

Index

- 3D-Var, 562
- 4D-Var, 563–567
- accuracy, 5
- activation function, 176
 - hyperbolic tangent function (tanh), 177–178, 183
 - logistic function, 176–178, 183
 - rectified linear unit (ReLU), 177–178, 496
 - softplus function, 177–178, 496, 539
 - swish function, 177–178, 496
- active remote sensing, 208
- adjoint, 562
- Advanced Very High Resolution Radiometer (AVHRR), 209
- aerosols, 209
- AI winters, 2
- air quality, 213
- Akaike information criterion (AIC), 275–278
- Akaike information criterion corrected (AIC_c), 277–278
- aliasing, 382–384
- alternative hypothesis, 101
- altimeter, 211
- analysis (in NWP), 562
- analysis of variance (ANOVA), 149–151
- Anderson–Darling test, 120–121, 123
- anomaly, 25
- Antarctic Oscillation (AAO), 313
- architecture-constrained NN (ACnet), 558–559
- Arctic Oscillation (AO), 171, 313
- arg min, 451
- artificial intelligence (AI), 2–4, 570
- artificial neural network (ANN)
 - (see neural network (NN))
- atmospheric radiation, 550–551
- attention U-net, 510
- attribute, 15
- augmented data matrix, 406
- auto-covariance, 387
- auto-regressive (AR) process, 37, 410–413
- auto-regressive integrated moving average (ARIMA) model, 415
- auto-regressive moving average (ARMA) model, 414–415
- autocorrelation, 36–38, 107–109, 405
- autoencoder, 347–362, 546
 - sparse, 349
 - denoising, 348
 - dimensionality reduction, 347–362
 - variational, 348
- automatic differentiation, 557
- automatic relevance determination (ARD), 467
- autospectrum, 379–387
- averaging in time, 404–405
- axon, 175
- back-propagation, 180, 221–225
- backward optimization, 280
- bagging (bootstrap aggregating), 214, 263, 266, 269, 482
 - confidence interval, 263
 - out-of-bag (OOB), 263
 - prediction interval, 263
- Bartlett method, 384
- base rate, 519
- batch learning, 27, 154, 224
- batch normalization, 497
- Bayes factor, 272
- Bayes theorem, 21, 49–53
- Bayes, Thomas, 21
- Bayesian, 49, 51
- back-propagation, 180, 221–225
- backward optimization, 280
- bagging (bootstrap aggregating), 214, 263, 266, 269, 482
 - confidence interval, 263
 - out-of-bag (OOB), 263
 - prediction interval, 263
- Bartlett method, 384
- base rate, 519
- batch learning, 27, 154, 224
- batch normalization, 497
- Bayes factor, 272
- Bayes theorem, 21, 49–53
- Bayes, Thomas, 21
- Bayesian, 49, 51

- Bayesian information criterion (BIC), 273–275
- Bayesian model averaging (BMA), 278–280
- Bayesian model selection, 272–273
- Bayesian neural network (BNN), 260
- Beijing, 79
- beluga whale, 498
- Bernoulli distribution, 66
- Bessel
 - Friedrich, 25
- best linear unbiased estimator (BLUE), 147, 561
- beta distribution, 80–82
- BFGS method, 233–234
 - limited memory, 234
- bias
 - frequency bias, 520
- bias error, 250–251, 264–266
- bias parameter, 175
- Big data, 570
- bike sharing system, 493
- binomial distribution, 66–68
- bits, 58
- bivariate Gaussian distribution, 76
- biweight midcorrelation, 40–41
- black box, 4
- boosting, 473, 487–492
 - AdaBoost, 488
 - CatBoost, 488
 - gradient, 488–492
 - LightBoost, 488
 - XGBoost, 491–492
- bootstrapping, 126–131, 133–134, 214, 263, 269, 482
 - block, 131
 - confidence interval (CI), 126–131, 263
 - basic method, 128–129
 - BCa method, 129
 - percentile method, 128–129
 - field significance, 133–134
 - moving block, 131
 - prediction interval, 263
 - serial correlation, 131
- bottleneck, 347
- Box–Cox transformation, 94
- Box–Jenkins models, 414–415
- boxplot, 45–46
- breakdown point, 30
- Brent’s method, 231
- Brier score (BS), 528
 - multiple classes, 531
- Brier skill score (BSS), 528
- Burgers’ equation, 556–557
- canonical correlation analysis (CCA), 320–327
 - pre-filter with PCA, 324–326
- canonical correlation coordinates, 321
- canonical variate, 321, 365
- CART (*see* classification and regression tree)
- categorical data, 6, 19, 156–158
- central limit theorem, 7–8, 72, 404
- centroid, 55
- channel, 501
- cherry-picking, 537
- chi-squared distribution, 97–98, 385
- chlorophyll, 209
- circular data, 156–157
- circular statistics, 82
- city-block distance, 332
- class imbalance, 436–438
- classical mechanics, 5
- classification, 6, 15–16, 50–54, 418–439
 - binary, 15
 - multi-class, 15, 461–462
- classification and regression tree (CART), 474–481
 - relative importance of predictors, 479–481
 - surrogate split, 480–481
- climate downscaling, 543–547
- climate extremes, 8, 405
- climate of extreme weather events (*see* climate extremes)
- climatological seasonal cycle, 396
- climatology forecast, 524, 533–534
- cloud, 209–210, 508, 516
- cloud cover, 80–82
- cloud resolving model (CRM), 551
- CloudSat, 517
- cluster analysis (*see* clustering)
- clustering, 16, 54–56, 330–343, 453
 - Calinski–Harabasz index, 333
 - external evaluation, 335–337
 - gap statistic, 333–334
 - Gaussian mixture model, 339
 - hierarchical, 339–343

- internal evaluation, 332–334
 - model evaluation, 332–337
 - non-hierarchical, 337–339
 - co-kriging, 467
 - code (*see* bottleneck)
 - coefficient of determination, 148
 - collinearity, 159, 163, 172
 - Columbus, 4
 - committee, 262
 - complementary cumulative distribution
 - function (CCDF), 23, 103
 - computational intelligence, 2, 570
 - condition number of a matrix, 163
 - conditional density network (CDN), 199–200, 544
 - conditional distribution, 21–23
 - conditional entropy, 59–61
 - conditional Gaussian distribution, 77
 - conditional probability, 21–22, 50
 - conditional probability density, 23
 - conditional probability distribution, 199–204
 - confidence interval (CI), 115–119
 - bagging, 263
 - bootstrapping, 126–131
 - correlation, 118–119
 - joint confidence region, 152
 - multiple linear regression, 151–153
 - population mean, 116–118
 - simple linear regression, 141–145
 - spectrum, 385
 - confidence level, 116
 - congruence, 317
 - conjugate gradient method, 229–233
 - contingency table, 519
 - continuous data, 19
 - continuous ranked probability score (CRPS), 535
 - continuous-time Fourier transform (CTFT), 375
 - convolution, 377
 - discrete, 499–502
 - missing data, 509
 - partial, 509
 - theorem, 391
 - convolutional layer, 496
 - convolutional LSTM (ConvLSTM) network, 512–513
 - convolutional neural network (CNN), 497, 499–510
 - fully (FCN), 508
 - temporal (TCN), 513–514
 - correlation, 33
 - a pair of correlations, 110–111, 119
 - confidence interval, 119
 - significance test, 110–111
 - biweight midcorrelation, 40–41
 - confidence interval, 118–119
 - Kendall, 39–40, 125
 - multiple correlation, 148
 - Pearson, 33–36
 - rank correlation, 38–40
 - relation to MSE, 533
 - sample correlation, 33
 - serial, 36–38, 67–68, 107–110
 - significance test, 109–111
 - Spearman, 38–39
- cosine taper, 379
 - cost function, 173, 181
 - covariance, 26–27
 - auto-covariance, 387
 - cross-covariance, 394
 - online algorithm, 28
 - covariance matrix, 26–27, 75, 155, 285, 300
 - covariate, 15, 137
 - critical success index (CSI), 520
 - cross-entropy, 60, 432–434, 477
 - cross-spectrum, 393–394
 - cross-validation, 164, 255–258, 477, 537–538
 - cumulative distribution function (CDF), 23, 71
 - curse of dimensionality, 17–18, 185, 430, 442, 450
 - cyclonic, 388
 - Daniell method, 384
 - dark energy, 5
 - dark matter, 5
 - dark universe, 5
 - Dartmouth Summer Research Project, 2
 - data
 - categorical, 6, 12–15, 19
 - continuous, 6, 12–13, 19
 - discrete, 6, 12–15, 19
 - nominal, 19
 - ordinal, 6, 19
 - structured, 497

- test, 6
- training, 6
- types, 19
- validation, 11
- data assimilation, 560–567
 - 3D-Var, 562
 - 4D-Var, 563–567
 - ensemble Kalman filter, 567
 - variational, 560–567
- data augmentation, 506
- data driven, 3, 570
- data mining, 3, 570
- data science, 3–5, 570
- decibel (dB), 378
- decile, 31
- decision boundary, 52, 179
- decision region, 52
- decision surface, 52
- decision tree, 473–481
 - C4.5, 478
 - C5.0, 478
 - classification and regression tree (CART), 474–481
- decoder, 347
- deep belief network, 494
- deep learning (DL), 494–517
- deep neural network (DNN), 494–517
- DeepLabv3, 510
- degrees of freedom, 97
- delay coordinate space, 406
- dendrite, 175
- denoising, 348
- dense layer, 503
- densely connected convolutional networks (DenseNet), 505
- density estimation, 65
- dependent variable, 15
- depth
 - of a NN model, 501
 - of a layer, 501
- design matrix, 146
- deterministic optimization, 236
- detrend, 382
- deviance, 477
- DFP method, 233
- Differential evolution (DE), 241–244
- dimensionality reduction
 - autoencoder, 347–362
 - non-linear, 347–364
 - principal component analysis (PCA), 300–301
- direct numerical simulation (DNS), 553
- directional statistics, 82
- discrete data, 19
- discrete Fourier transform (DFT), 374–375
- discrete-time Fourier transform (DTFT), 376
- discriminant analysis
 - Fisher, 421–423
 - linear (LDA), 419–423
 - quadratic (QDA), 419
 - regularized, 419
- discriminant function, 53–54, 419
- dissolved organic carbon (DOC), 82
- distance, 331–332
 - city-block, 332
 - Euclidean, 331
 - Hamming, 332
 - Minkowski, 332
- distribution, 12–15
 - beta, 80–82
 - chi-squared, 97–98
 - conditional, 199–204
 - exponential, 78
 - Fréchet, 84
 - gamma, 78–80, 98, 200
 - Gaussian, 7, 9, 13–15, 69–72, 75–77, 98
 - generalized extreme value (GEV), 84–86
 - Gumbel, 84
 - multivariate Gaussian, 75–76
 - Pearson type III, 78
 - Student t , 95–97
 - von Mises, 82–83
 - Weibull, 13, 84
- downsampling, 502–503
- downscaling, 543–547
 - climate, 543–547
 - climate of extremes, 544–545
 - dynamical, 543
 - expanded, 547
 - inflation, 546
 - precipitation, 544–545
 - quantile mapping, 546
 - randomization, 546
 - statistical, 543
 - streamflow, 546

- wind speed, 545–546
- dropout, 269–271, 497
- dual Lagrangian, 457
- dual solution, 444
- dual variable, 443
- Durbin–Watson statistic, 144–145
- early stopping, 224, 254–255, 263, 266
- effective sample size, 37–38, 46, 106, 108–109, 126
- efficiency
 - Legates–McCabe, 534
 - Nash–Sutcliffe, 534
- Ekman spiral, 319
- El Niño, 106, 113–114, 135, 290–295
- El Niño–Southern Oscillation (ENSO), 60, 121, 171, 213–214, 280, 290–295, 354–358, 366–367, 390–392, 401, 407–408, 498, 507, 513, 554–555
- empirical orthogonal function (EOF), 283, 289, 298
 - extended (EEOF), 408
- emulation, 550
- encoder, 347
- encoder–decoder network, 507–510
 - attention U-net, 510
 - DeepLabv3, 510
 - U-net, 508–510, 546
 - UNet 3+, 510
 - UNet++, 510, 546
- ensemble, 261–269
 - unequal weights, 266–269
- ensemble Kalman filter, 567
- entropy, 56–62
 - conditional, 59–61
 - cross, 60
 - joint, 59
 - relative, 60
- entropy impurity, 477
- environmental data science, 6–9
- environmental science (ES), 6
- EOF (*see* empirical orthogonal function)
- epoch, 224
- equitable threat score (ETS), 526
- error function, 173, 181
 - ϵ -insensitive, 463
- Euclidean distance, 331
- evolutionary algorithm (EA), 236–244
- evolutionary computation, 237
- expectation, 24
- expectation–maximization (EM) algorithm, 90–91, 339
- expert systems, 2
- explanatory variable, 137
- exploratory data analysis (EDA), 41–46
- exponential distribution, 78
- extended empirical orthogonal function (EEOF), 408
- extra trees (ET), 487
- extrapolation, 6–7, 12, 538–541
- extreme learning machine (ELM), 189–194, 196–199
 - classifier, 430–432
 - online sequential (OSELM), 193–194
- extreme values, 83–86
- extremely randomized trees (extra trees, ET), 487
- F distribution, 124
- F -measure, 522
- F -test, 124
- factor analysis (FA), 311–312
- false alarm rate (F), 521
- false alarm ratio (FAR), 521
- fast Fourier transform (FFT), 386
- feature, 6, 15, 52, 137, 471
- feature selection (*see* predictor selection)
- feature space, 52, 442, 460
- feed-forward neural network (FFNN), 174, 180
- field significance, 131–134
 - bootstrap method, 133–134
 - false discovery rate (FDR) method, 134
- filter, 395–403
 - 1-2-1 filter, 402–403
 - 3-point triangular, 402–403
 - finite impulse response (FIR), 398
 - moving average (MA), 402–403
 - periodic signal, 396
 - response function, 397
 - symmetric, 399
 - tidal, 396
 - windowing method, 399–401
- Fisher transformation, 110, 118–119

- Fisher–Snedecor distribution (*see F* distribution)
- Fletcher–Reeves method, 231
- folding frequency (*see Nyquist frequency*)
- forecast lead time, 518
- forecast rate, 520
- forecast verification, 518–535
 - binary classes, 519–526
 - multiple classes, 527–528
 - probabilistic binary classes, 528–530
 - probabilistic multiple classes, 531–532
- forest, 209, 420
- Fourier analysis, 372–376
- Fourier series, 373–374
 - discrete Fourier transform, 374–375
- Fourier transform, 375–376
 - continuous-time, 375
 - discrete, 374–375
 - discrete-time, 376
- fraction correct, 520
- Fréchet distribution, 84
- frequentist, 49, 51
- fully convolutional network (FCN), 508
- fundamental frequency, 382
- gamma distribution, 78–80, 98, 200
 - generalized, 78
- gamma function, 78, 97
- Gauss–Markov theorem, 147
- Gauss–Newton method, 235
- Gaussian distribution, 7, 9, 13–15, 42, 43, 69–72, 75–77, 98
 - bivariate, 76
 - conditional, 77
 - marginal, 77
 - multivariate, 75–76
 - sample mean, 72
 - standard, 71
- Gaussian kernel, 93
- Gaussian mixture model, 86–91, 201, 339
- Gaussian process (GP), 440, 442, 463–469
- general circulation model (GCM), 212, 367–369, 543
- general linear model, 153–154
- generalization error, 251
- generalized cross-validation, 258
- generalized extreme value (GEV) distribution, 84–86, 200
- generalized gamma distribution, 78
- generalized least squares (GLS), 168–170
- generative adversarial network (GAN), 514–517
 - conditional (CGAN), 516–517
- generative topographic mapping (GTM), 345
- genetic algorithm (GA), 237, 239–241
- geopotential height, 312
 - 500 hPa (Z500), 367–369
- Geostationary Operational Environmental Satellites (GOES), 210
- Gilbert skill score (GSS), 526
- Gini impurity index, 477
- Glacier, BC, 129
- global climate model (GCM) (*see general circulation model (GCM)*)
- global minimum, 461
- goodness-of-fit test, 119–123
 - two-sample, 121–123
- Google Trends, 3, 569
- gradient boosting machine (GBM), 488–492
 - CatBoost, 488
 - LightBoost, 488
 - XGBoost, 491–492
- gradient descent method, 221, 225–227, 233, 236
- gradient tree boosting (*see gradient boosting machine*)
- Gram matrix, 443
- greedy algorithm, 238
- grid search, 260
- Grouse Mountain, BC, 113, 126, 135, 172, 282
- Gumbel distribution, 84
- hail, 213, 521
- Hamming distance, 332
- Hamming window, 379
- Hann window, 379
- harmonic analysis, 396
- Heaviside step function, 175, 178
- Heidke skill score (HSS), 525
- Hessian matrix, 217
- heteroscedastic, 169, 486
- hidden layer, 180, 186
- hidden neuron (*see neuron, hidden*)
- hidden node (*see neuron, hidden*)

- hill climbing, 238–239
 random mutation, 238
 steepest ascent, 238
- Hinton, Geoffrey, 4, 16
- histogram, 41–42, 59, 61
- hit rate, 520
- Hotelling T^2 test, 105
- Huber function, 204–206, 250
- humidity, 34, 43, 45, 144, 149, 160
- hurricane, 67
- hybrid coupled model, 554–555
- hydrology, 212
- hyperbolic tangent function (tanh), 177–178, 183
- hyperparameter, 237, 252, 448
 Bayesian approach, 260–261
 grid search, 260
 random search, 260
 tuning, 258–261
- hyperplane, 54, 454–455
- hypothesis testing, 101–111
 field significance, 131–134
- i.i.d. (independent and identically distributed), 72
- ignorance score, 528–529, 535
 multiple classes, 531
- ill-conditioned matrix, 163, 198
- impurity, 477
- imputation, 304–305, 358, 509–510
- inconsistency index, 359, 362
- independent and identically distributed (i.i.d.), 72
- independent variable, 15
- index of agreement, 534
- indicator function, 435
- infiltration, 212
- informatics, 2, 570
- information criterion, 273–278, 360
 Akaike (AIC), 275–278
 Akaike, corrected (AIC_c), 277–278
 Bayesian (BIC), 273–275
- information theory, 56–62
- infrared, 209–210
- Intergovernmental Panel on Climate Change (IPCC), 262
- interpretability, 5
- interpretable AI, 280–281
- interpretable ML, 280–281
- interquartile range (IQR), 31
- inverse modelling, 560
- isomap, 363
- Jacobian, 562
- Johnson transformation, 95
- joint confidence region, 152
- joint entropy, 59
- joint probability, 20–22
- joint probability density, 23
- K -means clustering, 55, 89, 197, 337–338, 344, 453
 K -means++, 55
- K -nearest neighbours (KNN), 18, 61, 428–430
- Karush–Kuhn–Tucker (KKT) conditions, 456, 457, 459, 571
- Kendall's tau, 39–40, 125
- kernel, 201, 444–448, 469
 automatic relevance determination, 467
 Gaussian, 447, 460
 Matérn, 467
 polynomial, 447
 radial basis function (RBF), 447, 460
 ridge regression, 448–449
- kernel density estimation (KDE), 61, 91–93
- kernel function, 445, 460
- kernel matrix, 446
- kernel method, 440–471
- kernel principal component analysis, 451, 469–471
- kernel trick, 445, 450, 460
- Kolmogorov–Smirnov (KS) test, 119–123, 136
- kriging, 463
 co-kriging, 467
- Kronecker delta function, 49, 75, 416, 465
- Kullback–Leibler (KL) divergence, 60, 275
- kurtosis, 32
- L -moment, 32
- L_1 norm, 165
- L_2 norm, 165
- La Niña, 113–114, 135, 290–295
- Lagrange function (Lagrangian), 457, 570

- Lagrange multiplier, 164, 165, 286, 321, 456, 569–572
- Landsat, 209
- lasso, 165–166
- latent variable, 87–88
- layerwise relevance propagation (LRP), 280–281
- lead time, 518
- leaf angle, 81
- learning, 224
 - batch, 27
 - online, 27–28
- learning rate, 222, 260
- least squares, 198
 - generalized (GLS), 168–170
 - ordinary (OLS), 138, 145–147, 168
- leave-one-out cross-validation, 257–258
- Legates–McCabe efficiency, 534
- Levenberg–Marquardt method, 236
- Lighthill report, 2
- likelihood, 50–51, 73–74, 200, 246
- Lilliefors test, 121
- line search, 231
- linear discriminant analysis (LDA), 54, 419–423
- linear least squares, 138, 145–147
 - recursive, 154–156
- linear regression (*see* regression)
- linearly separable, 179
- loading, 289
- locally-linear embedding (LLE), 363
- log odds, 419
- logarithmic score, 528–529
- logistic function, 176–178, 183, 424, 436
- logistic regression, 424–427
 - multiclass, 425–427
 - multinomial, 425–427
- long short-term memory (LSTM) network, 510–513
- longwave radiation (LWR), 550
- loss function, 167–168, 173, 181
- Lotka–Volterra equations, 568
- L_p norm, 165
- M5 model tree, 479
- machine learning (ML), 1–5, 570
- Madden–Julian oscillation (MJO), 362
- Mahalanobis distance, 47–49, 75–76
- Mann–Kendall trend test, 125–126
- Mann–Whitney U -test (*see* Wilcoxon–Mann–Whitney test)
- Mann–Whitney–Wilcoxon test, 136
- marginal distribution, 20–23
- marginal Gaussian distribution, 77
- marginal likelihood, 272
- marginal probability, 20–22
- marginal probability density, 23
- Markov chain Monte Carlo method, 464, 467
- matrix
 - Hermitian, 288
 - orthonormal, 302
 - positive semi-definite, 288
- max pooling, 502
- maximum covariance analysis (MCA), 320, 326–327
- maximum likelihood, 78, 200, 245–247
- maximum likelihood estimation, 73–74
- maximum margin classifier, 454–462
- maximum norm (max-norm) constraint, 271–272
- McCulloch and Pitts model, 174–175
- mean, 24
 - conditional, 248
 - online algorithm, 27–28
 - population mean, 24
 - sample mean, 24
- mean absolute error (MAE), 247, 532
- mean error (ME), 532
- mean squared error (MSE), 10–11, 181, 245–247, 532–534
- mean squared error skill score (MSE SS), 533–534
- median, 29–30, 40, 41, 167–168
 - conditional, 250
- median absolute deviation (MAD), 30, 41
- Medium Resolution Imaging Spectrometer (MERIS), 209
- memory-based model, 428
- Mercer’s theorem, 446
- MeteoNet, 517
- method of moments, 70
- mini-batch, 229, 232, 270
- Minkowski distance, 332
- misclassification rate, 477
- missing data, 304–305, 509–510
- mixing coefficient, 86, 201

- mixture density network (MDN), 201–204
- mixture model, 86–87, 201–204
- MLP (see multi-layer perceptron), 180
- mode (of a distribution), 70
- model evidence, 272
- model output statistics (MOS), 542–543
 - updatable MOS (UMOS), 543
- model tree, 479
- model validation, 11
- Moderate-Resolution Imaging Spectroradiometer (MODIS), 517
- momentum, 227
- monotonic MLP (MonMLP), 187–189, 204
- Monte Carlo method, 108
- Moore–Penrose inverse, 146, 153, 190, 198
- moving average (MA) filter, 384, 402–403
- moving average (MA) process, 414
- multi-class classification, 461–462
- multi-layer perceptron (MLP), 3, 174, 180–186, 198, 365
 - classifier, 434–436
 - monotonic, 187–189
- multi-model ensemble, 267
- multichannel singular spectrum analysis (MSSA), 407–408
- multinomial distribution, 68–69
- multiple correlation coefficient, 148
- multiple linear regression (see regression)
- multivariate Gaussian distribution, 75–76
- multivariate linear regression (see regression)
- mutual information, 61–62
- naive Bayes classifier, 427–428
- Nash–Sutcliffe efficiency, 534
- nats, 58
- Navier–Stokes equation, 557–558
- neural network (NN), 3, 5, 12, 173–214
 - Bayesian, 260
 - confidence interval, 263
 - deep (DNN), 494–517
 - prediction interval, 263, 266
 - randomized, 189–195
 - recurrent, 180, 210
- neuron, 173, 174, 176
 - hidden, 180, 186–187
- Newton’s method, 220, 232
- Niño 1+2 region, 63, 291
- Niño 3 region, 291
- Niño 3.4 region, 63, 121, 136, 171, 215, 291
- Niño 4 region, 291
- node, 176
- nominal data, 19, 531
- non-linear canonical correlation analysis (NLCCA), 365–369
 - robust, 366
- non-linear complex principal component analysis, 358
- non-linear dimensionality reduction, 347–364
- non-linear least squares, 234–236
- non-linear principal component analysis (NLPKA), 349–362, 507, 509
 - closed curve, 360–362
 - complex variables, 358
 - hierarchical, 353–354
 - missing data imputation, 358
 - open curve, 350–360
 - overfitting, 358–360
- non-linear singular spectrum analysis (NLSSA), 408–410
- non-parametric distribution, 65
- non-parametric models, 428
- non-parametric statistics, 111–115
- non-stationary data, 395
- normal distribution (see Gaussian distribution)
- normal equations, 138, 146, 235
- North Atlantic Oscillation (NAO), 171, 313
- nowcasting, 508, 513
- null hypothesis, 101–105
- numerical weather prediction (NWP), 2, 541
- Nyquist frequency, 382
- objective function, 162, 173, 181, 247–250
- oblique rotation, 305
- observation operator, 562
- ocean colour, 209
- ocean heat content, 213
- odds ratio, 524

- offset parameter, 175
- oil spill, 212
- Old Faithful geyser, 171
- one-hot encoding, 157, 430, 434, 521
- one-of- C encoding (*see* one-hot encoding)
- one-tailed test, 103, 105, 108–109
- one-versus-one classification, 461
- one-versus-the-rest classification, 461
- online learning, 27–28, 154–156, 193–194, 229, 543
- online training, 224
- operation count, 147
- optimal fingerprinting, 170
- ordinal data, 6, 19, 528, 531
- ordinary least squares (OLS), 138, 145–147, 168
- orthonormal matrix, 302
- orthonormal rotation, 305–315
- out-of-bag (OOB) data, 263
- outlier, 6–7, 29, 247
- overfit, 3, 9–12, 158, 172, 224, 245, 249, 358
- ozone concentration, 172

- p -value, 103
- Pacific-North American (PNA) pattern, 171, 215, 313, 367
- pairwise classification, 461
- parametric density estimation, 92
- parametric distribution, 65
- parametric models, 428
- passive remote sensing, 208
- Pearson type III distribution, 78
- Peirce skill score (PSS), 526
- percentile, 30, 31, 71
- perceptron, 175–180
- perfect prog, 542–543
- periodogram, 380
- persistence, 36
- persistence forecast, 524, 533–534
- physics-informed machine learning (PIML), 556–559
- physics-informed neural network (PINN), 556–559
 - hard constraint, 558–559
 - soft constraint, 556–558
- phytoplankton, 209
- PM_{2.5} concentration, 7, 78, 514
- Poisson distribution, 68

- Polak–Ribiere method, 231
- polynomial, 9, 18, 185
- pooling, 502–503
- positive semi-definite, 446
- post-processing, 541–543
- power (of statistical test), 102–103, 120
- pre-image, 451–453
- precipitation, 16, 78, 96, 210, 508, 510, 513, 517
- Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PER-SIANN), 508
- precision, 522
- predictand, 15, 137
- prediction interval (PI)
 - bagging, 263
 - Gaussian process, 467
 - multiple linear regression, 151–153
 - neural network, 266
 - random forest, 485–486
 - simple linear regression, 143–145
- predictor, 6, 15, 137, 148–149
 - selection, 5, 62, 158–160
 - standardized, 148–149
- pressure, 34, 43, 89, 149, 160
- primal solution, 444
- principal component (PC), 286, 289
- principal component analysis (PCA), 16, 47–49, 75–76, 283–301
 - 2-dimensional vectors, 317–319
 - kernel PCA, 469–471
 - amplitudes, 289
 - Buell patterns, 315
 - combined PCA, 288, 318
 - complex variables, 288, 318, 358
 - correlation, 288
 - degenerate eigenvalues, 299
 - dimensionality reduction, 300–301
 - eigenvalues, 288
 - eigenvectors, 289
 - Hotelling scaling, 298
 - Kaiser test, 301
 - latitudinal scaling, 290
 - loadings, 289
 - Lorenz scaling, 298
 - missing data, 304–305
 - Monte Carlo test, 301
 - non-linear, 349–362

- number of modes to retain, 300–301
- orthogonality relations, 289–290
- principal components, 289
- removing temporal or spatial mean, 302
- rotated, 305–317
- rotated A-frame, 311–315
- rotated E-frame, 308–310
- Rule N, 301
- scaling, 298–299
- scores, 289
- scree graph, 301
- smaller covariance matrix, 299–300
- space–time PCA, 407–408
- principal vector, 289
- probabilistic forecast
 - binary classes, 528–530
 - continuous variable, 535
 - multiple classes, 531–532
- probability, 1–2, 19–22
 - p*-value, 103
 - a posteriori*, 50
 - a priori*, 50
 - Bayesian, 49, 51
 - complementary cumulative, 23, 103
 - conditional, 21–22, 50
 - cumulative, 23
 - density, 22–23
 - frequentist, 49, 51
 - joint, 20–22
 - marginal, 20–22
 - posterior, 50, 88
 - prior, 50, 88
 - tail, 23, 103
- probability density function (PDF), 23
- probability distribution (*see* distribution)
- probability of detection (POD), 520
- probability of false detection (POFD), 521
- proportion correct, 520

- QR decomposition, 147
- quadratic discriminant analysis (QDA), 419
- quadratic programming, 165, 456
- quadrature, 408
- quantile, 30–31, 71–72, 167–168
- quantile regression, 167–168, 188, 485–486
 - neural network (QRNN), 204–207, 544
- quantile regression forest, 485–486
- quantile–quantile (Q–Q) plot, 42–45
- quantization error (QE), 344
- quantum mechanics, 5
- quartile, 31, 45
- quasi-biennial oscillation (QBO), 362
- quasi-Newton method, 232–234
 - BFGS method, 233–234
 - DFP method, 233
 - limited memory, 234
- quasi-Newton methods, 236
- quintile, 31

- radar, 208, 211, 213
- radial basis functions (RBF), 195–199
 - normalized, 196
- radiometer, 208
- rainfall, 439
- random forecast, 524
- random forest (RF), 5, 473, 482–487
 - prediction interval, 485–486
 - quantile regression, 485–486
 - relative importance of predictors, 483–485
- random vector functional link (RVFL), 194–195
- randomized neural network, 189–195
- ranked probability score (RPS), 531–532
- re-expressing data, 93–95
- reanalysis, 404, 560
- recall, 522
- rectified linear unit (ReLU), 177–178, 496
- recurrent neural network (RNN), 180, 210, 510–513
- recursive least squares, 154–156
- regional climate model (RCM), 543
- regressand, 137
- regression, 6, 15–16
 - analysis of variance (ANOVA), 149–151
 - dual solution, 442–444
 - general linear model, 153–154
 - lasso, 165–166
 - linear, 137–172

- multiple linear regression (MLR),
 - 145–156
 - without intercept term, 146
- multivariate linear regression, 153–154
- online learning, 154–156
- partition of sums of squares, 139–141, 148
- quantile, 167–168
- relation with correlation, 139
- ridge, 161–165, 172, 252, 443, 448
- serial correlation, 144–145
- simple linear regression (SLR), 10, 137–145
- stepwise, 158–160
- regressor, 137
- regularization, 158, 251–254
 - Tikhonov, 161
- regularized least squares, 161
- reinforced learning, 17
- relative efficiency, 113
- relative entropy, 60
- relative humidity, 80–82, 96
- relative operating characteristic (ROC), 521, 522
- reliability diagram, 529–530, 532
- remote sensing, 208–212
 - active, 208
 - infrared, 209–210
 - microwave, 210–212
 - passive, 208
 - visible light, 209
- resampling, 126
- residual neural network (ResNet), 505–506
- resistant, 29, 30, 34, 38, 247–250
- response variable, 6, 15, 137
- retroactive validation, 538
- return level, 85
- return period, 85
- Reynolds stress, 553
- ridge regression, 161–165, 172, 252, 443
- River Thames, 85
- robust, 29, 30, 34, 38, 247–250
- robust statistics, 29
- rogues wave, 85
- rotated principal component analysis (RPCA), 305–317
 - A-frame rotation, 311–315
 - E-frame rotation, 308–310
 - oblique rotation, 305
 - orthonormal rotation, 305–315
 - promax rotation, 308
 - S-mode decomposition, 315
 - T-mode decomposition, 315
 - varimax rotation, 308–315
- rotation
 - orthonormal, 305–315
- running mean (*see* moving average)
- runoff, 212
- saddle point, 218
- sample size, 3, 15
 - effective, 37–38, 145
- scatterometer, 211
- scatterplot, 33–34
- score, 289
- scree graph, 301
- sea level, 214
- sea level pressure (SLP), 366–367
- sea surface temperature (SST), 185, 211, 213–214, 291, 354–358, 366–369
- Sea-viewing Wide Field-of-view Sensor (SeaWiFS), 209
- self-organizing map (SOM), 343–347, 360
- semantic segmentation, 507–510
- sensitivity, 223
- sequential training, 224
- serial correlation, 36–38, 46, 67–68, 106–110, 113, 144–145, 257
- sgn function, 125
- Shapiro–Wilk test, 121
- shortwave radiation (SWR), 550
- shrinkage method, 159–166
- sigmoidal function, 176
- significance level, 103
- significance test, 101–104
- Silverman’s rule of thumb, 93
- Sinclair Pass, BC, 127
- singular spectrum analysis (SSA), 405–408
 - non-linear, 408–410
 - rotation, 408
- singular value, 302
- singular value decomposition (SVD), 147, 302–304, 326
- singular vector, 302, 327
- skewness, 32, 67
- skill score, 525

- binary classes, 524–526
- climatology as reference, 533–534
- continuous variable, 533–534
- equitable, 527–528
- equitable threat score, 526
- Gilbert, 526
- Hanssen and Kupers, 526
- Heidke, 525
- Legates–McCabe efficiency, 534
- multiple classes, 527–528
- Nash–Sutcliffe efficiency, 534
- Peirce, 526
- persistence as reference, 533–534
- skip connection, 194, 505–506, 508
- slack variable, 458
- Snedecor's F distribution (*see F* distribution)
- snow, 210–211
- snow water equivalent (SWE), 113, 126, 129, 172, 212, 282
- soft computing, 2, 570
- softmax function, 201, 435–436
- softplus function, 177–178, 496, 539
- soil properties, 81
- Southern Oscillation, 295
- Southern Oscillation Index (SOI), 63, 295
- sparsity, 166, 458
- Special Sensor Microwave Imager (SSM/I), 210–211
- spectrum, 379–394
 - amplitude spectrum, 393
 - autospectrum, 379–387
 - band averaging, 384
 - Bartlett method, 384
 - co-spectrum, 393
 - coherence spectrum, 393
 - confidence interval, 385
 - cross-spectrum, 393–394
 - Daniell method, 384
 - ensemble averaging, 384
 - phase spectrum, 393
 - quadrature spectrum, 393
 - rotary, 387–390
 - smoothing, 384–385
 - Welch method, 384
- SSE (sum of squared errors), 138–140
- SSR (sum of squares, regression), 140
- SST (*see sea surface temperature*)
- SST (sum of squares), 140
- stacked generalization, 268–269
- stacking, 268–269
- standard deviation, 25
 - population standard deviation, 25
 - sample standard deviation, 25
- standard error, 72
- standard Gaussian distribution, 71
- standard score, 25, 71
- standardized anomaly, 25
- standardized variable, 25, 71, 183, 253
- stationary process, 395, 412
- statistical inference, 101–111
- statistics, 1–5
- Stave River, B.C., 79
- steepest descent method, 221, 225
- stepwise regression, 158–160
- stochastic gradient descent (SGD), 221, 227–229
- stochastic optimization, 236–244
- stochastic training, 224
- stream function, 557
- streamflow, 78, 85, 212, 511, 514
- stride, 501
- structured datasets, 497
- Student t -distribution (*see t*-distribution)
- Student t -test (*see t*-test)
- sub-grid processes, 517
- subsurface temperature, 214
- super-resolution, 510
- supervised learning, 15–17, 54, 173
- support vector, 455, 457, 460
- support vector machine (SVM), 440, 453–463
- support vector regression (SVR), 462–463
 - least squares, 448
- surface gravity wave, 555
- surface wind speed, 210
- suspended sediment, 209
- swish function, 177–178, 496
- Sydney airport, 438
- symmetric function, 445
- synapse, 175
- synthetic aperture radar (SAR), 211–212
- t -distribution, 95–97, 109, 117
- t -test, 104–110, 533
 - correlation, 109–110
 - cross-validated, 533

- dependent paired samples, 107–109
- independent two-sampled, 105–109
- one-sampled, 105
- tail distribution, 23, 103
- target, 15, 55, 176, 181
- teleconnection, 171, 312–315, 367–369
- temperature, 34, 38, 43, 45, 108–109, 117–118, 149, 160
- temporal convolutional network (TCN), 513–514
- tercile, 31
- terminology
 - trends, 3, 569
- test data, 6, 11, 258
- Thematic Mapper, 209
- threat score (TS), 520
- threshold parameter, 176
- tidal harmonic analysis, 396
- Tikhonov regularization, 161
- time-averaging, 404–405
- topographic error (TE), 345
- tornado, 213
- Toronto, Ontario, 44, 45, 171
- total least squares (TLS), 170
- training data, 6, 256
- training error, 251
- transfer learning, 498–499
- transformation
 - Box–Cox, 94
 - Johnson, 95
 - Yeo–Johnson, 94–95
- transforming data, 93–95
- tree (*see* decision tree)
- trend removal, 382
- trend test, 125–126
- tropical cyclone, 213
- Tropical Ocean Global Atmosphere (TOGA)
 - program, 312
- true skill statistics, 526
- turbulent flux, 553–554
- two-tailed test, 104–105, 108–109, 116
- Type I error, 103
- Type II error, 103

- U-net, 508–510, 546
- underfit, 11, 245
- UNet 3+, 510
- UNet++, 510, 546
- unit (in neural network model), 176
- universal approximator, 186, 187

- unmanned aerial system (UAS), 487
- unstructured datasets, 497
- unsupervised learning, 15–17, 54, 173, 196
- upsampling, 507
- urban land cover, 209

- validation, 11, 224, 253–255
 - cross-validation, 255–258
 - double cross-validation, 258
 - generalized cross-validation, 258
 - leave-one-out, 257–258
 - retroactive, 538
- validation data, 11, 164, 172, 191, 256
- Vancouver, BC, 34, 38–39, 42, 45, 78, 82, 89, 96, 108–109, 117–118, 144, 149, 160, 168, 171
- vanishing gradient problem, 494
- variance, 25
 - online algorithm, 28
 - population variance, 25
 - sample variance, 25
- variance compensation factor, 381
- variance error, 250–251, 264–266
- variance matrix, 26–27
- variational autoencoder, 348
- variational data assimilation, 560–567
- varimax rotation, 308–315
- verification data, 258
- Victoria, BC, 45, 108–109, 117–118
- visibility, 144, 149, 160–162
- von Mises distribution, 82–83
- voting, 432

- water wave, 85, 555
- watershed, 212
- wavelet, 390–392
- weak learner, 262, 473
- Weibull distribution, 13, 84
- weight, 175, 222
- weight penalty, 163, 251–254, 359–360, 459, 462, 477
- Welch method, 384
- Welford algorithm, 28
- Western Pacific pattern, 313
- whale, 498, 507
- white noise, 410
- Wilcoxon rank sum test (*see* Wilcoxon–Mann–Whitney test)
- Wilcoxon signed-rank test, 114–115

Index

627

- Wilcoxon–Mann–Whitney (WMW) test, 111–114
- wind direction, 172, 211
- wind speed, 34, 43, 45, 89, 96, 160, 168, 185, 211, 212
- wind stress, 367
- wind wave, 555–556
- window, 376–381
 - cosine-tapered, 379
 - Hamming, 379
 - Hann, 379
- Winnipeg, Manitoba, 135
- Woodbury matrix identity, 156
- Yeo–Johnson transformation, 94–95
- yin and yang, 4
- Yule–Walker equations, 413
- z-score, 25, 71, 118
- zero-padding, 386
- zonal, 362