression, 46  
 LeNet, 113  
 Lifting, 40  
 Linearly independent, 7  
 Linear operators, 17  
 Link analysis, 135  
 Lipschitz continuous, 4  
 Logistic regression, 48  
 Logit, 49  
 Long-term depression (LTD), 82  
 Long-term potentiation (LTP), 82  
 Loss landscape, 232  
 Loss surfaces, 227  
 Lyapunov function, 235  
 Lyapunov global asymptotic stability theorem, 236  
 Lyapunov stability analysis, 235

**M**

Manifold, 221  
 Masked self-attention, 184  
 Matrix factorization, 146  
 Matrix inversion lemma, 13  
 Matrix norm, 13  
 Maximum margin linear classifier, 31  
 Max pooling, 120  
 Mean squared error (MSE), 47  
 Metabotropic receptors, 164  
 Metric, 3  
 Metric space, 3  
 Mini-batch, 126  
 Momentum method, 101  
 Multi-input multi-output (MIMO) convolution, 120  
 Multi-input single-output (MISO) convolution, 119  
 Multiset, 149

**N**

Natural language processing (NLP), 139  
 Neural tangent kernel (NTK), 237

Neuron, 80  
 Neurotransmitters, 81  
 Node, 135  
 Node classification, 135  
 Node coloring, 139  
 Node2vec, 147  
 Nonlinear independent component estimation (NICE), 299  
 Norm, 5  
 Normalizing flow (NF), 298  
 Null space, 11  
 Numerator layout, 15

**O**

Odds, 49  
 Optimal transport, 277  
 Orthogonal complement, 5  
 Over-parameterization, 227  
 Overfitting, 126

**P**

PAC–Bayes bounds, 254  
 Perceptual loss, 116  
 Perfect reconstruction condition, 210  
 Piecewise linear partition, 223  
 Pix2pix, 302  
 Point evaluation function, 66  
 Polyak–Łojasiewicz (PL) condition, 229  
 Pooling, 120  
 Population risk, 244  
 Positional encoding, 177  
 Positive definite, 19  
 Positive definite kernel, 41, 65  
 Positive semidefinite, 18  
 Probability measure, 9  
 Probability space, 9  
 Probably approximately correct (PAC) learning, 245  
 Projector, 12  
 Push-forward of a measure, 269

**Q**

Query, 166

**R**

Rademacher complexity, 251  
 Radon–Nikodym derivative, 10, 270  
 Random variable, 11, 269  
 Random walks, 147  
 Range space, 11  
 Rank, 11

Real-valued non-volume-preserving (real NVP) transformations, 300  
 Receptive field, 87  
 Rectified linear unit (ReLU), 94  
 Regression analysis, 45  
 Regularization, 55  
 Relative entropy, 274  
 Reparameterization trick, 295  
 Representer theorem, 68  
 Reproducing kernel Hilbert space (RKHS), 62  
 Reproducing property, 67  
 Residual Network (ResNet), 117  
 Ridge regression, 51

## S

Sauer's lemma, 250  
 Score functions, 167  
 Self-attention GAN (SAGAN), 171  
 Shattering, 248  
 Sigmoid function, 49  
 Single-input single-output (SISO) circular convolution, 204  
 Singular value decomposition (SVD), 12  
 Sinkhorn distance, 281  
 Skip connections, 117, 124  
 Skip-gram, 141  
 Softmax loss, 125  
 Soma, 80  
 Span, 8  
 Sparse recovery, 200  
 Spatial attention, 166  
 Spectral normalization, 289  
 Squeeze and excitation network (SENet), 168  
 StarGAN, 307  
 Statistical distances, 272  
 Stochastic gradient descent (SGD), 98  
 Strong duality, 26  
 Strongly convex (SC), 229  
 StyleGAN, 169, 290  
 Style image, 161  
 Sub-differential, 21  
 Subspace, 12  
 Support vector, 35  
 Support vector machine (SVM), 35  
 Symmetrization lemma, 247  
 Synapse, 80  
 Synaptic terminals, 80

## T

Telodendria, 80  
 Test, 126  
 Tight frame, 9  
 Tikhonov regularization, 51  
 Trace, 11  
 Training, 126  
 Transformer, 174  
 Transportation cost, 277  
 Transport map, 277

## U

Unbiased estimator, 56  
 U-Net, 118  
 Union bound, 246  
 Union of subspaces, 201  
 Universal approximation theorem, 197  
 Unpooling, 122  
 Unsupervised learning, 301  
 Unweighted graph, 138

## V

Validation, 126  
 Vanishing gradient problem, 117  
 Variational autoencoder (VAE), 291  
 VC bound, 249  
 VC dimension, 249  
 Vector space, 4  
 Vertices, 138  
 VGGNet, 116  
 Vision Transformer (ViT), 185

## W

Wasserstein GAN (W-GAN), 288  
 Wasserstein metric, 274  
 Wasserstein-1 metric, 276  
 Wavelet frame, 200  
 Wavelet shrinkage, 200  
 Weight clipping, 289  
 Weighted graphs, 138  
 Weisfeiler–Lehman (WL) isomorphism test, 150  
 W-GAN with the gradient penalty (WGAN-GP), 289  
 Whitening and coloring transform (WCT), 163  
 Word embedding, 140  
 Word2vec, 141