

# Index

---

*Note:* Page numbers followed by “*f*” indicate figures; and “*t*” indicate tables; and “*b*” indicate boxes.

## A

- ABIDE. *See* Autism Brain Imaging Data Exchange (ABIDE)
- Accountability, 332–333  
gap, 332–333, 333f
- Accuracy, 37–38, 313  
score, 238, 238t
- Activation functions, 159–160, 159f
- Activity space, 268–269
- AD. *See* Alzheimer’s disease (AD)
- ADHD. *See* Attention-deficit/hyperactivity disorder (ADHD)
- ADNI. *See* Alzheimer’s Disease Neuroimaging Initiative (ADNI)
- AE. *See* Autoencoders (AE)
- Agglomerative clustering (AG clustering), 240–241
- Aggregated voxels, 319
- Aggregating features, 314–315
- Aggression prediction in ADHD, 135–136
- AlexNet, 173
- AlphaGo model, 174
- Alzheimer’s disease (AD), 4, 46, 51, 74, 109, 148–149, 166, 184, 215–216, 283–284, 330–331  
diagnostic classification, 149–150  
early detection in, 167–168  
multimodal neuroimaging studies with deep learning in, 299–301  
predicting cognitive decline in, 76–77  
predicting conversion from mild cognitive impairment to, 110–112  
prediction, 184–185
- Alzheimer’s Disease Neuroimaging Initiative (ADNI), 58, 149, 220, 250–251
- Anaconda (programming language), 345
- Analysis of variance (ANOVA), 25–27, 242  
RB kernels, 129
- Aphasia. *See* Language impairment
- Area under curve (AUC), 38–39
- Area under ROC curve (AUC-ROC), 38–39
- Artificial intelligence, 1–2, 4
- Artificial neural networks, 3
- ASD. *See* Autism spectrum disorder (ASD)
- Asymmetric data fusion, 287
- Atlas-based ROIs, 320–321
- Attention-deficit/hyperactivity disorder (ADHD), 124, 135–136  
aggression prediction in, 135–136
- AUC. *See* Area under curve (AUC)
- AUC-ROC. *See* Area under ROC curve (AUC-ROC)
- Augmentation  
data, 259  
simulation-based, 259–260
- Autism Brain Imaging Data Exchange (ABIDE), 133–134, 250–251  
ABIDE I, 204
- Autism spectrum disorder (ASD), 84, 166, 168–169, 203  
classification of, 93  
diagnosis of, 203–205  
predicting depressive states, 205–206  
predicting symptoms severity and mathematical abilities, 133–135  
pretraining and fine-tuning process, 204f
- Autoencoders (AE), 24–25, 193–194, 299  
adversarial, 198–201  
network, 199f  
representations in latent space, 198f  
structure, 200f  
application to brain disorders, 201–206  
denoising, 196–197, 197f  
method description, 194–201  
sparse, 197–198  
structure of, 195f  
undercomplete, 196
- Automated feature reduction methods, 314–315
- Automated feature selection, 25–28, 26t  
embedded methods, 27–28  
filter methods, 25–27  
selection methods, 34  
wrapper methods, 27

**B**

- bac. *See* Balanced accuracy (bac)
- Backpropagation, 161
- Balanced accuracy (bac), 37–38, 101, 358
- Bayes rule, 89
- Bayesian logistic regression, 89, 93
- Bayesian variant of sparse logistic regression, 89–90
- Behaviorism, 2–3
- Bias, 308–309
  - problem of, 337–338
  - sources in machine learning, 309–311
    - choice of features, 310–311
    - choice of machine learning algorithm, 311
    - choice of target variable, 309–310
    - data processing, 311
- Bias–variance trade-off, 34–35
- Big data, 7–8
- Bilingualism, 77–78
- Binary labels, 85
- Biological age score, 221
- Biomarker, 51, 329–330
  - absence of, 51
- Biomedical data, ontology of, 335
- Bipolar disorder, 94
  - data fusion in, 298
- Black box, 165
- Blind source separation problem, 287–288, 288f
- Block-wise missing data, 150
- Blood oxygen level dependent activity (BOLD activity), 267–268
- Boolean class labels, 109
- Brain aging, 221
  - in schizophrenia, 132–133
- Brain atrophy, 75
- Brain disorders, 283–284, 329–330
  - applications in clustering analysis, 240–243
    - functional connectivity states investigation in fMRI, 242–243
    - identification of cross-diagnostic neurocognitive profiles, 242
    - identifying disorder subtypes, 240–242
  - applications to, 318–323
    - using clinical variables for prognostic prediction, 322–323
  - in CNNs, 184–188
  - language as marker of psychiatric and developmental disorders, 321–322
    - of machine learning to, 330–331
    - in multimodal integration, 296–301
    - using neuroimaging to predicting brain age, 319–321
    - in voxel-wise encoding models, 278–279
- autoencoders application to, 201–206
  - predicting prognosis from EHRs, 201–203
- challenges in machine learning studies of psychiatric and neurological disorders, 50–55
  - absence of biomarkers, 51
  - heterogeneity, 53–55
  - reliability of diagnosis, 52–53
- cost of misclassification, 56, 57f
- DNN applications to, 166–169
  - autism spectrum disorders, 168–169
  - diagnostic classification in schizophrenia, 166–167
  - early detection in Alzheimer’s disease, 167–168
- ethical tensions from using machine learning in, 331–338
  - challenges in interaction between humans and intelligent systems, 332–333
- data security and patient privacy, 334–335
- interpretability of machine learning and problem of bias, 337–338
- mental privacy and question of “neurorights”, 335–336
- ontology of biomedical data, 335
- transparency of machine learning, 336–337
- linear classification model applications, 91–96
  - classification of ASD, 93
  - classification of schizophrenia, 94–96
- linear regression applications, 74–79
- machine learning applications, 45
- machine learning relevant to
  - data-driven models and big data, 7–8
  - focus on prediction and generalizability, 6–7
  - group-level to individual-level inferences, 5
  - recognizing heterogeneity, 8
  - univariate to multivariate inferences, 5–6

- MKL applications, 148–153  
 AD, diagnostic classification of, 149–150  
 PD, diagnostic classification of, 150–151  
 psychosis, diagnostics classification of, 151–153  
 optimism, 55  
 PCA applications to, 218–221  
 healthy aging and disease process, 221  
 neurological disorders, 219–221  
 psychiatric disorders, 218–219  
 reason for people interested in machine learning, 45–50  
 diagnostic evaluation, 48–49  
 prediction of illness onset, 46–48  
 prediction of outcomes, 49–50  
 result of machine learning applied in psychiatry and neurology, 57–59  
 SVM applications to, 109–116  
 brain-based diagnostics of schizophrenia, 112–115  
 predicting conversion from MCI to AD, 110–112  
 predicting treatment response in depression, 115–116  
 SVR applications to, 131–136  
 aggression prediction in ADHD, 135–136  
 predicting deviations from “brain age” in schizophrenia, 132–133  
 predicting symptoms severity and mathematical abilities in ASD, 133–135  
 Brain imaging machine learning, 318  
 Brain morphometry to classifying patients with SZ and HC, 345–347  
 Brain tumor segmentation, 185–187  
 Brain volume measures, 75  
 BRATS. *See* Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)  
 By-chance heterogeneity, 260
- C**  
 Camberwell Assessment of Need Short Appraisal Schedule (CANSAS), 322–323  
 Canonical correlation analysis (CCA), 289–290, 290f  
 Canonical correlation coefficient (CCC), 289–290  
 Canonical variate, 289–290  
 Canonical variate analysis (CVA), 214, 215t  
 CANSAS. *See* Camberwell Assessment of Need Short Appraisal Schedule (CANSAS)  
 Care, 90–91  
 CCA. *See* Canonical correlation analysis (CCA)  
 CCC. *See* Canonical correlation coefficient (CCC)  
 CDI. *See* Communicative Development Inventories (CDI)  
 Centroid-based clustering, 229  
 Cerebrospinal fluid (CSF), 148–149  
 Classes, 10  
 Classical statistics, 6–7  
 statistical methods, 90–91  
 Classification, 101  
 algorithms, 10  
 function, 86  
 threshold, 38–39  
 Classifier, 101  
 Clinical and behavioral scores, 141–142  
 Clinical neuroimaging, 84, 150  
 Clinical translation, 47–48  
 Cluster validation, 230–234  
 Clustering analysis, 11, 227  
 applications to brain disorders, 240–243  
 choosing  $K$  with silhouette scores, 232–233  
 with density-based spatial clustering, 237f  
 $K$ -means, 230–240  
 multipeak and uniform distributions, 228f  
 SSE, 231–232  
 testing for multipeak distribution, 233–234  
 CNNs. *See* Convolutional neural networks (CNNs)  
 “Cocktail party” problem, 287–288  
 Cognitive impairment, 148–149  
 Comma-separated values (CSV), 349  
 Communicative Development Inventories (CDI), 322  
 Competitive approach, 146–147  
 Complete case analysis/listwise deletion, 255  
 Computational models, 3  
 Computed tomography (CT), 259–260  
 Confidence intervals, 259

Confounding variable, 23  
 Confounds, 261–264  
 Confusion matrix, 36–37, 37f  
 Control-based harmonization model, 262  
 Convergence to local minima, 232  
 ConvNets. *See* Convolutional neural networks (CNNs)  
 Convolution operation, 177, 178f  
 Convolutional neural networks (CNNs), 173, 299, 321  
   applications to brain disorders, 184–188  
   architecture, 182–183  
   convolutional layer, 176–181  
     filter count, 180  
     filter size, 180  
     padding, 181  
     stride size, 180, 180f  
   flexibility, 186–187  
   fully connected layer for, 182  
   layers, 176–182  
   method description, 175–183  
   structure, 174f  
 Cooperative approach, 146–147  
 Coordinate descent with early stopping, 276  
 Core object, 237–238  
 Core observations, 237–238  
 Cortical surface, 319–320  
 Cost of misclassification, 56, 57f  
 Covariance matrices, 212–213  
 Cross-diagnostic neurocognitive profiles, identification of, 242  
 Cross-validation (CV), 29–33, 106, 141–142, 345–347, 357–359  
   feature selection scheme, 113–115  
   final performance in, 36f  
   nested, 31–32  
   parameter selection using, 131  
   performance, 41  
   stratified, 33  
   types, 32f  
 CSF. *See* Cerebrospinal fluid (CSF)  
 CSV. *See* Comma-separated values (CSV)  
 CT. *See* Computed tomography (CT)  
 Curse of dimensionality, 72–73, 243  
 CV. *See* Cross-validation (CV)  
 CVA. *See* Canonical variate analysis (CVA)

## D

### Data

  augmentation, 259

  data-driven models, 7–8  
   data-driven unsupervised techniques, 314–315  
   efficient learning, 260  
   fusion, 285–291  
     in schizophrenia and bipolar disorder, 298  
   harmonization, 262  
   prediction of, 6–7  
   preparation, 23  
   security, 334–335  
   transformations, 34  
     implementing in machine learning pipeline, 34  
 Data processing, 311–318  
   addressing high dimensionality, 314–318  
   feature aggregation, 315–318  
   unsupervised methods, 318  
   measuring separability between two groups, 313–314  
 Dataframe, 349  
 DBSCAN. *See* Density-based spatial clustering of applications with noise (DBSCAN)  
 DD. *See* Developmental dyslexia (DD)  
 Decision-support system (DSS), 332–333  
 Decoder, 194–195  
 Deep learning (DL), 3, 157–158, 273, 285–286, 293–294, 321, 336–337  
   approaches, 257  
   encoding models from Gabor feature space to, 271–275  
   multimodal neuroimaging studies with deep learning in AD, 299–301  
 Deep neural networks (DNNs), 158, 161–162, 166, 175, 194, 299. *See also* Convolutional neural networks (CNNs)  
   applications to brain disorders, 166–169  
   hyperparameter tuning, 165–166  
   interpretability, 165  
   learning rate on weight optimization, 164f  
   loss landscape, 162f  
   overfitting, 164  
   starting weights, multiple local minima, and saddle points, 163–164  
   structure, 159–160, 159f  
   training, 160–164  
 Deepnet-fwRF models, 273, 274f–275f  
 Dendrogram, 240–241

- Density-based spatial clustering of applications with noise (DBSCAN), 236–238  
 comparison of  $K$ -means with, 238–240  
 Density-reachable points, 237–238  
 Dependent variable, 22–23  
 Depression, predicting treatment response in, 115–116  
 Developmental dyslexia (DD), 322  
 Diagnostic and Statistical Manual of Mental Disorders (DSM), 52, 309  
 DSM-5, 52  
 Diagnostic labels, 52–53  
 labeling method, 309  
 Diagonal matrix, 213–214  
 Diffusion MRI, 131–132  
 Diffusion tensor imaging (DTI), 297  
 Dimension, 22–23  
 reduction, 213–214  
 Dimensionality, 22–23, 196  
 curses of, 72–73, 243  
 of image data, 87  
 reduction, 12, 24–25, 34  
 Dip test, 233–234  
 Discriminative mapping, 84  
 Discriminator, 199  
 Disease progression, 49  
 DL. *See* Deep learning (DL)  
 DNNs. *See* Deep neural networks (DNNs)  
 Dot products, 177  
 Double dipping, 34  
 DSM. *See* Diagnostic and Statistical Manual of Mental Disorders (DSM)  
 DSS. *See* Decision-support system (DSS)  
 DTI. *See* Diffusion tensor imaging (DTI)
- E**  
 “Earlier IQ”, 71–73  
 Early multimodal data integration, 285, 287–291. *See also* Intermediate multimodal integration; Late multimodal integration  
 jICA, 287–289  
 mCCA, 289–290  
 mCCA + jICA, 290–291  
 EDSS. *See* Expanded Disability Status Scale (EDSS)  
 EEG. *See* Electroencephalography (EEG)  
 Effect sizes, 67–68  
 eHealth technology. *See* Electronic health technology (eHealth technology)  
 EHRs. *See* Electronic health records (EHRs)  
 Elastic net, 88  
 Electroencephalography (EEG), 141–142, 184, 283–284, 307–308, 329–330  
 Electronic health records (EHRs), 201–202  
 deep patient model, 202f  
 diagnosis of ASD, 203–205  
 predicting prognosis from, 201–203  
 Electronic health technology (eHealth technology), 334–335  
 Electrophysiological methods, 329–330  
 Embedded methods, 27–28, 105, 258  
 Empirical risk  
 functional, 87  
 minimization, 85–87  
 loss functions, 86f  
 Encoder, 194–195  
 Encoding models from Gabor feature space, 271–275  
 Enhancing Neuroimaging Genetics through Meta-analysis (ENIGMA), 250–251  
 ENIGMA. *See* Enhancing Neuroimaging Genetics through Meta-analysis (ENIGMA)  
 ENIGMA Bipolar Disorder working group (ENIGMA-BP), 58  
 ENIGMA-BP. *See* ENIGMA Bipolar Disorder working group (ENIGMA-BP)  
 Ensemble feature selection, 113–115  
 Ensemble methods, 294–296  
 $\epsilon$ -SVR model, 125  
 Euclidian distances, 243  
 Event-related potentials, 319  
 Expanded Disability Status Scale (EDSS), 75  
 “Expected risk”, 86  
 Explanatory variables. *See* Feature set
- F**  
 False negatives (FN), 36  
 False positives (FP), 36  
 Family risk (FR), 322  
 FC matrices. *See* Functional connectivity matrices (FC matrices)  
 Feature engineering, 21–28, 157–158, 193, 202–203  
 dimensionality reduction, 24–25  
 feature extraction, 24  
 feature selection, 25–28  
 Feature scaling, 360  
 feature scaling/normalization, 28

Feature selection, 25–28, 105–106, 229, 359–360  
 automated, 25–28, 26t  
 embedded methods, 105  
 feature scaling/normalization, 28  
 feature selection/transformation, 258  
 filter methods, 105–106  
 manual, 25  
 method, 146–147  
 wrapper methods, 106  
 Feature-weighted receptive field model (fwRF model), 277, 278f  
 Features, 101–102  
 aggregation, 315–318  
 extraction, 24, 357  
 hierarchical features, 175  
 map, 177, 273  
 reduction using ROI, 112b  
 set, 21–23  
 vector, 22–23  
 FEP. *See* First episode of psychosis (FEP)  
 Filter, 177  
 count, 180  
 methods, 25–27, 105–106, 258  
 size, 180  
 sliding, 177  
 First episode of psychosis (FEP), 152–153  
 First principal axis, 210  
 Flatness, 126  
 fMRI. *See* Functional magnetic resonance imaging (fMRI)  
 FN. *See* False negatives (FN)  
 “For loop”, 359, 365  
 Forward propagation, 160  
 FP. *See* False positives (FP); Frontal pole (FP)  
 FR. *See* Family risk (FR)  
 Fractional anisotropy, 107  
 FreeSurfer, 345–347  
 Frontal pole (FP), 135–136  
 Functional connectivity matrices (FC matrices), 166  
 Functional magnetic resonance imaging (fMRI), 108–109, 131–132, 141–142, 240, 267–268, 285–286, 319  
 functional connectivity states  
 investigation in, 242–243  
 Functioning, 11  
 fwRF model. *See* Feature-weighted receptive field model (fwRF model)

## G

Gabor feature space to deep learning networks, encoding models from, 271–275  
 Gabor wavelets, 271–272  
 Gabor-fwRF models, 273, 274f–275f  
 GAINs. *See* Generative adversarial imputation networks (GAINs)  
 GANs. *See* Generative adversarial networks (GANs)  
 Gaussian mixture model (GMM), 236–237  
 comparison of *K*-means with, 238–240  
 Gaussian noise, 268–269  
 Gaussian prior, 89–90  
 GD. *See* Gradient descendant (GD)  
 General linear model (GLM), 112b  
 General linear regression, 70–71  
 Generalizability, 6–7  
 Generalized linear model, 147–148  
 Generative adversarial imputation networks (GAINs), 257  
 Generative adversarial networks (GANs), 257, 259–260  
 Generative modelling, 194  
 Generator, 199  
 Genetic and epigenetic information, 329–330  
 Genomic data, 141–142  
 GL-MKL. *See* Group lasso MKL (GL-MKL)  
 GLM. *See* General linear model (GLM)  
 Global minima, 161–163  
 GM-f model, 75  
 GMM. *See* Gaussian mixture model (GMM)  
 Goodness of fit, 21  
 Gradient, 161–162  
 Gradient descendant (GD), 160, 163–164  
 Graduate Record Examinations (GRE), 67  
 Grid search, 277–278  
 GridSearchCV, 362  
 Group convolution network, 260  
 Group lasso MKL (GL-MKL), 146–147  
 Group-level statistics, 25–27  
 Group-level to individual-level inferences, 5  
 Grouping criterion, 229

## H

Hard clustering, 229  
 HCs. *See* Healthy controls (HCs)  
 Healthy aging, 218, 221

Healthy controls (HCs), 106–107, 143–144, 217, 345–347  
 using brain morphometry to classifying patients with, 345–347  
 Heterogeneity, 53–55, 249–251  
 algorithms and procedures, 260–262  
 data simulation, 251–253  
 recognizing, 8  
 High-dimensional data, 160, 267  
 Hold-out method, 29  
 Homogeneity score, 238, 238t  
 Human learning, 1–2  
 to machine learning, 2–4  
 Hyperbolic tangent activation, 159–160  
 Hyperparameters, 31, 107  
 optimization, 360–362  
 tuning, 32, 160, 165–166  
 Hyperplane, 101–103, 102f, 106–107, 109, 125–126

## I

ICA. *See* Independent component analysis (ICA)  
 ICD. *See* International Classification of Diseases (ICD)  
 ICs. *See* Independent components (ICs)  
 ICV. *See* Intracranial volume (ICV)  
 Illness onset prediction, 46–48  
 ImageNet competition. *See* “ImageNet Large Scale Visual Recognition Challenge” event  
 “ImageNet Large Scale Visual Recognition Challenge” event, 173  
 Imputation methods, 253–254, 257  
 “In-sample” analyses, 73  
 Indentation, 359  
 Independent component analysis (ICA), 24–25, 152, 242–243, 287–288  
 Independent components (ICs), 288  
 Independent variables. *See* Feature set  
 Inference stage, 89–90  
 Inner CV, 32, 345–347  
 Input space, 268–269  
 Intelligence, 68  
 Intelligence quotient (IQ), 67–68  
 explanation of, 73–74  
 prediction of, 73–74  
 test, 70  
 Intercept, 69

Intermediate multimodal integration, 285, 291–294. *See also* Early multimodal data integration; Late multimodal integration  
 deep learning, 293–294  
 kernel-based methods, 291–292  
 Intermediate-level approaches, 256–257  
 International Classification of Diseases (ICD), 52, 309  
 Interpretability, 165, 311–313, 318–320  
 Interpretability of machine learning, 337–338  
 Interpretation, 216–217  
 Alzheimer’s disease pattern, 217f  
 Interrater reliability, 52  
 Intracranial volume (ICV), 75  
 IQ. *See* Intelligence quotient (IQ)  
 isnull() function, 351

## J

Joint independent component analysis (jICA), 287–289, 289f

## K

*k*-fold CV, 29–31  
*K*-means, 230–240, 242–243  
 algorithm, 230  
 cluster validation, 230–234  
*K*-nearest neighbours (KNNs), 212, 256–257  
 Kernel, 141–142  
 functions, 143–144  
 kernel-based data integration  
 approaches, 285–286  
 kernel-based machine learning  
 algorithms, 141–142  
 kernel-based methods, 291–292  
 MKL, 292  
 unweighted simple sum of kernels, 292  
 linear, 141–142  
 methods, 104  
 space, 127–128  
 SVR, 127–129  
 trick, 71, 91, 105, 127–128, 143–144  
 types, 150–151  
 KNNs. *See* *K*-nearest neighbours (KNNs)  
 Knowledge leakage, 345–347

## L

Labels, 10

- Language as marker of psychiatric and developmental disorders, 321–322
- Language impairment, 74, 77  
prediction after stroke, 77–79
- Lasso penalty, 88
- Late multimodal integration, 285, 294–296.  
*See also* Early multimodal data integration; Intermediate multimodal integration  
majority voting, 296  
soft voting, 295–296
- “Later IQ”, 71–73
- Lateral orbitofrontal (IOFC), 135–136
- Laterality index (LI), 318
- Layer-wise structure, 159–160
- Learning  
rate, 161–162  
decay, 161–162  
types, 9
- Leave-one-out cross-validation approach (LOOCV approach), 29–31, 30f, 77–78
- Leave-one-site-out procedure (LOSO procedure), 94
- Leave-one-subject-out cross-validation, 134–135
- LI. *See* Laterality index (LI)
- Linear classification models, 83–84  
applications to brain disorders, 91–96  
empirical risk minimization, 85–87  
extensions to nonlinear models, 91  
fundamentals and notation, 85  
mapping discriminating pattern, 90–91  
optimization, 88–89  
probabilistic classification, 89–90  
regularization, 87–88, 87t
- Linear classifier, 102–103
- Linear discriminant analysis, 83–84
- Linear Gaussian process models, 83–84
- Linear kernel, 141–142
- Linear MKL combination algorithms, 145–147
- Linear regression analysis, 67, 123–124, 268–269  
applications to brain disorders, 74–79  
predicting cognitive decline in AD, 76–77  
predicting disease progression in MS, 74–76  
predicting language impairment after stroke, 77–79  
general linear regression, 70–71  
limitations, 72–73  
multiple regression, 71–72  
prediction *vs.* explanation, 73–74  
scatter plots of earlier IQ, 68f  
simple regression, 69–70
- Linear SVM, 83–84, 102–103
- Linear SVR model, 135–136
- Linear  $\epsilon$ -SVR model, 126–127
- Linear  $\nu$ -SVR algorithm, 132–133
- Local connectivity, 174–175
- Local kernel approach, 147–148
- Local minima, 163
- IOFC. *See* Lateral orbitofrontal (IOFC)
- Log loss, 85–86
- Logistic activation function, 86
- Logistic regression, 85–86, 141–142
- LOOCV approach. *See* Leave-one-out cross-validation approach (LOOCV approach)
- LOSO procedure. *See* Leave-one-site-out procedure (LOSO procedure)
- Loss function, 85–86, 160–161, 268–269
- Loss landscape, 161–162
- Low signal-to-noise ratio, 258–259
- Lower-dimensional space, 196
- Lp-norm MKL, 146–147
- ## M
- Machine learning, 1–2, 4–5, 21, 101, 231, 267, 329–330, 343  
algorithms, 107, 360–362  
applications to brain disorders, 330–331  
data preparation, 23  
diagnostic evaluation, 48  
ethical tensions from using machine learning in brain disorders, 331–338  
feature engineering, 23–28  
human learning to, 2–4  
machine learning–based decision support system, 55–56  
methods, 123–124, 157–158, 167, 307–308  
models, 249  
evaluation, 36–41  
training, 28–35  
pipeline implementing in tutorial, 346f  
post hoc analysis, 41–42  
problem formulation, 22–23  
relevant to brain disorders, 5–8  
steps of standard supervised machine learning pipeline, 22f  
terminology in, 22f  
tools, 210–212



- types, 9–13, 9f
  - reinforcement learning, 13
  - semisupervised learning, 12
  - supervised learning, 9–11
  - unsupervised learning, 11–12
- MAE. *See* Mean absolute error (MAE)
- Magnetic resonance imaging (MRI),
  - 22–23, 47, 72, 85, 124, 176, 218,
  - 257–259, 307–308, 329–330,
  - 345–347
- acquisition techniques, 131–132
- Major depressive disorder (MDD),
  - 106–107
- Manhattan distances, 243
- MANOVA. *See* Multivariate analysis of variance (MANOVA)
- Manual feature selection, 25
- Mapping discriminating pattern, 90–91
- MAR. *See* Missing at random (MAR)
- Markov chain Monte Carlo methods,
  - 89–90
- Mass-univariate methods, 5–6
- Mathematics, 1–2
- Matplotlib, 344–345, 347–348, 352–353
- Max pooling layer, 182, 182f
- Maximum margin principle, 143–144
- Maximum mean discrepancy (MMD), 150
- MCAR. *See* Missing completely at random (MCAR)
- mCCA. *See* Multimodal canonical correlation analysis (mCCA)
- MCCA. *See* Multiset canonical correlation analysis (MCCA)
- mCCAR. *See* Multimodal canonical correlation analysis with reference (mCCAR)
- MCI. *See* Mild cognitive impairment (MCI)
- MDD. *See* Major depressive disorder (MDD)
- Mean absolute error (MAE), 40
- Mean rule, 145
- Mean squared error (MSE), 40
- Mean/median/mode imputation, 256
- Measurements, 309–313
  - instruments, 307–308
  - suboptimal, 310–311
- Medial orbitofrontal (mOFC), 135–136
- Medial prefrontal cortex (mPFC), 135–136
- Medial temporal lobe (MTL), 216
- Mental imagery, 279
- Mental privacy, 335–336
- mHealth technology. *See* Mobile health technology (mHealth technology)
- Mild cognitive impairment (MCI), 4,
  - 46–47, 76, 110–111, 149, 168, 299
- to AD, 110–112
  - A-priori regions of interest, 111f
  - feature reduction using ROI, 112b
- Min-max scaling, 28
- Mini-Mental State Examination (MMSE), 76
- Missing at random (MAR), 254
- Missing completely at random (MCAR),
  - 254
- Missing data, 249, 251
  - algorithms and procedures, 253–257
  - mechanism and model, 254
  - simulation, 251–253
- Missing not at random (MNAR), 254
- MKL. *See* Multi-kernel learning (MKL); Multiple kernel learning (MKL)
- MLM. *See* Multivariate linear models (MLM)
- MMD. *See* Maximum mean discrepancy (MMD)
- MMSE. *See* Mini-Mental State Examination (MMSE)
- MNAR. *See* Missing not at random (MNAR)
- MNI. *See* Montreal Neurological Institute (MNI)
- Mobile health technology (mHealth technology), 334–335
- Mobility measures, 334–335
- Model
  - averaging, 259
  - class, 268–269
  - coefficients, 362
  - fitting, 272–273
  - selection, 32
- Model evaluation, 36–41
  - classification, 36–40
    - accuracy, 37–38
    - multiclass classification, 39–40
    - ROC and area under curve, 38–39
    - sensitivity/true positive rate/recall, 38
    - specificity/true negative rate, 38
  - regression, 40–41
- Model training, 28–35
  - bias–variance trade-off, 34–35
  - cross-validation, 29–33
  - implementing data transformations in machine learning pipeline, 34
  - training and test sets, 29

- mOFC. *See* Medial orbitofrontal (mOFC)
- Montreal Neurological Institute (MNI), 215–216
- mPFC. *See* Medial prefrontal cortex (mPFC)
- MRI. *See* Magnetic resonance imaging (MRI)
- MS. *See* Multiple sclerosis (MS)
- MSE. *See* Mean squared error (MSE)
- MTL. *See* Medial temporal lobe (MTL)
- Multiclass classification, 39–40
- Multicollinearity, 72
- Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), 186
- Multimodal canonical correlation analysis (mCCA), 287, 289–290, 291f  
mCCA + jICA, 290–291, 292f
- Multimodal canonical correlation analysis with reference (mCCAR), 298
- Multimodal diagnostic system, 113–115
- Multimodal integration, 283–284, 286f  
application to brain disorders, 296–301  
data fusion in schizophrenia and bipolar disorder, 298  
multimodal neuroimaging studies with deep learning in AD, 299–301  
multimodal studies with kernel-based and ensemble methods, 297–298  
early multimodal data integration, 287–291  
intermediate multimodal integration, 291–294  
late multimodal integration, 294–296
- Multimodal machine learning studies, 286–287
- Multimodality approach, 46–47
- Multiparametric maps, 221
- Multipeak distribution, testing for, 233–234  
alternatives to *K*-means, 236–240  
main drawbacks of *K*-means, 234–236
- Multiple kernel learning (MKL), 141–148, 142f, 292, 293f  
advanced MKL methods, 147–148  
applications to brain disorders, 148–153  
linear MKL combination algorithms, 145–147
- Multiple local minima, 163–164
- Multiple regression, 71–72
- Multiple sclerosis (MS), 74  
predicting disease progression in, 74–76
- Multiset canonical correlation analysis (MCCA), 113–115
- Multivariable regression model, 75
- Multivariate analysis of variance (MANOVA), 5–6
- Multivariate approach, 5–6
- Multivariate linear models (MLM), 214, 218–219, 221
- Multivoxel pattern analysis (MVPA), 108–109
- ## N
- Natural images, 176
- Nested cross-validation, 31–32
- “Neural convergence” theories, 77–78
- Neuroanatomical age, 133–134
- Neurocognitive profiles, 240, 242
- Neuroethics, 335–336
- Neuroimaging, 5–6, 329–331, 335–336, 356  
data, 148–149  
to predicting brain age, 319–321  
SVM in, 108–109  
techniques, 283–284, 284f
- Neurological disorders, 218–221, 250, 331  
challenges in machine learning studies, 50–55
- Neurology, 50
- NeuroMiner, 343
- Neuropsychologically normal cluster, 242
- “Neurorights”, 335–336
- Neuroscience, 1–2
- Noise, 268–269, 308–309  
sources in machine learning, 309–311  
choice of features, 310–311  
choice of machine learning algorithm, 311  
choice of target variable, 309–310  
data processing, 311  
distribution, 268–269
- Nonlinear function, 181
- Nonlinear mapping, 91, 271
- Nonlinear MMD mapping function, 150
- Nonlinear models, extensions to, 91
- Nonlinear relationship, 269
- Nonlinear *e*-SVR, 128
- Nonlinearity layer for CNNs, 181
- Normalization, 360
- Numerical Operations score, 134–135
- Numpy, 344–345, 347–348
- “Nurture” theory, 68–69

## O

OASIS database. *See* Open Access Series of Studies database (OASIS database)  
 OLS. *See* Ordinary least squares (OLS)  
 Omic data, 212  
 Open Access Series of Studies database (OASIS database), 150  
 Optimal hyperplane, 143–144  
 Optimization, 88–89  
 Ordinary least squares (OLS), 70  
 “Out-of-sample” analyses, 73  
 Overfitting, 6–7, 24, 34–35, 103–104, 108–109, 164

## P

Padding, 181, 181f  
 Pairwise deletion, 255–256  
 Pandas, 344–345, 347–348  
 PANSS. *See* Positive and Negative Syndrome Scale (PANSS)  
 Parameter selection using cross-validation, 131  
 Parkinson’s disease (PD), 148–149, 219  
   diagnostic classification of, 150–151  
 Parkinson’s Progression Markers Initiative (PPMI), 58, 150–151  
 Pars orbitalis (Pars Orb), 135–136  
 Partial least squares (PLS), 214, 215t  
 Pathlib module, 347–348  
 Patient  
   future disease, 203  
   privacy, 334–335  
 Patient Health Questionnaire depression scale (PHQ-8 depression scale), 205–206  
 PCA. *See* Principal component analysis (PCA)  
 PD. *See* Parkinson’s disease (PD)  
 PDAT. *See* Probable dementia of Alzheimer type (PDAT)  
 Pearson correlation coefficient, 25–27  
 Penalized linear models, 87, 89–90  
 Penalized logistic regression, 83–84  
 Penalized regression methods, 84  
 Permutation testing, 41, 364–365  
 PET. *See* Positron emission tomography (PET)  
 PHQ-8 depression scale. *See* Patient Health Questionnaire depression scale (PHQ-8 depression scale)  
 Pixel space, 268–269  
 Plateau, 34–35

PLS. *See* Partial least squares (PLS)  
 pMCI. *See* Progressive MCI (pMCI)  
 pNC. *See* Progressive normal controls (pNC)  
 Pooling, 250–251  
   layer for CNNs, 182  
   pooled sample of controls, 262  
 Positive and Negative Syndrome Scale (PANSS), 312–313, 323  
 Positron emission tomography (PET), 148–149, 167–168  
 Post hoc analysis, 41–42, 364–368  
   identifying most informative features, 41–42  
   significance testing, 41  
 PPMI. *See* Parkinson’s Progression Markers Initiative (PPMI)  
 Pretraining, 194, 201  
 Principal component analysis (PCA), 12, 24–25, 132–133, 194, 209–210, 209f, 215t, 216f–217f, 242–243, 314–315, 318, 321  
   applications to brain disorders, 218–221  
   based MLM analysis, 222f  
   dimension reduction and variance, 213–214  
   data, analysis, and results, 215–216  
   dataset with dementia patients, 215–217  
   for discriminative analyses, 219f  
   extensions of, 214  
   interpretation, 216–217  
   method description, 212–217  
     singular value decomposition and covariance matrices, 212–213  
   rank-1 decomposition, 211f  
 Probabilistic approaches, 84, 89–90  
 Probabilistic classification, 89–90  
 Probable dementia of Alzheimer type (PDAT), 110–111  
 Prognosis of brain disorders, 109  
 Prognostic prediction, using clinical variables for, 322–323  
 Progressive MCI (pMCI), 168  
 Progressive normal controls (pNC), 168  
 PRoNTto, 343  
 Psychiatric disorders, 49, 218–219, 250, 331  
   challenges in machine learning studies, 50–55  
 Psychiatry, 50  
 Psychology, 1–2

- Psychosis  
   diagnostics classification of, 151–153  
   multimodal studies with kernel-based  
     and ensemble methods in, 297–298  
 Python, 344–345
- R**
- R-squared approach ( $R^2$  approach), 41  
 $R^2$  approach. *See* R-squared approach  
   ( $R^2$  approach)  
 rACC. *See* Rostral anterior cingulate cortex  
   (rACC)  
 Radial basis function kernels (RBF  
   kernels), 129  
 Radical behaviorism, 2–3  
 RBF kernels. *See* Radial basis function  
   kernels (RBF kernels)  
 read\_csv() function, 349  
 Receiving operating curve (ROC), 38–39  
 Receptive field, 174–175, 178, 180, 277  
 Recursive feature elimination (RFE), 106,  
   113–115  
   algorithm, 27  
 Regions of interests (ROIs), 109, 144, 318,  
   320–321  
   feature reduction using, 112b  
 Regression, 40–41  
   algorithm, 11  
   analysis, 123  
   coefficients, 90–91  
   MAE, 40  
   models, 67–69, 74  
   MSE, 40  
    $R^2$  approach, 41  
   RMSE, 40  
 Regularization, 87–88, 87t, 258, 268–269  
   penalty, 27–28  
   techniques, 27–28  
 Reinforcement learning, 2–3, 13, 101  
 “Relevance vector machine”, 89–90  
 Reliability, 308–309, 319  
   of diagnosis, 52–53  
 ReLU function’s gradient, 181  
 Representation learning methods,  
   193–194  
 Response mechanism or model, 254  
 Resting state fMRI (rsfMRI), 113–115  
 RFE. *See* Recursive feature elimination  
   (RFE)  
 Ridge regression, 87  
 RMSE. *See* Root mean squared error  
   (RMSE)
- ROC. *See* Receiving operating curve  
   (ROC)  
 ROIs. *See* Regions of interests (ROIs)  
 Root mean squared error (RMSE), 40  
 Rostral anterior cingulate cortex (rACC),  
   135–136  
 rsfMRI. *See* Resting state fMRI (rsfMRI)
- S**
- Saddle points, 163–164  
 Sample code, 347–368  
   data preparation, 349–357  
     class imbalance, 352  
     confounding variables, 352–356  
     feature set and target, 356–357  
     loading data, 349–351  
     missing data, 351–352  
   feature engineering, 357–360  
   importing libraries, 347–348  
   model evaluation, 363–364  
   model training, 360–362  
   organizing workspace, 348–349  
   post hoc analysis, 364–368  
   problem formulation, 349  
   results, 368  
   setting random seed, 348  
 Scaling/normalization, 34  
 Schizophrenia (SZ), 309, 345–347  
   using brain morphometry to classifying  
     patients with, 345–347  
   brain-based diagnostics of, 112–115  
   classification of, 94–96, 166–167  
   data fusion in, 298  
   predicting deviations from “brain age” in,  
     132–133  
 “School ranking”, 71–72  
 Scikit-learn (sklearn), 344–345, 347–348,  
   361  
 SD. *See* Standard deviation (SD)  
 Seaborn, 344–345, 347–348  
 Searchlight approach, 113  
 Searchlight MVPA approach, 113  
 Second principal axis, 209f, 210  
 Seizure detection, 187–188  
 Semisupervised learning, 12, 101, 228–229  
 Sensitivity (sens), 358  
   sensitivity/true positive rate/recall, 38  
 Separability, 315–316  
   measurement between two groups,  
     313–314  
 Sequential feature selection methods,  
   113–115

- Shapiro–Wilk test, 354–355
- Shared agency, 332–333
- Sigmoid function, 181
- Significance testing, 41
- Silhouette scores, choosing  $K$  with, 232–233
- Similar-go-together principle, 267–268
- Similarity metric, choice of, 229
- Simple models, 258
- Simple regression, 69–70
- Simplest loss function, 85
- Simulation-based augmentation, 259–260
- Single dot product, 177, 177f
- Single neuroimaging technique, 283–284
- Single Photon Emission Computerized Tomography (SPECT), 150–151
- Single-nucleotide polymorphism (SNP), 297, 307–308
- Singular value decomposition (SVD), 212–213
- sklearn. *See* Scikit-learn (sklearn)
- Small sample sizes, 249, 251
- algorithms and procedures, 257–260
- data simulation, 251–253
- sMCI. *See* Stable MCI (sMCI)
- sMRI. *See* Structural magnetic resonance imaging (sMRI)
- sNC. *See* Stable normal controls (sNC)
- Snippets, 345
- SNP. *See* Single-nucleotide polymorphism (SNP)
- Sociability measures, 334–335
- Soft clustering, 229
- Soft margin, 103–104
- Soft voting, 295–296
- Softmax function, 160
- Space-feature separability, 277
- Sparse penalties, 88
- Sparsity, 84, 88
- constraint, 197
- Spatial invariance, 175
- Specificity (spec), 358
- specificity/true negative rate, 38
- SPECT. *See* Single Photon Emission Computerized Tomography (SPECT)
- Speech comprehension, 77
- Speech production, 77
- SPM12. *See* Statistical Parametric Mapping software (SPM12)
- SSE. *See* Sum of squared error (SSE)
- Stable MCI (sMCI), 168
- Stable normal controls (sNC), 168
- Stacked feature maps, 178, 179f
- Standard deviation (SD), 360
- Starting weights, 163–164
- State-of-the-art machine learning approaches, 3
- Statistical methods, 67, 147–148
- Statistical Parametric Mapping software (SPM12), 215–216
- Statistical tests, 23
- Statistics, 1–2
- Step-by-step tutorial, 343–344
- using brain morphometry to classifying patients with SZ and HC, 345–347
- installing Python and main libraries, 344–345
- sample code, 347–368
- Stimulus, 275–276
- Stratified cross-validation (Stratified CV), 33
- Stride size, 180, 180f
- Stroke, predicting language impairment after, 77–79
- Structural equation modeling, 287
- Structural magnetic resonance imaging (sMRI), 72, 141–142, 167–168, 185–186, 283–284, 307–308, 319
- Structural neuroimaging, 74
- Structural regularization, 276–278
- Structured sparsity, 94
- “Suboptimal” measurements, 310–311
- Sum of squared error (SSE), 231–232
- Supervised learning, 9–11, 21, 22f, 101.
- See also* Unsupervised learning
- classification, 10
- models, 309–310
- regression, 11
- Support vector machine (SVM), 83–84, 88, 94–96, 101–104, 123, 141–142, 157, 212, 285–286, 292, 322
- applications to brain disorders, 109–116
- model, 345–347, 361–362
- in neuroimaging, 108–109
- stages, 104–108
- evaluating SVM performance, 107–108
- feature selection, 105–106
- training and testing classifier, 106–107
- Support vector regression (SVR), 123, 125–126
- applications to brain disorders, 131–136

- Support vector regression (SVR) (*Continued*)  
 graphic representation of linear  $\epsilon$ -SVR models, 125f  
 kernel SVR, 127–129  
 linear  $\epsilon$ -SVR model, 126–127  
 parameter selection using cross-validation, 131  
 V-SVR model, 129–131
- SVD. *See* Singular value decomposition (SVD)
- SVM. *See* Support vector machine (SVM)
- SVR. *See* Support vector regression (SVR)
- Symmetric data fusion, 287
- Systematic heterogeneity, 260
- SZ. *See* Schizophrenia (SZ)
- T**
- t-tests, 25–27  
 filtering, 113–115
- T1-weighted MR images, 133–134
- T1-weighted structural MRI, 131–132
- T2-weighted structural MRI, 131–132
- tanh function, 181
- Target variable, 22–23
- Targeted question, 257–258
- Task, 22–23
- TD children. *See* Typically developing children (TD children)
- TDC. *See* Typically developed controls (TDC)
- Test set, 29
- TF. *See* True negatives (TF)
- Theory-driven approaches, 7
- This-then-that principle, 267–268
- 3D CNN model, 299
- Tikhonov regularization. *See* Ridge regression
- Top-down approaches, 7
- TP. *See* True positives (TP)
- Tractography, 135–136
- Training, 32  
 procedure, 268–269  
 regularization, 275–276  
 set, 29
- Transfer learning, 259, 263
- Transparency of machine learning, 336–337
- Treatment response, 49
- True negatives (TF), 36
- True positives (TP), 36
- Two-dimension (2D)  
 distributions, 228–229  
 grid structure, 176
- Typically developed controls (TDC), 93
- Typically developing children (TD children), 322
- U**
- U-Net, 187
- UK Biobank, 8, 250–251, 259
- Uncorrelated PANSS Score Matrix (UPSM), 323
- Underfitting, 34–35
- Unsupervised dimensionality reduction, 321
- Unsupervised learning, 11–12, 101.  
*See also* Supervised learning  
 cluster analysis, 11  
 dimensionality reduction, 12
- Unsupervised methods, 318
- Unweighted simple sum of kernels, 292, 293f
- $\nu$ -SVR model, 129–130
- UPSM. *See* Uncorrelated PANSS Score Matrix (UPSM)
- V**
- Validation, 32. *See also* Cross-validation (CV)  
 cluster, 230–234
- Vapnik-Chervonenkis theory (VC theory), 125
- Variance, 213–214
- VGG model, 182–183
- VGG16 model, 182–183, 183f
- “Vibration” effects, 91–92
- Voxel-based gray matter density (VBM), 319–321
- Voxel-wise encoding models, 267–278  
 applications to brain disorders, 278–279  
 encoding models from Gabor feature space to deep learning networks, 271–275  
 fitting encoding model  
 structural regularization, 276–278  
 training regularization, 275–276
- Voxels, 108–109
- W**
- Ward clustering, 241–242

Ward's linkage, 240–241  
Weighting function, 147–148  
Weights, 145  
  of kernels, 150–151  
WM-f model, 75  
Wrapper methods, 27, 106, 258

## Y

Yale–Brown Obsessive–Compulsive  
  Scale (Y-BOCS), 241–242

## Z

z-score normalization, 28