

INDEX

A

- activation function, 131–132
- Adam, 267, 278, 296
- ANN (artificial neural network), 132
- anomaly detection, 397
- area under the ROC curve (AUC),
224–225, 305
- array processing, 2

B

- backpropagation, 134, 148, 157, 160
- bagging, 83
- batch normalization, 354
- batch size, 186
- batch training, 152
- Bayes, Thomas, 73
- big-O notation, 115
- bootstrap sample, 83
- bounding box, 292, 408
- box plot, 39
 - IQR (interquartile range), 41
 - whiskers, 41
- Bunker, Chang and Eng, 454
- Burges, Christopher, 85

C

- capsule networks, 255
- Cartesian plane, 4
- case study, 315
 - classical models, 323
 - classifier
 - audio features, 323
 - neural network, 326
 - spectrograms, 335
 - CNN (convolutional neural network)
 - ensembles, 339
 - one-dimensional, 327
 - two-dimensional, 335

- dataset

- augmentation, 317
 - ESC-10, 316
 - preprocessing, 321
 - spectrogram, 332–333
 - sox (command line tool), 333

- categorical value, 16

- centroid, 69

- CIFAR-10 dataset, 50, 296

- binary vs. multiclass, 307

- building models, 297

- one-versus-rest, 307

- classes, 13–14

- classification, 13

- multilabel, 432

- Claude 3 model, 508–509

- CNN (convolutional neural network),
6, 237

- activations, 260

- anatomy of, 243

- architecture, 243–244

- bounding box, 292, 408

- building, 264

- constructing in Keras, 266

- convolution, 239

- exact, 242

- filter, 248

- valid, 242

- zero-padding, 242

- convolutional layer

- effect, 250

- initialization, 253

- operation, 247

- stacking, 252

- data flow, 246

- effective receptive field,
252–253

- end-to-end training, 261

- evaluating, 267

- CNN (*continued*)
 - experiments, 272
 - architecture, 272
 - epochs, 275
 - minibatches, 275
 - optimizers, 278
 - training-set size, 275
 - feature maps, 245
 - filter, 248
 - fully convolutional network, 257, 263, 280, 282
 - building and training, 280
 - graymap, 289
 - heatmap, 288
 - history, 239
 - kernel, 239
 - layers
 - convolutional, 244, 247
 - dense, 244
 - dropout, 244
 - flatten, 244
 - fully connected, 245, 255
 - fully convolutional, 256
 - inner product, 245
 - inputs, 245
 - pooling, 244, 254
 - ReLU, 244
 - layer types, 244
 - loss function, 267
 - max pooling, 254
 - motivation, 238
 - neocognitron, 239
 - optimizer keyword, 267
 - optimizers, 278
 - outputs, 258
 - plotting, 270
 - pooling layer, 254
 - probability map, 288
 - receptive field, 252–253
 - scaling data, 265
 - spatial invariance, 238
 - stride, 239, 254
 - training, 267
 - YOLO model, 292
 - Cohen’s kappa, 215
 - comparing models, 222
 - Compute Unified Device Architecture (CUDA), 9
 - confusion matrix, 206, 207
 - multiclass, 227
 - contingency table, 207
 - convolution, 239
 - exact, 242
 - image processing, 242
 - valid, 242
 - zero-padding, 242
 - convolutional neural network. *See* CNN
 - Cortes, Corinna, 83
 - cosine distance, 399
 - covariate shift, 355
 - Creative Solution Diagnosis Scale (CSDS), 536
 - Cropley, David, 536
 - CUDA (Compute Unified Device Architecture), 9
 - curse of dimensionality, 17, 19
 - curve fitting, 169
- ## D
- DAT (divergent association task), 532
 - data augmentation, 52, 171
 - approaches, 54
 - images, 60
 - rationale, 53
 - using PCA, 57
 - datasets, 7, 19
 - augmentation, 52
 - bagging, 83
 - balanced, 26
 - breast cancer dataset, 45
 - categorical value, 16
 - cautionary tales, 41
 - CIFAR-10 dataset, 50
 - CIFAR-100 dataset, 440
 - classes, 13
 - confusers, 23–24
 - curse of dimensionality, 17
 - discrete value, 15
 - ESC-10 dataset, 316
 - Fashion MNIST (FMNIST)
 - dataset, 485
 - feature selection, 17
 - feature vectors, 14–15
 - floating-point number, 15
 - hard negative, 24
 - ImageNet dataset, 354

- interval value, 15
- irises dataset, 43
- k*-fold validation, 36, 100
- label, 14
- M2NIST dataset, 422
- mean centering, 26
- mislabeled data, 38
- missing features, 29
- MNIST dataset, 47, 110
- normalizing, 27–28, 95
- one-hot encoding, 17
- ordinal value, 16
- outliers, 38, 41
- partitioning, 31–32, 35
- partitioning by class, 32
- pitch shifting, 317
- potential problems, 37
- preprocessing, 25, 43
- random sampling, 33
- sample, 4
- scaling, 25–26
- spectrogram, 332–333
- standardizing, 27–28, 95
- summarizing, 38
- test data, 30
- time shifting, 317
- training data, 30
- UC Merced Land Use
 - dataset, 409
 - validation data, 30
- DCGAN (deep convolutional GAN), 470
- decision tree, 123
- deep learning, xxviii
- depthwise convolution, 371
- derivative, 150
 - local minimum, 151
 - minimum, 150
 - partial, 158
 - tangent line, 149
- descriptive statistics, 7
- detection, 407
- divergent association task (DAT), 532
- domain adaptation, 22, 283
- dot product, 5
- dropout, 173
 - spatial dropout, 355
- dying ReLU problem, 365

E

- effective receptive field, 252–253
- ELM (extreme learning machine), 445
- embedding, 391, 515
- ensemble, 82
- epoch, 153
- Euclidean distance, 70, 399
- experiments
 - breast cancer, 95
 - CIFAR-10
 - analyzing models, 300
 - animal or vehicle, 302
 - binary versus multiclass, 307
 - building models, 297
 - irises, 89, 140
 - classical models, 90
 - MNIST, 110
 - activation function, 179
 - activation results, 183
 - architecture, 179
 - architecture results, 183
 - base learning rate, 190
 - batch size, 186
 - classical models, 110
 - code, 179
 - initialization, 197
 - L2 regularization, 193
 - momentum, 195
 - neural networks, 178
 - PCA (principal component analysis), 117
 - scrambled, 120, 201
 - training-set size, 192
 - MNIST CNN
 - basic experiments, 272
 - building models, 266
 - dataset, 264
 - epochs, 275
 - fully convolutional, 280
 - minibatches, 275
 - optimizers, 278
 - scrambled, 292
 - training-set size, 275
- extrapolation, 20
- extreme learning machine (ELM), 445

F

- F1 score, 214
- false-negative rate (FNR), 211
- false-positive rate (FPR), 211
- features, 14
- feature selection, 17
- feature types, 15
- feature vectors, 14–15
- fine-tuning, 380, 452, 461
- Fisher, R.A., 43
- Flesch, Rudolf, 536
- Flesch readability score, 536
- floating-point number, 15, 45, 320
- Fourier transform, 332
 - power spectrum, 332
- Frank, Eibe, 117
- fully connected layer, 255
- fully convolutional layer, 256
- fully convolutional network, 280

G

- Gaussian distribution, 8
- Gemma model, 512
- generative adversarial network (GAN), 463
 - conditional, 464, 480
 - discriminator, 464
 - generator, 464
 - latent vector, 464
 - mode collapse, 476
 - training, 467
 - unconditional, 464–465
- generative model, xxviii, 463
- Gini index, 82
- Glorot, Xavier, 198
- Goodfellow, Ian, 463
- GPT-4, 494, 505
- GPU (graphics processing unit), 1, 9
- gradient, 148
 - intercept, 149
 - slope, 149
- gradient descent, 148
 - algorithm, 152
 - batch training, 152
 - epoch, 153
 - first-order, 154
 - learning rate, 151, 190
 - local minimum, 151

- minibatch, 152–154
- momentum, 157, 195
- stochastic, 152

- grand mean, 103
- graph, 130
- graphics processing unit (GPU), 1, 9

H

- Hadamard, Jacques, 163
- Hadamard product, 163
- He, Kaiming, 168, 363
- heatmap, 288
- Hinton, Geoffrey E., 157, 255
- hit rate, 209
- hyperbolic tangent, 133
- hyperparameters, 104
- hyperplane, 84
- hypothesis testing, 8

I

- ID (intrinsic dimensionality), 448
- image processing, 242
- image retrieval, 399
- in-context learning, 504
- informedness, 213–214
- inner product, 5
- interpolation, 20, 488
- intersection over union (IoU), 414, 430
- interval value, 15
- intrinsic dimensionality (ID), 448

J

- joint probability, 74

K

- k-d-tree, 117
- Keras, 2
 - applications, 382
 - constructing CNNs with, 266
 - documentation, 265
 - functional API, 350
- kernel, 239
- k-fold validation, 36, 100
- Kibriya, Ashraf M., 117
- k-means, 395
- k-nearest neighbor (*k*-NN), 72–73
- k-NN classifier, 72–73, 99, 104, 114
- Krizhevsky, Alex, 173

L

- L2 regularization, 171–172
- label, 14
- large language model (LLM), 493
 - block diagram, 496
 - Claude 3, 508
 - embedding, 515
 - GPT-4, 494, 505
 - in-context learning, 504
 - Llama2, 510
 - MLP block, 501
 - Ollama, 494, 509
 - positional matrix, 498
 - quantization, 510
 - retrieval-augmented generation (RAG), 521
 - self-attention, 495
 - multihead, 499
 - semantic search, 519
 - temperature, 502
 - tokenization, 497
 - transformer, 499
- latent space, 473, 486
 - interpolation, 488
- layer normalization, 501
- learning rate, 151, 156, 190
- LeNet-5, 351
- linear algebra, 4. *See also* matrix; vector
 - scalar, 5
 - tensor, 6
- linear function, 132
- linear regression, 20
- Linux, 2
- Llama2 model, 510, 512
- LLaVa model, 512
- localization, 408
- local minimum, 151
- local response normalization (LRN), 354
- long short-term memory (LSTM), 495
- loss function, 148, 150, 164, 165, 267
 - absolute error, 165
 - cross-entropy loss, 166, 267
 - mean squared error (MSE), 165

M

- macOS, 3
- main function, 58

- Mann–Whitney U test, 8, 106
- markedness, 213
- matrix
 - multiplication, 6, 139
 - transpose, 143
 - vector multiplication, 5
- Matthews correlation coefficient (MCC), 215
 - multiclass, 227, 232
- max pooling, 254
- mean, 6
- median, 7
- metrics, 203
 - 2×2 table, 207
 - accuracy, 91, 204
 - accuracy matrix, 228
 - advanced metrics, 213
 - area under the ROC curve (AUC), 224
 - CIFAR-10 example, 304
 - Cohen’s kappa, 215
 - confusion matrix, 206–207
 - multiclass, 227
 - contingency table, 207
 - derived metrics, 208
 - F1 score, 214
 - false negative (FN), 206
 - false-negative rate (FNR), 211
 - false positive (FP), 206
 - false-positive rate (FPR), 211
 - grand mean, 103
 - hit rate, 209
 - implementation, 215
 - informedness, 213–214
 - interpreting models, 211
 - intersection over union (IoU), 414, 430
 - intrinsic dimensionality (ID), 448
 - markedness, 213
 - Matthews correlation coefficient (MCC), 215
 - multiclass, 227, 232
 - negative predictive value (NPV), 210
 - per-class accuracy, 205
 - positive predictive value (PPV), 209
 - precision, 209
 - precision-recall (PR) curve, 226

- metrics (*continued*)
 - purity, 394, 448
 - recall, 209
 - receiver operating characteristics (ROC), 217
 - elements of, 220
 - generating, 224
 - score, 91
 - sensitivity, 209
 - specificity, 209
 - true negative (TN), 206
 - true-negative rate (TNR), 209
 - true positive (TP), 206
 - true-positive rate (TPR), 209
 - t-SNE, 261, 393, 448
 - weighted accuracy, 230
 - Youden's *J* statistic, 213
- minibatch, 152–154, 186
- missing features, 29
- Mistral model, 512
- MobileNet, 373
- mode collapse, 476
- model
 - capacity, 24
 - classical models, 68
 - computational requirements, 125
 - explainability, 125
 - small datasets, 125
 - summary, 121
 - vector inputs, 125
 - when to use, 124
- Claude 3, 508
- comparing, 222
- DCGAN (deep convolutional GAN), 470
- decision tree, 77–78
 - construction, 81
 - summary, 123
- ensemble, 82, 375
- extreme learning machine (ELM), 445
- functional API, 353
- Gemma, 512
- generative adversarial network (GAN), 463
- Gini index, 82
- GPT-4, 494, 505
- hyperparameters, 104
 - optimizing, 104
- k*-means, 395
- LeNet5, 351
- Llama2, 510, 512
- LLaVa, 512
- long short-term memory (LSTM), 495
- Mistral, 512
- MobileNet, 373
- multiheaded, 408
- naive Bayes, 73–74, 114, 122
 - Gaussian, 91, 104, 114
 - multinomial, 91
 - summary, 122
- nearest centroid, 68, 71
 - summary, 121
- nearest neighbor, 4, 72
 - optimizing, 104
 - summary, 122
- Nomic, 512
- Ollama, 494
- one-class SVM, 397
- overfitting, 148, 168, 169
- PatchRot, 446
- random forest, 77, 82
 - optimizing, 105
 - summary, 123
- recurrent neural network (RNN), 495
- ResNet, 363
- RotNet, 441
- Siamese network, 454
- SimCLR, 462
- support vector machine (SVM), 83, 124
 - kernel, 86
 - margin, 84–85
 - optimizing, 86, 106
 - summary, 124
 - support vectors, 85
- template matching, 71
- U-net, 422
- VGG (Visual Geometry Group), 354
- YOLO, 420
- momentum, 157, 195
- multilabel classification, 432

N

- naive Bayes, 73–74, 114, 122
 - Gaussian, 91, 104, 114
 - multinomial, 91
 - summary, 122
- nearest centroid, 68–71, 121
- nearest neighbor, 4, 122
- negative predictive value (NPV), 210
- neural network, 130
 - activation function, 131–132
 - hyperbolic tangent, 133
 - identity, 137
 - ReLU, 134
 - sigmoid, 133
 - anatomy, 130
 - ANN (artificial neural network), 132
 - architecture, 135
 - backpropagation, 148, 157
 - backward pass, 158
 - bias, 132
 - essence, 132
 - feedback, 130
 - forward pass, 158
 - fully connected feed-forward,
 - 129, 130
 - function approximation, 130
 - hidden layer, 130
 - implementation, 140–141
 - multilayer perceptron (MLP), 132
 - neuron, 130–131
 - output layer, 132, 137
 - regularization, 148, 168, 171
 - dropout, 173
 - L2, 171–172
 - weight decay, 172, 193
 - representation, 139
 - rules of thumb, 136
 - softmax, 137
 - traditional, 129
 - training, 175
 - batch, 152
 - epoch, 153
 - gradient descent, 148
 - loss function, 148, 150,
 - 164, 165
 - minibatch, 152, 154
 - momentum, 157
 - overview, 148

- stochastic gradient descent (SGD), 152
 - stopping, 155
 - weight update, 151
- weight initialization, 148, 167, 197
- weights, 131
- Nomic model, 512
- nominal value, 16
- nonlinear function, 132
 - transcendental, 133
 - trigonometric, 133
- nonparametric test, 8
- normal distribution, 7, 8
- normalizing, 27–28, 95
- NPV (negative predictive value), 210
- NumPy, 2
 - argsort function, 33
 - convert arrays to images, 63
 - convert images to arrays, 63
 - histogram, 306
 - pseudorandom seed, 98
 - where function, 32
- NVIDIA, 9

O

- Ollama, 494, 509
- Olson, Jay, 532
- one-class SVM, 397
- one-hot encoding, 17
- operating environment, 1
- ordinal value, 16
- outer product, 5
- outliers, 41
- overfitting, 53, 148, 155, 168, 169

P

- parametric test, 8
- parent distribution, 7, 22, 204
- partial derivative, 158
- Pasteur, Louis, 126
- PatchRot model, 446
- PCA (principal component analysis),
 - 55, 113, 117
- Pillow library, 3
- Plato, 7, 22
- pointwise convolution, 371
- pooling, 254
- positional matrix, 498

- positive predictive value (PPV), 209
- Powers, David Martin, 214
- power spectrum, 332
- precision, 209
- precision-recall (PR) curve, 226
- preprocessing, 43
- primary visual cortex, 252
- principal component analysis (PCA), 55, 113, 117
- prior class probability, 22, 204
- probability, 6
 - Bayes' theorem, 73
 - distribution, 7
 - Gaussian distribution, 8
 - normal distribution, 7, 8, 76
 - parent distribution, 7, 22
 - uniform distribution, 7, 8, 167
 - joint probability, 74
 - likelihood, 73
 - posterior probability, 73
 - prior class probability, 22
 - prior probability, 74
- probability map, 288
- pseudorandom sequence, 99
- purity, 394, 448
- p-value, 8
- Python
 - libraries
 - Pillow, 3
 - scikit-dimension (skdim), 449
 - scikit-learn (sklearn), 2
 - SciPy, 318
 - modules
 - librosa, 318
 - pickle, 142

Q

- quantization, 510

R

- RAG (retrieval-augmented generation), 521
- random forest, 82, 123
- recall, 209
- receiver operating characteristics (ROC), 217, 220–226
- receptive field, 252–253
- rectified linear unit (ReLU), 134
 - dying ReLU problem, 365
 - leaky, 465

- recurrent neural network (RNN), 495
- recursion, 80
- regularization, 53, 148, 168, 171
 - dropout, 173
 - L2, 171–172, 193
 - weight decay, 172, 193
- reinforcement learning, xxviii
- ResNet model, 363
 - architecture, 366
- retrieval-augmented generation (RAG), 521
- RMSprop, 278
- RNN (recurrent neural network), 495
- ROC (receiver operating characteristics), 217, 220–226
- Ronneberger, Olaf, 258
- RotNet model, 441
- Rumelhart, David E., 157

S

- sample, 4
- scalar, 5
- scikit-dimension (skdim) library, 449
- scikit-learn (sklearn) library, 2
 - DecisionTreeClassifier class, 90
 - GaussianNB class, 90
 - KNeighborsClassifier class, 90
 - make_classification function, 32
 - MLPClassifier class, 132, 144, 178
 - MultinomialNB class, 90
 - NearestCentroid class, 90
 - RandomForestClassifier class, 90
 - SVC class, 90
- SciPy library, 318
 - wavfile module, 318
- self-attention, 495, 499
 - multihead, 499
- self-supervised learning, xxviii, 439
- semantic search, 519
- semantic segmentation, 407
- sensitivity, 209
- sentiment analysis, 523
- SGD (stochastic gradient descent), 152, 180, 267, 278, 296
- Siamese network, 454
- sigmoid, 133
- SimCLR, 462
- Simony, Karen, 354
- skip connection, 363
- softmax, 137

- spatial dropout, 355
- specificity, 209
- spectrogram, 333
- standard deviation, 7–8, 27, 76, 167
- standard error, 6
- standardizing, 27–28, 95
- statistically significant, 8
- statistical test, 8
- statistics, 6
 - bootstrap sample, 83
 - descriptive, 6–7
 - grand mean, 103
 - hypothesis testing, 8
 - Mann–Whitney U test, 8, 106
 - mean, 6
 - median, 7
 - nonparametric test, 8
 - parametric test, 8
 - p-value, 8
 - quartile, 40
 - standard deviation, 7–8, 27, 76, 167
 - standard error, 6, 39
 - statistically independent, 74
 - statistically significant, 8
 - t-test, 8
 - variance, 7
- stochastic gradient descent (SGD), 152, 180, 267, 278, 296
- supervised learning, xxviii, 19
- support vector machine (SVM), 83, 124
 - kernel, 86
 - margin, 84–85
 - one-versus-one, 115
 - one-versus-rest, 115
 - optimizing, 86, 106
- support vectors, 85

T

- tangent line, 149
- temperature, 502
- tensor, 6
- TensorFlow, 2
- test data, 30
- tokenization, 497
- toolkits, 2–3
- topic modeling, 523
- training data, 30
- training set, xxviii
- transcendental functions, 133

- transfer learning, 391
- transformer, 499
- trigonometric functions, 133
- trilobite, 421
- true-negative rate (TNR), 209
- true-positive rate (TPR), 209
- t-SNE, 261, 393, 448
- t-test, 8

U

- Ubuntu, 1, 2–3
- UCI Machine Learning Repository, 45
- U-net, 422
- uniform distribution, 7–8, 167
- unsupervised learning, xxviii

V

- validation data, 30
- vanishing gradient problem, 364
- Vapnik, Vladimir, 83
- variance, 7
- Vaswani, Ashish, 494
- vector, 4
 - column vector, 4
 - dot product, 5
 - inner product, 5
 - multiplication, 5
 - outer product, 5
 - row vector, 4
- VGG (Visual Geometry Group), 354

W

- weight decay, 172, 193
- weighted accuracy, 230
- weight initialization, 148, 167, 197
 - Glorot, 168
 - He, 168
 - Xavier, 168
- weights, 131
- weight update, 151
- Windows, 3

Y

- YOLO model, 292, 420
- Youden's *J* statistic, 213

Z

- zero-padding, 242
- Zisserman, Andrew, 354