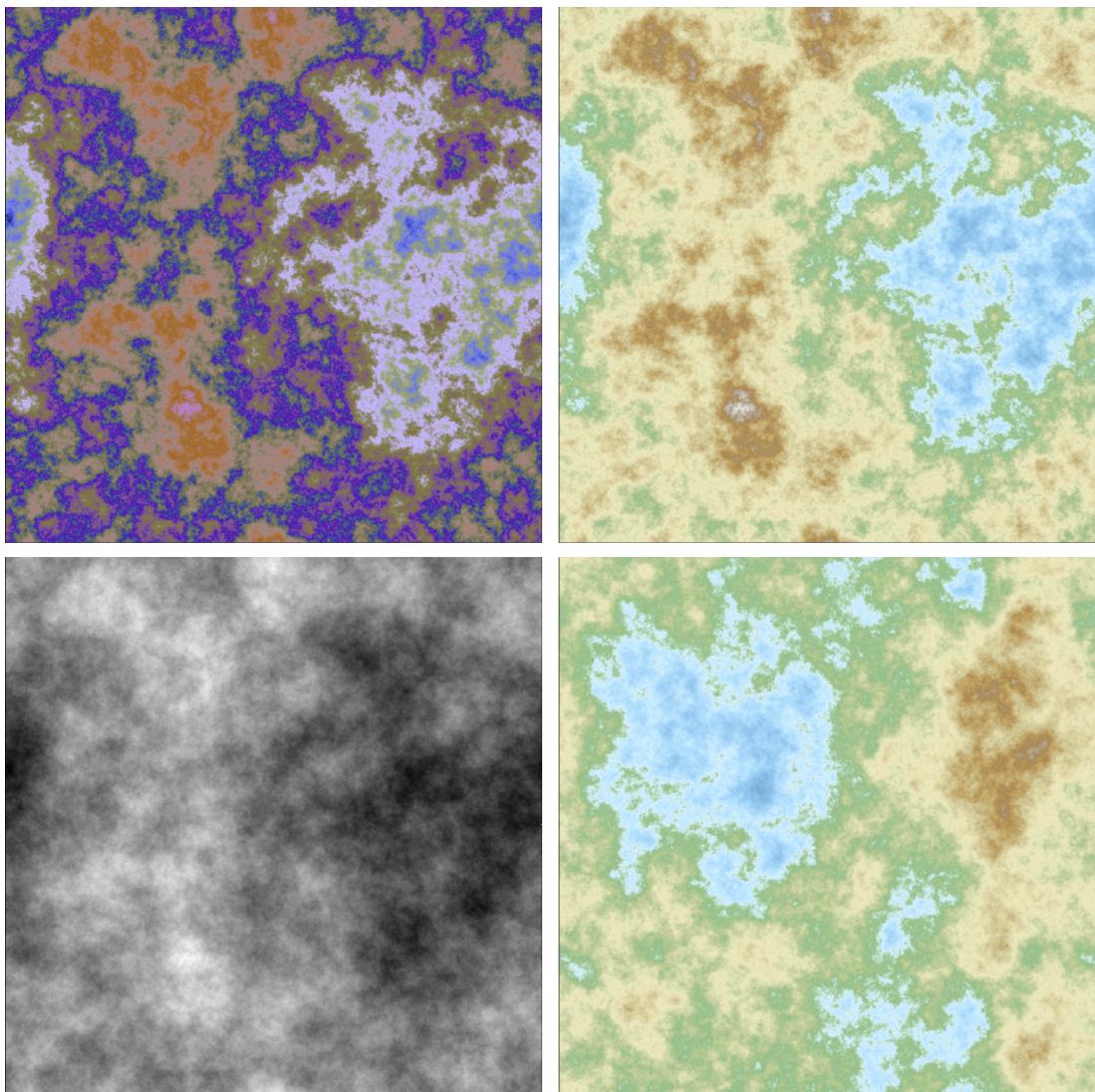

Synthetic Data and Generative AI



Preface

This book covers the foundations of machine learning, with modern approaches to solving complex problems and the systematic generation and use of synthetic data. Emphasis is on scalability, automation, testing, optimizing, and interpretability (explainable AI). For instance, regression techniques – including logistic and Lasso – are presented as a single method, without using advanced linear algebra. There is no need to learn 50 versions when one does it all and more. Confidence regions and prediction intervals are built using parametric bootstrap, without statistical models or probability distributions. Models (including generative models and mixtures) are mostly used to create rich synthetic data to test and benchmark various methods.

Topics covered include clustering and classification, GPU machine learning, ensemble methods including an original boosting technique, elements of graph modeling, deep neural networks, auto-regressive and non-periodic time series, Brownian motions and related processes, simulations, interpolation, random numbers, natural language processing (smart crawling, taxonomy creation and structuring unstructured data), computer vision (shapes generation and recognition), curve fitting, cross-validation, goodness-of-fit metrics, feature selection, curve fitting, gradient methods, optimization techniques and numerical stability.

Several chapters focus on synthetic data, agent-based modeling and GIS applications: fractal-like terrain generation with the diamond-square algorithm, disaggregation of ocean tides time series, geospatial interpolation of temperatures in the Chicago area, and synthetic star clusters evolving over time and bound by gravity. The latter provides great insights to explore the past and future of our universe or studying collision graphs. It also allows you to explore alternative universes, for instance with negative masses. Chapters 15 and 16 are more advanced and may be skipped in introductory classes. The former focuses on point process applications, while the later focuses on applications a machine learning methods to discover new insights in a famous mathematical conjecture: the Riemann Hypothesis. Section 17.7.2 illustrates the use of copulas to produce synthetic data, applied to a well-known insurance dataset.

Methods are accompanied by enterprise-grade Python code, replicable datasets and visualizations, including data animations (gifs, videos, even sound done in Python). The code uses various data structures and library functions sometimes with advanced options. It constitutes a solid introduction to scientific programming. The code, datasets, spreadsheets and data visualizations are also on GitHub, spread across the following repositories: [Machine Learning](#), [Point Processes](#), [Visualizations](#), and [Experimental Math](#). Chapters are mostly independent from each other, allowing you to read in random order. A glossary, index and numerous cross-references make the navigation easy and unify all the chapters.

The style is very compact, getting down to the point quickly, and suitable to business professionals. Jargon and arcane theories are absent, replaced by simple English to facilitate the reading by non-experts, and to help you discover topics usually made inaccessible to beginners. While state-of-the-art research is presented in all chapters, the prerequisites to read this book are minimal: an analytic professional background, or a first course in calculus and linear algebra. The original presentation avoids all unnecessary math and statistics, yet without eliminating advanced topics. Finally, this book is the main reference for my course on intuitive machine learning. For details about the classes, see [here](#).

About the Author

Vincent Granville is a pioneering data scientist and machine learning expert, co-founder of Data Science Central (acquired by a publicly traded company in 2020), founder of [MLTechniques.com](#), former VC-funded executive, author and patent owner. Vincent's past corporate experience includes Visa, Wells Fargo, eBay, NBC, Microsoft, and CNET.



Vincent is also a former post-doc at Cambridge University, and the National Institute of Statistical Sciences (NISS). He published in *Journal of Number Theory*, *Journal of the Royal Statistical Society* (Series B), and *IEEE Transactions on Pattern Analysis and Machine Intelligence*. He is also the author of multiple books, available [here](#). He lives in Washington state, and enjoys doing research on stochastic processes, dynamical systems, experimental math and probabilistic number theory.

Contents

List of Figures	9
List of Tables	11
1 Machine Learning Cloud Regression and Optimization	12
1.1 Introduction: circle fitting	12
1.1.1 Previous versions of my method	13
1.2 Methodology, implementation details and caveats	14
1.2.1 Solution, R-squared and backward compatibility	14
1.2.2 Upgrades to the model	15
1.3 Case studies	16
1.3.1 Logistic regression, two ways	16
1.3.2 Ellipsoid and hyperplane fitting	17
1.3.2.1 Curve fitting: 250 examples in one video	17
1.3.2.2 Confidence region for the fitted ellipse: application to meteorite shapes	18
1.3.2.3 Python code	19
1.3.3 Non-periodic sum of periodic time series: ocean tides	25
1.3.3.1 Numerical instability and how to fix it	26
1.3.3.2 Python code	27
1.3.4 Fitting a line in 3D, unsupervised clustering, and other generalizations	28
1.3.4.1 Example: confidence region for the cluster centers	29
1.3.4.2 Exact solution and caveats	30
1.3.4.3 Comparison with K-means clustering	31
1.3.4.4 Python code	33
1.4 Connection to synthetic data: meteorites, ocean tides	35
2 A Simple, Robust and Efficient Ensemble Method	36
2.1 Introduction	36
2.2 Methodology	37
2.2.1 How hidden decision trees (HDT) work	37
2.2.2 NLP case study: summary and findings	38
2.2.3 Parameters	39
2.2.4 Improving the methodology	39
2.3 Implementation details	39
2.3.1 Correcting for bias	39
2.3.1.1 Time-adjusted scores	40
2.3.2 Excel spreadsheet	40
2.3.3 Python code and dataset	40
2.4 Model-free confidence intervals and perfect nodes	44
2.4.1 Interesting asymptotic properties of confidence intervals	44
3 Gentle Introduction to Linear Algebra – Synthetic Time Series	46
3.1 Power of a matrix	46
3.2 Examples, generalization, and matrix inversion	47
3.2.1 Example with a non-invertible matrix	48
3.2.2 Fast computations	48
3.2.3 Square root of a matrix	48
3.3 Application to machine learning problems	49
3.3.1 Markov chains	49
3.3.2 Time series: auto-regressive processes	49

3.3.3	Linear regression	50
3.4	Mathematics of auto-regressive time series	50
3.4.1	Simulations: curious fractal time series	51
3.4.1.1	White noise: Fréchet, Weibull and exponential cases	51
3.4.1.2	Illustration	51
3.4.2	Solving Vandermonde systems: a numerically stable method	52
3.5	Math for Machine Learning: Must-Read Books	53
4	The Art of Visualizing High Dimensional Data	54
4.1	Introduction	54
4.2	Applications	55
4.2.1	Spatial time series	55
4.2.2	Prediction intervals in any dimensions	55
4.2.3	Supervised classification of an infinite dataset	56
4.2.3.1	Machine learning perspective	56
4.2.3.2	Six challenging problems	57
4.2.3.3	Mathematical background: the Riemann Hypothesis	57
4.2.3.4	Partial solutions to the six challenging problems	58
4.2.4	Algorithms with chaotic convergence	59
4.3	Python code	59
4.3.1	Path simulation	59
4.3.2	Visual convergence analysis in 2D	62
4.3.3	Supervised classification	63
4.4	Visualizations	66
5	Fast Classification and Clustering via Image Convolution Filters	69
5.1	Introduction	69
5.2	Generating the synthetic data	70
5.2.1	Simulations with logistic distribution	70
5.2.2	Mapping the raw observations onto an image bitmap	71
5.3	Classification and unsupervised clustering	71
5.3.1	Supervised classification based on convolution filters	72
5.3.2	Clustering based on histogram equalization	72
5.3.3	Fractal classification: deep neural network analogy	73
5.3.4	Generalization to higher dimensions	74
5.3.5	Towards a very fast implementation	74
5.4	Python code	75
5.4.1	Fractal classification	76
5.4.2	GPU classification and clustering	78
5.4.3	Home-made graphic library	80
6	Shape Classification and Synthetization via Explainable AI	83
6.1	Introduction	83
6.2	Mathematical foundations	83
6.3	Shape signature	84
6.3.1	Weighted centroid	84
6.3.2	Computing the signature	85
6.3.3	Example	86
6.4	Shape comparison	86
6.4.1	Shape classification	87
6.5	Application	87
6.6	Exercises	88
7	Synthetic Data, Interpretable Regression, and Submodels	89
7.1	Introduction	89
7.2	Synthetic data sets and the spreadsheet	90
7.2.1	Correlation structure	90
7.2.2	Standardized regression	91
7.2.3	Initial conditions	91
7.2.4	Simulations and Excel spreadsheet	91
7.3	Damping schedule and convergence acceleration	92
7.3.1	Spreadsheet implementation	92

7.3.2	Interpretable regression with no overfitting	93
7.3.3	Adaptive damping	93
7.4	Performance assessment on synthetic data	93
7.4.1	Results	94
7.4.2	Distribution-free confidence intervals	96
7.4.2.1	Parametric bootstrap	97
7.5	Feature selection	97
7.5.1	Combinatorial approach	97
7.5.2	Stepwise approach	98
7.6	Conclusion	99
8	From Interpolation to Fuzzy Regression	101
8.1	Introduction	101
8.2	Original version	102
8.3	Full, non-linear model in higher dimensions	102
8.3.1	Geometric proximity, weights, and numerical stability	103
8.3.2	Predicted values and prediction intervals	103
8.3.3	Illustration, with spreadsheet	104
8.3.3.1	Output fields	105
8.4	Results	105
8.4.1	Performance assessment	105
8.4.2	Visualization	106
8.4.3	Amplitude restoration	106
8.5	Exercises	107
8.6	Python source code and datasets	108
9	New Interpolation Methods for Synthetization and Prediction	112
9.1	First method	112
9.1.1	Example with infinite summation	113
9.1.2	Applications: ocean tides, planet alignment	114
9.1.3	Problem in two dimensions	115
9.1.4	Spatial interpolation of the temperature dataset	116
9.2	Second method	118
9.2.1	From unstable polynomial to robust orthogonal regression	119
9.2.2	Using orthogonal functions	119
9.2.3	Application to regression	119
9.3	Python code	120
9.3.1	Time series interpolation	120
9.3.2	Geospatial temperature dataset	123
9.3.3	Regression with Fourier series	126
10	Detecting Subtle Departures from Randomness	128
10.1	Introduction	128
10.2	Pseudo-random numbers	129
10.2.1	Strong pseudo-random numbers	129
10.2.1.1	New test of randomness for PRNGs	130
10.2.1.2	Theoretical background: the law of the iterated logarithm	130
10.2.1.3	Connection to the Generalized Riemann Hypothesis	130
10.2.2	Testing well-known sequences	131
10.2.2.1	Reverse-engineering a pseudo-random sequence	132
10.2.2.2	Illustrations	133
10.3	Python code	135
10.3.1	Fixes to the faulty random function in Python	135
10.3.2	Prime test implementation to detect subtle flaws in PRNG's	135
10.3.3	Special formula to compute 10 million digits of $\sqrt{2}$	138
10.4	Military-grade PRNG Based on Quadratic Irrationals	141
10.4.1	Fast algorithm rooted in advanced analytic number theory	141
10.4.2	Fast PRNG: explanations	142
10.4.3	Python code	142
10.4.4	Computing a digit without generating the previous ones	144
10.4.5	Security and comparison with other PRNGs	144
10.4.5.1	Important comments	144

10.4.6 Curious application: a new type of lottery	145
11 Some Unusual Random Walks	146
11.1 Symmetric unbiased constrained random walks	146
11.1.1 Three fundamental properties of pure random walks	146
11.1.2 Random walks with more entropy than pure random signal	147
11.1.2.1 Applications	147
11.1.2.2 Algorithm to generate quasi-random sequences	148
11.1.2.3 Variance of the modified random walk	148
11.1.3 Random walks with less entropy than pure random signal	149
11.2 Related stochastic processes	150
11.2.1 From Brownian motions to clustered Lévy flights	150
11.2.2 Integrated Brownian motions and special auto-regressive processes	151
11.3 Python code	152
11.3.1 Computing probabilities and variances attached to S_n	152
11.3.2 Path simulations	153
12 Empirical Optimization with Divergent Fixed Point Algorithm	155
12.1 Introduction	155
12.1.1 The problem, with illustration	156
12.2 Non-converging fixed-point algorithm	157
12.2.1 Trick leading to intuitive solution	157
12.2.2 Root detection: method and parameters	157
12.2.3 Case study: factoring a product of two large primes	158
12.3 Generalization with synthetic random functions	158
12.3.1 Example	160
12.3.2 Connection to the Poisson-binomial distribution	161
12.3.2.1 Location of next root: guesstimate	161
12.3.2.2 Integer sequences with high density of primes	161
12.3.3 Python code: finding the optimum	162
12.4 Smoothing highly chaotic curves	163
12.4.1 Python code: smoothing	163
12.5 Connection to synthetic data: random functions	166
13 Synthetic Terrain Generation and AI-generated Art	167
13.1 Introduction	167
13.2 Terrain generation and the evolutionary process	169
13.2.1 Morphing and non-linear palette operations	169
13.2.2 The diamond-square algorithm	169
13.2.3 The evolutionary process	170
13.2.4 Finding optimum parameters	170
13.2.5 Mimicking real terrain: the synthesis step	170
13.3 Python code	171
13.3.1 Producing data videos with four sub-videos in parallel	171
13.3.2 Main program	172
13.4 AI-generated art with 3D contours	176
13.4.1 Python code using Matplotlib	177
13.4.2 Python code using Plotly	178
13.4.3 Tips to quickly solve new problems	179
14 Synthetic Star Cluster Generation with Collision Graphs	180
14.1 Introduction	180
14.2 Model parameters and simulation results	181
14.2.1 Explanation of color codes	181
14.2.2 Detailed description of top parameters	181
14.2.3 Interesting parameter sets	182
14.3 Analysis of star collisions and collision graph	183
14.3.1 Weighted directed graphs: visualization with NetworkX	184
14.3.2 Interesting findings: how the universe got started	184
14.4 Animated data visualizations	185
14.5 Python code and computational issues	186
14.5.1 Simulating the real and synthetic universes	186

14.5.2	Visualizing collision graphs	190
15	Perturbed-Lattice Point Process: Inference, Nearest Neighbor Graph	192
15.1	Perturbed lattices: definition and properties	192
15.1.1	Point counts distribution	193
15.1.2	Periodicity and amplitude of point count expectations	193
15.1.3	Testing the independence of point counts	194
15.1.3.1	Results and Interpretation	195
15.1.3.2	About the Spreadsheet	196
15.2	Cluster processes and nearest neighbor graphs	196
15.2.1	Synthetic, semi-rigid cluster structures	196
15.2.2	Python code to generate cluster processes	198
15.2.3	References on cluster processes	198
15.2.4	Superimposed perturbed lattices: an alternative to mixture models	199
15.2.4.1	Hexagonal lattice, nearest neighbors	200
15.2.4.2	Exercises: nearest neighbor graphs, size of connected components	201
15.2.4.3	Python code to compute connected components	202
15.3	Statistical inference for point processes	204
15.3.1	Estimation of Core Parameters	204
15.3.1.1	Intensity	205
15.3.1.2	Scaling factor	205
15.3.1.3	Alternative estimation method	205
15.3.2	Spatial statistics, nearest neighbors, clustering	206
15.3.2.1	Inference for two-dimensional processes	206
15.3.2.2	Other possible tests	206
15.3.2.3	Rayleigh test	207
15.3.2.4	Exercises	208
15.4	Special topics	209
15.4.1	Minimum contrast estimation and explainable AI	209
15.4.2	Model identifiability, hard-to-detect patterns	210
15.4.2.1	Stochastic residues	210
15.4.3	Hidden model and random permutations	210
15.4.4	Retrieving the F distribution	212
15.4.4.1	Theoretical values obtained by simulations	212
15.4.4.2	Retrieving F from the interarrival times distribution	213
15.4.5	Record distances between an observed point and its vertex	213
15.4.5.1	Distribution of records	214
15.4.5.2	Distribution of arrival times for records	215
16	New Perspective on the Riemann Hypothesis	216
16.1	Introduction	216
16.1.1	Key concepts and terminology	217
16.1.2	Orbits and holes	217
16.1.3	Industrial Applications	217
16.2	Euler products	218
16.2.1	Finite Euler Products	218
16.2.1.1	Generalization using Dirichlet characters	219
16.2.2	Infinite Euler products	220
16.2.2.1	Special products	220
16.2.2.2	Probabilistic properties and conjectures	221
16.3	Finite Dirichlet series and generalizations	222
16.3.1	Finite Dirichlet series	222
16.3.2	Non-trivial cases with infinitely many primes and a hole	224
16.3.2.1	Sums of two cubes, or cuban primes	224
16.3.2.2	Primes associated to elliptic curves	224
16.3.2.3	Analytic continuation, convergence, and functional equation	225
16.3.2.4	Hybrid Dirichlet-Taylor series	225
16.3.3	Riemann Hypothesis with cosines replaced by wavelets	226
16.3.4	Riemann Hypothesis for Beurling primes	227
16.3.5	Stochastic Euler products	228
16.4	Exercises	229

16.5	Python code	232
16.5.1	Computing the orbit of various Dirichlet series	232
16.5.2	Creating videos of the orbit	235
17	Misc Topics Including Copulas to Synthetize Data	238
17.1	The sound that data makes	238
17.1.1	From data visualizations to videos to data music	238
17.1.2	References	239
17.1.3	Python code	239
17.2	Data videos and enhanced visualizations in R	240
17.2.1	Cairo library to produce better charts	240
17.2.2	AV library to produce videos	241
17.3	Dual confidence regions	242
17.3.1	Case study	242
17.3.2	Standard confidence region	242
17.3.3	Dual confidence region	243
17.3.4	Simulations	243
17.3.5	Original problem with minimum contrast estimators	244
17.3.6	General shape of confidence regions	245
17.4	Fast feature selection based on predictive power	246
17.4.1	How cross-validation works	247
17.4.2	Measuring the predictive power of a feature	247
17.4.3	Efficient implementation	248
17.5	Natural language processing: taxonomy creation	249
17.5.1	Designing a keyword taxonomy	249
17.5.2	Fast clustering algorithm for keyword data	250
17.5.2.1	Computational complexity	250
17.5.2.2	Smart crawling of the whole Internet and a bit of graph theory	251
17.6	Automated detection of outliers and number of clusters	252
17.6.1	Black-box elbow rule to detect outliers	252
17.7	New developments in explainable AI and synthetic data	253
17.7.1	Sensitivity analysis, bias reduction and other uses of synthetic data	254
17.7.2	Using copulas to generate synthetic data	254
17.7.2.1	The insurance dataset: Python code and results	255
17.8	Advice to beginners	257
17.8.1	Getting started and learning how to learn	257
17.8.1.1	Getting help	258
17.8.1.2	Beyond Python	258
17.8.2	Automated data cleaning and exploratory analysis	258
17.8.3	Example of simple analysis: marketing attribution	259
17.8.4	Upcoming books and courses	259
Glossary		260
Bibliography		263
Index		268

List of Figures

1.1	Fitted ellipse (blue), given the training set (red) distributed around a partial arc	18
1.2	Confidence region in blue, $n = 30$ training set points; 50 training sets (left) vs 150 (right)	19
1.3	Three non-periodic time series made of periodic terms (see section 16.2.2.1)	25
1.4	Training set (red), validation set (orange), fitted curve (blue) and model (gray)	26
1.6	Biased confidence region for (θ_A^*, θ_B^*) ; same example as in Figure 1.5; true value is $(0.5, 1.0)$	29
1.5	Finding the two centers θ_A^*, θ_B^* in sample 39; $n = 1000$	30
1.7	Challenging mixture, requiring $p_A = 3, p_B = 1$ to identify the two cluster centers	31
2.1	Output from the Excel version of HDT	41
3.1	AR models, classified based on the types of roots of the characteristic polynomial	52
4.1	Scatterplot observations vs. predicted values, with prediction intervals (in any dimension)	66
4.2	Comets orbiting the sun: simulation	66
4.3	Comets orbiting the sun: snapshot in time	67
4.4	Three orbits of $\eta(\sigma + it)$: $\sigma = 0.5$ (red), 0.75 (blue) and 1.25 (yellow)	67
4.5	Sample orbit points of $\eta(\sigma + it)$: $\sigma = 0.5$ (red), 0.75 (blue) and 1.25 (yellow)	67
4.6	Sample orbit points of $\eta(\sigma + it)$: $\sigma = 0.5$ (red), 0.75 (blue) and 1.25 (yellow)	68
4.7	Raw orbit points of $\eta(\sigma + it)$: $\sigma = 0.5$ (red), 0.75 (blue) and 1.25 (yellow)	68
4.8	Convergence of partial sums of $\eta(z)$, for six $z = \sigma + it$ in the complex plane	68
5.1	Special interlacing of 4 lattice processes with $s = 0$	71
5.2	Classification of left dataset; $s = 0.15, w = 10$. One loop (middle) vs 3 (right).	72
5.3	Clustering of left dataset; $s = 0.15$, 3 loops, $w = 10$ (middle) vs 20 (right).	73
5.4	Classification ($w = 10$) and clustering ($w = 20$); $s = 0.05$, three loops.	73
5.5	Fractal classification, $s = 0.15$. Loop 6, 250 and 400.	74
5.6	Fractal classification, $s = 0.05$.Loop: 6 and 60.	74
5.7	Fast (left) vs standard method (right), 3 loops, $s = 0.15, w = 10$	75
5.8	Fast method, $s = 0.05, w = 20$. Three loops (middle), one loop (right).	75
6.1	Comparing two shapes	84
6.2	Weighted centroid, shape signature	85
6.3	Weight function used in Figure 6.2	86
6.4	Another interesting shape	87
7.1	Regression coefficients oscillating when using adaptive damping	94
7.2	Convergence of regression coefficients (left) and distribution of residual error (right)	95
7.3	Goodness-of-fit: training set (right) versus validation set (left)	95
8.1	Fuzzy regression with prediction intervals, original version, 1D	102
8.2	Fuzzy regression with prediction intervals, full model, 2D	104
8.3	Scatterplots: median vs weighted method, on validation (left) vs training set (right)	106
8.4	Dirichlet eta function (real part, bottom) and interpolation error (top)	108
9.1	Interpolating the real part of $\zeta(\frac{1}{2} + it)$ based on orange points	113
9.2	Tides at Dublin (5-min data), with 80 mins between interpolating nodes	116
9.3	Temperature data: interpolation with my method (observed values at dots)	117
9.4	My method: round dots represent observed values, “+” are interpolated	117
9.5	Temperature dataset: interpolation using ordinary kriging	118

10.1	Orbit of $L(z, \chi)$ at $\sigma = \frac{1}{2}$, with $0 < t < 200$ and $\chi = \chi_4$ (left) versus pseudo-random χ (right)	131
10.2	$L_3^*(n)$ test statistic for four sequences: Python[200] and SQRT[90,91] fail	133
10.3	$ L_3(n) $ test statistic for four sequences: Python[200] and SQRT[90,91] fail	133
10.4	Correlations are computed on sequences consisting of 300 binary digits	145
11.1	Typical path S_n with $0 \leq n \leq 50,000$ for four types of random walks	147
11.2	$\delta_n = 1 - \text{Var}[S_{n+1}] + \text{Var}[S_n]$ for four types of random walks, with $0 \leq n \leq 5000$	148
11.3	Same as Figure 11.2, using a more aesthetic but less meaningful chart type	149
11.4	Clustered Brownian process	151
11.5	AR models, classified based on the types of roots of the characteristic polynomial	152
12.1	Function $f(b)$ as a better alternative to $g(b)$ in Figure 12.2. Root at $b = 3083$	156
12.2	Function $g(b) = 2 - \cos(2\pi b) - \cos(2\pi a/b)$, with $a = 3083 \times 7919$	156
12.3	Transformed function f_3 , amplifying the root at $b = 3083$	157
12.4	Signal strength ρ_n , first 130 fixed-point iterations; $n = 31$ leads to a root.	160
12.5	(b_n, ρ_n) plot. Yellow and orange dots linked to roots.	160
12.6	Signal strength ρ_n , first 130 fixed-point iterations; $n = 87$ leads to a root.	160
12.7	Random function from section 12.3.1, with root at $b = 5646$	163
13.1	Six frames from the terrain video, each one containing four images	168
13.2	Contour plot, 3D mixture model, produced with Plotly	176
13.3	Same as Figure 13.2, produced with Matplotlib	177
14.1	Collisions graph for the biggest star eater (star 47) in video 7	184
14.2	Summary statistics for the whole collision structure: the X axis represents the time	185
14.3	Snapshots of universe 4 (left) and universe 7 (right)	186
15.1	Period and amplitude of $\phi_\tau(t)$; here $\tau = 1, \lambda = 1.4, s = 0.3$	194
15.2	A new test of independence (R-squared version)	194
15.3	Radial cluster process ($s = 0.2, \lambda = 1$) with centers in blue; zoom in on the left	197
15.4	Radial cluster process ($s = 2, \lambda = 1$) with centers in blue; zoom in on the left	197
15.5	Manufactured marble lacking true lattice randomness (left)	197
15.6	Four superimposed Poisson-binomial processes: $s = 0$ (left), $s = 5$ (right)	200
15.7	Rayleigh test to assess if a point distribution matches that of a Poisson process	208
15.8	Realization of a 5-interlacing with $s = 0.15$ and $\lambda = 1$: original (left), modulo $2/\lambda$ (right)	211
15.9	Locally random permutation σ ; $\tau(k)$ is the index of X_k 's closest neighbor to the right	211
15.10	Each arrow links a point (blue) to its vertex (red): $s = 0.2$ (left), $s = 1$ (right)	214
15.11	Distance between a point and its vertex ($\lambda = s = 1$)	215
16.1	Three orbits ($\sigma = 0.5, 0.75, 1.25$) with finite Euler product: $P = \{2, 3\}$ (left) vs $\{2, 3, 5\}$ (right)	219
16.2	Distance between orbit and location $(c, 0)$ depending on t on the X-axis	221
16.3	Distance between orbit and location $(c, 0)$ depending on t on the X-axis	221
16.4	Distance between orbit and location $(c, 0)$ depending on t on the X-axis	221
16.5	Four orbits where the “hole” (repulsion basin) is apparent	223
16.6	Three orbits with “hole” closer to the origin, showing impact of $\beta > \frac{1}{2}$ and larger n	223
16.7	Orbit of Dirichlet eta $\eta(z)$ when cosines are replaced by other periodic functions	227
17.1	Data linked to the melody: red curve for note frequencies, blue curve for note durations	239
17.2	R plot before Cairo (left), and after (right)	240
17.3	Intermediate (left) and last frame (right) of the video	241
17.4	Example of 90% dual confidence region for (p, q)	243
17.5	Minimum contrast estimation for (λ, s) using (p, q) as proxy stats	244
17.6	Non-elliptic confidence regions with various confidence levels	245
17.7	Elbow rule (right) finds $m = 3$ clusters in Brownian motion (left)	253

List of Tables

1.1	Estimated ellipse parameters vs true values ($n = 30$), for shape in Figure 1.2	19
1.2	First and last step of <code>curve_fitting</code> , approaching the model.	27
1.3	MSE for different methods and θ s, same data set as in Figure 1.5	32
1.4	MSE for different methods and θ s, same data set as in Figure 1.7	32
2.1	List of potential features to use in the model	37
2.2	Statistics for selected HDT nodes (Excel version)	40
2.3	Order of magnitude for the expectation and standard deviation of the range R_n	44
3.1	Characteristic polynomials used in the simulations	51
7.1	Regression coefficients and performance metrics r, s based on methodology	96
7.2	Correlation matrix	96
7.3	Best performance given m (number of features)	97
7.4	Feature comparison table (top 32 feature combinations)	99
7.5	Feature comparison table (bottom 31 feature combinations)	100
8.1	R -squared ρ^2 and slope β , on training and validation sets, median vs weighted	106
10.1	$L_3^*(n)$, for various sequences ($n = 20,000$); “Fail” means failing the prime test	134
12.1	High ρ_n at iterations $n = 31$ and $n = 127$ points to roots 3083 and 7919	159
14.1	Description of top parameters used in the star cluster simulator	182
14.2	Eight selected parameter sets covering various situations	183
15.1	Variance attached to F_s , as a function of s	193
15.2	Poisson process ($s = \infty$) versus $s = 39.85$	213
17.1	Extract of the mapping table used to recover (λ, s) from (p, q)	245
17.2	Eight bins: 2 features (A, B) times 2 outcomes (Good/Bad)	247
17.3	Amount of data collected at each level, when crawling the Internet	251
17.4	Comparing real data with two different synthetic copies	255

Glossary

Autoregressive process	Auto-correlated time series , as described in section 3.4. Time-continuous versions include Gaussian processes and Brownian motions , while random walks are a discrete example; two-dimensional versions exist. These processes are essentially integrated white noise . See pages 49, 97, 151
Binning	Feature binning consists of aggregating the values of a feature into a small number of bins, to avoid overfitting and reduce the number of nodes in methods such as naive Bayes , neural networks , or decision trees . Binning can be applied to two or more features simultaneously. I discuss optimum binning in this book. See pages 37, 73, 247
Boosted model	Blending of several models to get the best of each one, also referred to as ensemble methods . The concept is illustrated with hidden decision trees in this book. Other popular examples are gradient boosting and AdaBoost . See pages 36, 260
Bootstrapping	A data-driven, model-free technique to estimate parameter values, to optimize goodness-of-fit metrics. Related to resampling in the context of cross-validation . In this book, I discuss parametric bootstrap on synthetic data that mimics the actual observations. See pages 15, 96, 208, 260
Confidence Region	A confidence region of level γ is a 2D set of minimum area covering a proportion γ of the mass of a bivariate probability distribution. It is a 2D generalization of confidence intervals . In this book, I also discuss dual confidence regions – the analogous of credible regions in Bayesian inference. See pages 12, 15, 18, 20, 29, 205, 206, 242, 245
Cross-validation	Standard procedure used in bootstrapping , and to test and validate a model, by splitting your data into training and validation sets . Parameters are estimated based on training set data. An alternative to cross-validation is testing your model on synthetic data with known response. See pages 15, 37, 93, 99, 183, 247, 260
Decision trees	A simple, intuitive non-linear modeling techniques used in classification problems. It can handle missing and categorical data, as well as a large number of features, but requires appropriate feature binning. Typically one blends multiple binary trees each with a few nodes , to boost performance. See pages 36, 37, 39, 41, 260, 261
Dimension reduction	A technique to reduce the number of features in your dataset while minimizing the loss in predictive power. The most well known are principal component analysis and feature selection to maximize goodness-of-fit metrics. See pages 12, 16, 261, 262
Empirical distribution	Cumulative frequency histogram attached to a statistic (for instance, nearest neighbor distances), and based on observations. When the number of observations tends to infinity and the bin sizes tend to zero, this step function tends to the theoretical cumulative distribution function of the statistic in question. See pages 16, 96, 120, 129, 192, 195, 201, 207, 212, 214, 226, 254
Ensemble methods	A technique consisting of blending multiple models together, such as many decision trees with logistic regression , to get the best of each method and outperform each method taken separately. Examples include boosting , bagging, and AdaBoost. In this book, I discuss hidden decision trees . See pages 36, 83, 260
Explainable AI	Automated machine learning techniques that are easy to interpret are referred to as interpretable machine learning or explainable artificial intelligence. As much as possible, the methods discussed in this book belong to that category. The goal is to design black-box systems less likely to generate unexpected results with unintended consequences. See pages 13, 35, 69, 74, 83, 90, 155, 171, 209, 253

Feature selection	Features – as opposed to the model response – are also called independent variables or predictors. Feature selection, akin to dimensionality reduction , aims at finding the minimum subset of variables with enough predictive power . It is also used to eliminate redundant features and find causality (typically using hierarchical Bayesian models), as opposed to mere correlations. Sometimes, two features have poor predictive power when taken separately, but provide improved predictions when combined together. See pages 12 , 15 , 37 , 94 , 97 , 238 , 246 , 260 , 262
Generative model	Bayesian Gaussian mixtures (GMM) combined with kernel density estimation and the EM algorithm is a classic modeling tool. In this book, I used m-interlacings instead. Generative adversarial networks (GAN) work as follows: the generator creates new observations and the discriminator tests whether the new observations are statistically indistinguishable from training set data. When this goal is achieved, the new observations is your synthetic data. In this book, new observations are generated with parametric bootstrap instead. See pages 35 , 52 , 99 , 166 , 167 , 169 , 176 , 183 , 262
Goodness-of-fit	A model fitting criterion or metric to assess how a model or sub-model fits to a dataset, or to measure its predictive power on a validation set . Examples include R-squared , Chi-squared, Kolmogorov-Smirnov, error rate such as false positives and other metrics discussed in this book. See pages 15 , 56 , 93 , 94 , 247 , 260 , 262
Gradient methods	Iterative optimization techniques to find the minimum of maximum of a function, such as the maximum likelihood . When there are numerous local minima or maxima, use swarm optimization . Gradient methods (for instance, stochastic gradient descent or Newton's method) assume that the function is differentiable. If not, other techniques such as Monte Carlo simulations or the fixed-point algorithm can be used. Constrained optimization involves using Lagrange multipliers . See pages 15 , 31 , 55 , 89
Graph structures	Graphs are found in decision trees , in neural networks (connections between neurons), in nearest neighbors methods (NN graphs), in hierarchical Bayesian models , and more. See pages 70 , 74 , 184 , 250 , 251
Hyperparameter	An hyperparameter is used to control the learning process: for instance, the dimension, the number of features, parameters, layers (neural networks) or clusters (clustering problem), or the width of a filtering window in image processing. By contrast, the values of other parameters (typically node weights in neural networks or regression coefficients) are derived via training. See pages 29 , 56 , 70 , 75 , 101 , 170 , 261
Link function	A link function maps a nonlinear relationship to a linear one so that a linear model can be fit, and then mapped back to the original form using the inverse function. For instance, the logit link function is used in logistic regression . Generalizations include quantile functions and inverse sigmoids in neural network to work with additive (linear) parameters. See pages 13 , 16 , 261
Logistic regression	A generalized linear regression method where the binary response (fraud/non-fraud or cancer/non-cancer) is modeled as a probability via the logistic link function. Alternatives to the iterative maximum likelihood solution are discussed in this book. See pages 16 , 33 , 36 , 40 , 260 , 261
Neural network	A blackbox system used for predictions, optimization, or pattern recognition especially in computer vision. It consists of layers, neurons in each layer, link functions to model non-linear interactions, parameters (weights associated to the connections between neurons) and hyperparameters . Networks with several layers are called deep neural networks . Also, neurons are sometimes called nodes. See pages 69 , 73 , 75 , 83 , 101 , 260 , 261
NLP	Natural language processing is a set of techniques to deal with unstructured text data, such as emails, automated customer support, or webpages downloaded with a crawler. The example discussed in section 17.5 deals with creating a keyword taxonomy based on parsing Google search results pages. See pages 36 , 249
Numerical stability	This issue occurring in unstable optimization problems typically with multiple minima or maxima, is frequently overlooked and leads to poor predictions or high volatility. It is sometimes referred to as ill-conditioned problems . I explain how to fix it in several examples in this book, for instance in section 3.4.2 . Not to be confused with numerical precision. See pages 12 , 14 , 59

Overfitting	Using too many unstable parameters resulting in excellent performance on the training set , but poor performance on future data or on the validation set . It typically occurs with numerically unstable procedures such as regression (especially polynomial regression) when the training set is not large enough, or in the presence of wide data (more features than observations) when using a method not suited to this situation. The opposite is underfitting. See pages 15, 92, 101, 254, 255, 260, 262
Predictive power	A metric to assess the goodness-of-fit or performance of a model or subset of features, for instance in the context of dimensionality reduction or feature selection . Typical metrics include R-squared , or confusion matrices in classification. See pages 38, 40, 44, 246, 248, 253, 261
R-squared	A goodness-of-fit metric to assess the predictive power of a model, measured on a validation set . Alternatives include adjusted R-squared, mean absolute error and other metrics discussed in this book. See pages 12, 15, 35, 56, 90, 93, 95, 97, 104, 261, 262
Random number	Pseudo-random numbers are sequences of binary digits, usually grouped into blocks, satisfying properties of independent Bernoulli trials. In this book, the concept is formally defined, and strong pseudo-number generators are built and used in computer-intensive simulations. See pages 29, 128, 135, 252
Regression methods	I discuss a unified approach to all regression problems in chapter 1. Traditional techniques include linear, logistic, Bayesian, polynomial and Lasso regression (to deal with numerical instability and overfitting), solved using optimization techniques, maximum likelihood methods, linear algebra (eigenvalues and singular value decomposition) or stepwise procedures. See pages 12, 13, 15, 16, 19, 27, 36, 40, 46, 50, 52, 56, 89, 95, 101, 108, 261, 262
Supervised learning	Techniques dealing with labeled data (classification) or when the response is known (regression). The opposite is unsupervised learning , for instance clustering problems. In-between, you have semi-supervised learning and reinforcement learning (favoring good decisions). The technique described in chapter 1 fits into unsupervised regression. Adversarial learning is testing your model against extreme cases intended to make it fail, to build better models. See pages 262
Synthetic data	Artificial data simulated using a generative model , typically a mixture model , to enrich existing datasets and improve the quality of training sets . Called augmented data when blended with real data. See pages 12, 13, 15, 17, 27, 29, 33, 35, 48, 52, 55, 69, 70, 75, 88, 94, 105, 112, 118, 141, 147, 155, 169, 176, 183, 243, 252, 254, 260
Tensor	Matrix generalization with three or more dimensions. A matrix is a two-dimensional tensor. A triple summation with three indices is represented by a three-dimensional tensor, while a double summation involves a standard matrix. See pages 69, 74
Training set	Dataset used to train your model in supervised learning . Typically, a portion of the training set is used to train the model, the other part is used as validation set . See pages 13, 15, 17, 20, 29, 36, 40, 56, 72, 88, 95, 101, 105, 183, 247, 260, 262
Validation set	A portion of your training set , typically 20%, used to measure the actual performance of your predictive algorithm outside the training set. In cross-validation and bootstrapping, the training and validation sets are split into multiple subsets to get a better sense of variations in the predictions. See pages 15, 27, 41, 56, 93, 101, 183, 247, 254, 255, 260, 261, 262

Bibliography

- [1] Weighted percentiles using numpy. *Forum discussion*, 2020. StackOverflow [\[Link\]](#). 101
- [2] Jan Ackmann et al. Machine-learned preconditioners for linear solvers in geophysical fluid flows. *Preprint*, pages 1–19, 2020. arXiv:2010.02866 [\[Link\]](#). 93
- [3] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. Wiley, fourth edition, 2016. [\[Link\]](#). 201
- [4] José M. Amigó, Roberto Dale, and Piergiulio Tempesta. A generalized permutation entropy for random processes. *Preprint*, pages 1–9, 2012. arXiv:2003.13728 [\[Link\]](#). 212
- [5] Luc Anselin. *Point Pattern Analysis: Nearest Neighbor Statistics*. The Center for Spatial Data Science, University of Chicago, 2016. Slide presentation [\[Link\]](#). 199
- [6] Adrian Baddeley. Spatial point processes and their applications. In Weil W., editor, *Stochastic Geometry. Lecture Notes in Mathematics*, pages 1–75. Springer, Berlin, 2007. [\[Link\]](#). 198
- [7] David Bailey and Richard Crandall. Random generators and normal numbers. *Experimental Mathematics*, 11, 2002. Project Euclid [\[Link\]](#). 144
- [8] N. Balakrishnan and C.R. Rao (Editors). *Order Statistics: Theory and Methods*. North-Holland, 1998. 201, 215
- [9] Christopher Beckham and Christopher Pal. A step towards procedural terrain generation with GANs. *Preprint*, pages 1–5, 2017. arXiv:1707.03383 [\[Link\]](#). 168
- [10] Rabi Bhattacharya and Edward Waymire. *Random Walk, Brownian Motion, and Martingales*. Springer, 2021. 146
- [11] Barbara Bogacka. *Lecture Notes on Time Series*. 2008. Queen Mary University of London [\[Link\]](#). 49
- [12] B. Bollobas and P. Erdős. Cliques in random graphs. *Mathematical Proceedings of the Cambridge Philosophical Society*, 80(3):419–427, 1976. [\[Link\]](#). 202
- [13] Miklos Bona. *Combinatorics of Permutations*. Routledge, second edition, 2012. 212
- [14] Peter Borwein, Stephen K. Choi, and Michael Coons. Completely multiplicative functions taking values in $\{-1, 1\}$. *Transactions of the American Mathematical Society*, 362(12):6279–6291, 2010. [\[Link\]](#). 220
- [15] Peter Borwein and Michael Coons. Transcendence of power series for some number theoretic functions. *Proceedings of the American Mathematical Society*, 137(4):1303–1305, 2009. [\[Link\]](#). 222
- [16] Oliver Bröker and Marcus J. Groteb. Sparse approximate inverse smoothers for geometric and algebraic multigrid. *Applied Numerical Mathematics*, 41(1):61–80, 2002. 90
- [17] H. M. Bui and M. B. Milinovich. Gaps between zeros of the Riemann zeta-function. *Quarterly Journal of Mathematics*, 69(2):402–423, 2018. [\[Link\]](#). 232
- [18] Bartłomiej Błaszczyzyn and Dhandapani Yogeshwaran. Clustering and percolation of point processes. *Preprint*, pages 1–20, 2013. Project Euclid [\[Link\]](#). 198
- [19] Bartłomiej Błaszczyzyn and Dhandapani Yogeshwaran. On comparison of clustering properties of point processes. *Preprint*, pages 1–26, 2013. arXiv:1111.6017 [\[Link\]](#). 198
- [20] Bartłomiej Błaszczyzyn and Dhandapani Yogeshwaran. Clustering comparison of point processes with applications to random geometric models. *Preprint*, pages 1–44, 2014. arXiv:1212.5285 [\[Link\]](#). 198
- [21] Oliver Chikumbo and Vincent Granville. Optimal clustering and cluster identity in understanding high-dimensional data spaces with tightly distributed points. *Machine Learning and Knowledge Extraction*, 1(2):715–744, 2019. 253
- [22] Keith Conrad. *L-functions and the Riemann Hypothesis*. 2018. 2018 CTNT Summer School [\[Link\]](#). 131, 217, 220, 225
- [23] Noel Cressie. *Statistic for Spatial Data*. Wiley, revised edition, 2015. 198

- [24] D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*. Springer, second edition, 2002. Volume 1 – Elementary Theory and Methods. [150](#)
- [25] D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*. Springer, second edition, 2014. Volume 2 – General Theory and Structure. [150](#)
- [26] Tilman M. Davies and Martin L. Hazelton. Assessing minimum contrast parameter estimation for spatial and spatiotemporal log-Gaussian Cox processes. *Statistica Neerlandica*, 67(4):355–389, 2013. [244](#)
- [27] Marc Deisenroth, A. Faisal, and Cheng Soon Ong. *Mathematics for Machine Learning*. Cambridge University Press, 2020. [\[Link\]](#). [53](#)
- [28] Harold G. Diamond and Wen-Bin Zhang. *Beurling Generalized Numbers*. American Mathematical Society, 2016. Mathematical Surveys and Monographs, Volume 213 [\[Link\]](#). [132](#), [228](#)
- [29] D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes – Volume I: Elementary Theory and Methods*. Springer, second edition, 2013. [199](#)
- [30] D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes – Volume II: General Theory and Structure*. Springer, second edition, 2014. [199](#)
- [31] David Coupier (Editor). *Stochastic Geometry: Modern Research Frontiers*. Wiley, 2019. [209](#)
- [32] Ding-Geng Chen (Editor), Jianguo Sun (Editor), and Karl E. Peace (Editor). *Interval-Censored Time-to-Event Data: Methods and Applications*. Chapman and Hall/CRC, 2012. [200](#)
- [33] Khaled Emam, Lucy Mosquera, and Richard Hoptroff. *Practical Synthetic Data Generation*. O'Reilly, 2020. [99](#)
- [34] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. In *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, volume 5, pages 17–61, 1960. [\[Link\]](#). [202](#)
- [35] Achim Zeileis et al. Colorspace: A toolbox for manipulating and assessing colors and palettes. *Preprint*, pages 1–45, 2019. arXiv:1903.06490 [\[Link\]](#) [\[R Library\]](#). [168](#)
- [36] Arash Farahmand. *Math 55 Lecture Notes*. 2021. University of Berkeley [\[Link\]](#). [48](#), [53](#)
- [37] W. Feller. On the Kolmogorov-Smirnov limit theorems for empirical distributions. *Annals of Mathematical Statistics*, 19(2):177–189, 1948. [\[Link\]](#). [201](#), [208](#)
- [38] Nikos Frantzikinakis. Ergodicity of the Liouville system implies the Chowla conjecture. *Preprint*, pages 1–41, 2016. arXiv [\[Link\]](#). [222](#)
- [39] P. M. Gauthier. Approximating the Riemann zeta-function by polynomials with restricted zeros. *Canadian Mathematical Bulletin*, 62(3):475–478, 2018. [\[Link\]](#). [232](#)
- [40] P. A. Van Der Geest. The binomial distribution with dependent Bernoulli trials. *Journal of Statistical Computation and Simulation*, pages 141–154, 2004. [\[Link\]](#). [147](#)
- [41] Stamatia Giannarou and Tania Stathaki. Shape signature matching for object identification invariant to image transformations and occlusion. 2007. ResearchGate [\[Link\]](#). [84](#)
- [42] Minas Gjoka, Emily Smith, and Carter Butts. Estimating clique composition and size distributions from sampled network data. *Preprint*, pages 1–9, 2013. arXiv:1308.3297 [\[Link\]](#). [202](#)
- [43] B.V. Gnedenko and A. N. Kolmogorov. *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, 1954. [151](#)
- [44] Manuel González-Navarrete and Rodrigo Lambert. Non-markovian random walks with memory lapses. *Preprint*, pages 1–14, 2018. arXiv [\[Link\]](#). [146](#)
- [45] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. [\[Link\]](#). [53](#)
- [46] Vincent Granville. Estimation of the intensity of a Poisson point process by means of nearest neighbor distances. *Statistica Neerlandica*, 52(2):112–124, 1998. [\[Link\]](#). [199](#)
- [47] Vincent Granville. *Applied Stochastic Processes, Chaos Modeling, and Probabilistic Properties of Numeration Systems*. MLTechniques.com, 2018. [\[Link\]](#). [132](#)
- [48] Vincent Granville. *Stochastic Processes and Simulations: A Machine Learning Perspective*. MLTechniques.com, 2022. [\[Link\]](#). [51](#), [59](#), [151](#), [161](#), [192](#), [193](#), [194](#), [196](#), [200](#), [202](#), [228](#), [232](#), [246](#), [259](#)
- [49] Vincent Granville, Mirko Krivanek, and Jean-Paul Rasson. Simulated annealing: A proof of convergence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16:652–656, 1996. [72](#)
- [50] Vincent Granville and Richard L Smith. Disaggregation of rainfall time series via Gibbs sampling. *NISS Technical Report*, pages 1–21, 1996. [\[Link\]](#). [107](#)
- [51] Kristen Grauman. Shape matching. 2008. University of Texas, Austin [\[Link\]](#). [87](#)
- [52] Hui Guo et al. Eyes tell all: Irregular pupil shapes reveal gan-generated faces. *Preprint*, pages 1–7, 2021. arXiv:2109.00162 [\[Link\]](#). [254](#)

- [53] Aurélien Géron. *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*. O'Reilly, third edition, 2023. [35](#)
- [54] Radim Halíř and Jan Flusser. Numerically stable direct least squares fitting of ellipses. *Preprint*, pages 1–8, 1998. [\[Link\]](#). [17](#), [19](#)
- [55] Peter Hall. *Introduction to the theory of coverage processes*. Wiley, 1988. [209](#)
- [56] Adam J. Harper. Moments of random multiplicative functions, II: High moments. *Algebra and Number Theory*, 13(10):2277–2321, 2019. [\[Link\]](#). [128](#), [228](#)
- [57] Adam J. Harper. Moments of random multiplicative functions, I: Low moments, better than squareroot cancellation, and critical multiplicative chaos. *Forum of Mathematics, Pi*, 8:1–95, 2020. [\[Link\]](#). [128](#), [130](#), [228](#)
- [58] Adam J. Harper. Almost sure large fluctuations of random multiplicative functions. *Preprint*, pages 1–38, 2021. arXiv [\[Link\]](#). [130](#), [222](#), [228](#)
- [59] K. Hartmann, J. Krois, and B. Waske. *Statistics and Geospatial Data Analysis*. Freie Universität Berlin, 2018. E-Learning Project SOGA [\[Link\]](#). [195](#)
- [60] D. R. Heath-Brown. Primes represented by $x^3 + 2y^3$. *Acta Mathematica*, 186:1–84, 2001. [\[Link\]](#). [224](#)
- [61] T. W. Hilberdink and M. L. Lapidus. Beurling Zeta functions, generalised primes, and fractal membranes. *Preprint*, pages 1–31, 2004. arXiv [\[Link\]](#). [131](#), [132](#), [228](#)
- [62] Christian Hill. *Learning Scientific Programming with Python*. Cambridge University Press, 2016. [\[Link\]](#). [19](#)
- [63] Robert V. Hogg, Joseph W. McKean, and Allen T. Craig. *Introduction to Mathematical Statistics*. Pearson, eighth edition, 2016. [\[Link\]](#). [53](#)
- [64] Zhiqiu Hu and Rong-Cai Yang. A new distribution-free approach to constructing the confidence region for multiple parameters. *PLOS One*, pages 1–13, 2013. [\[Link\]](#). [243](#)
- [65] Peter Humphries. The distribution of weighted sums of the Liouville function and Pólya's conjecture. *Preprint*, pages 1–33, 2011. arXiv [\[Link\]](#). [229](#)
- [66] Timothy D. Johnson. Introduction to spatial point processes. *Preprint*, page 2008. NeuroImaging Statistics Oxford (NISOx) group [\[Link\]](#)[\[Mirror\]](#). [199](#)
- [67] Chigozie Kelechi. Towards efficiency in the residual and parametric bootstrap techniques. *American Journal of Theoretical and Applied Statistics*, 5(5), 2016. [\[Link\]](#). [97](#)
- [68] Denis Kojevnikov, Vadim Marmer, and Kyungchul Song. Limit theorems for network dependent random variables. *Journal of Econometrics*, 222(2):419–427, 2021. [\[Link\]](#). [199](#)
- [69] Samuel Kotz, Tomasz Kozubowski, and Krzysztof Podgorski. *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Springer, 2001. [213](#)
- [70] Faraj Lagum. *Stochastic Geometry-Based Tools for Spatial Modeling and Planning of Future Cellular Networks*. PhD thesis, Carleton University, 2018. [\[Link\]](#). [198](#)
- [71] Günther Last and Mathew Penrose. *Lectures on the Poisson Process*. Cambridge University Press, 2017. [198](#)
- [72] Yuk-Kam Lau, Gerald Tenenbaum, and Jie Wu. On mean values of random multiplicative functions. *Proceedings of the American Mathematical Society*, 142(2):409–420, 2013. [\[Link\]](#). [128](#), [130](#)
- [73] Gary R. Lawlor. A l'Hospital's rule for multivariable functions. *Preprint*, pages 1–13, 2013. arXiv:1209.0363 [\[Link\]](#). [113](#)
- [74] Jing Lei et al. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113:1094–1111, 2018. [\[Link\]](#). [97](#)
- [75] G. Last M.A. Klatt and D. Yogeshwaran. Hyperuniform and rigid stable matchings. *Random Structures and Algorithms*, 2:439–473, 2020. [\[Link\]](#)[\[PowerPoint\]](#). [198](#)
- [76] Jorge Mateu, Frederic P Schoenberg, and David M Diez. On distances between point patterns and their applications. *Preprint*, pages 1–29, 2010. [\[Link\]](#). [199](#)
- [77] Natarajan Meghanathan. Distribution of maximal clique size of the vertices for theoretical small-world networks and real-world networks. *Preprint*, pages 1–20, 2015. arXiv:1508.01668 [\[Link\]](#). [202](#)
- [78] Masahiro Mine. Probability density functions attached to random Euler products for automorphic L-functions. *Preprint*, pages 1–38, 2020. arXiv [\[Link\]](#). [228](#), [229](#)
- [79] Christoph Molnar. *Interpretable Machine Learning*. ChristophMolnar.com, 2022. [\[Link\]](#). [97](#), [253](#)

- [80] Marc-Andreas Muendler. Linear difference equations and autoregressive processes. 2000. University of Berkeley [Link]. 49
- [81] V. Kumar Murty. Seminar on Fermat's last theorem. In *Canadian Mathematical Society – Conference Proceedings*, volume 17, Toronto, Canada, 1995. [Link]. 225
- [82] Peter Mörters and Yuval Peres. *Brownian Motion*. Cambridge University Press, 2010. Cambridge Series in Statistical and Probabilistic Mathematics, Volume 30 [Link]. 146, 150
- [83] Jesper Møller. Introduction to spatial point processes and simulation-based inference. In *International Center for Pure and Applied Mathematics (Lecture Notes)*, Lomé, Togo, 2018. [Link][Mirror]. 199, 212, 244
- [84] Jesper Møller and Rasmus P. Waagepetersen. *An Introduction to Simulation-Based Inference for Spatial Point Processes*. Springer, 2003. 199
- [85] Jesper Møller and Rasmus P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. CRC Press, 2007. 199
- [86] S. Ghosh N., Miyoshi, and T. Shirai. Disordered complex networks: energy optimal lattices and persistent homology. *Preprint*, pages 1–44, 2020. arXiv:2009.08811. 192
- [87] Saralees Nadarajah. A modified Bessel distribution of the second kind. *Statistica*, 67(4):405–413, 2007. [Link]. 213
- [88] Hasan Nasab, Mahdi Tavana, and Mohsen Yousefu. A new heuristic algorithm for the planar minimum covering circle problem. *Production and Manufacturing Research*, pages 142–155, 2014. [Link]. 209
- [89] Guillermo Navas-Palencia. Optimal binning: mathematical programming formulation. *Preprint*, pages 1–21, 2020. arXiv:2001.08025 [Link]. 37
- [90] Nathan Ng. Large gaps between the zeros of the Riemann zeta function. *Journal of Number Theory*, 128(3):509–556, 2007. [Link]. 232
- [91] Yosihiko Ogata. Cluster analysis of spatial point patterns: posterior distribution of parents inferred from offspring. *Japanese Journal of Statistics and Data Science*, 3:367–390, 2020. 198
- [92] Fred Park. Shape descriptor / feature extraction techniques. 2011. UCI iCAMP 2011 [Link]. 84
- [93] Yuval Peres and Allan Sly. Rigidity and tolerance for perturbed lattices. *Preprint*, pages 1–20, 2020. arXiv:1409.4490 [Link]. 192, 198
- [94] Carl Rasmussen and Christopher Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006. [Link]. 52
- [95] Alfred R.Osborne. Multidimensional Fourier series. *International Geophysics*, 97:115–145, 2010. [Link]. 120
- [96] Kamron Saniee. A simple expression for multivariate Lagrange interpolation. *SIAM Undergraduate Research Online*, 2007. SIURO [Link]. 103
- [97] Mahesh Shivanand and all. Fitting random regression models with Legendre polynomial and B-spline to model the lactation curve for Indian dairy goat of semi-arid tropic. *Journal of Animal Breeding and Genetics*, pages 414–422, 2022. [Link]. 120
- [98] Karl Sigman. Notes on the Poisson process. New York NY, 2009. IEOR 6711: Columbia University course [Link]. 198
- [99] Joshua Snoke et al. General and specific utility measures for synthetic data. *Journal of the Royal Statistical Society Series A*, 181:663–688, 2018. arXiv:1604.06651 [Link]. 35
- [100] Luuk Spreeuwiers. *Image Filtering with Neural Networks: Applications and Performance Evaluation*. PhD thesis, University of Twente, 1992. 73
- [101] J. Michael Steele. Le Cam's inequality and Poisson approximations. *The American Mathematical Monthly*, 101(1):48–54, 1994. [link]. 161
- [102] Dietrich Stoyan, Wilfrid S. Kendall, Sung Nok Chiu, and Joseph Mecke. *Stochastic Geometry and Its Applications*. Wiley, 2013. 209
- [103] E.C. Titchmarsh and D.R. Heath-Brown. *The Theory of the Riemann Zeta-Function*. Oxford Science Publications, second edition, 1987. 58, 131, 217
- [104] Chris Tofallis. Fitting equations to data with the perfect correlation relationship. *Preprint*, pages 1–11, 2015. Hertfordshire Business School Working Paper[Link]. 13
- [105] D. Umbach and K.N. Jones. A few methods for fitting circles to data. *IEEE Transactions on Instrumentation and Measurement*, 52(6):1881–1885, 2003. [Link]. 14, 17

- [106] D. A. Vaccari and H. K. Wang. Multivariate polynomial regression for identification of chaotic time series. *Mathematical and Computer Modelling of Dynamical Systems*, 13(4):1–19, 2007. [[Link](#)]. 17
- [107] Remco van der Hofstad. *Random Graphs and Complex Networks*. Cambridge University Press, 2016. [[Link](#)]. 201
- [108] Yu Vizilter and Sergey Zheltov. Geometrical correlation and matching of 2D image shapes. 2012. ResearchGate [[Link](#)]. 86
- [109] Fengyun Wang and all. Bivariate Fourier-series-based prediction of surface residual stress fields using stresses of partial points. *Mathematics and Mechanics of Solids*, 2018. [[Link](#)]. 120
- [110] Luyao Wang and Hai Cheng. Pseudo-random number generator based on logistic chaotic system. *Entropy*, 21, 2019. [[Link](#)]. 144
- [111] Mingguang Wu, Yanjie Sun, and Yaqian Li. Adaptive transfer of color from images to maps and visualizations. *Cartography and Geographic Information Science*, pages 289–312, 2021. [[Link](#)]. 168
- [112] Lan Wua, Yongcheng Qi, and Jingping Yang. Asymptotics for dependent Bernoulli random variables. *Statistics and Probability Letters*, pages 455–463, 2012. [[Link](#)]. 146
- [113] Oren Yakir. Recovering the lattice from its random perturbations. *Preprint*, pages 1–18, 2020. arXiv:2002.01508 [[Link](#)]. 198
- [114] Ruqiang Yan, Yongbin Liub, and Robert Gao. Permutation entropy: A nonlinear statistical measure for status characterization of rotary machines. *Mechanical Systems and Signal Processing*, 29:474–484, 2012. 212
- [115] Shaohong Yan, Aimin Yang, et al. Explicit algorithm to the inverse of Vandermonde matrix. In *2009 International Conference on Test and Measurement*, 2009. IEEE [[Link](#)]. 47
- [116] D. Yogeshwaran. Geometry and topology of the boolean model on a stationary point processes : A brief survey. *Preprint*, pages 1–13, 2018. Researchgate [[Link](#)]. 199
- [117] Tonglin Zhang. A Kolmogorov-Smirnov type test for independence between marks and points of marked point processes. *Electronic Journal of Statistics*, 8(2):2557–2584, 2014. 194

Index

- α -compositing, 57
- m -interlacing, 71, 199, 207, 210, 261
- A/B testing, 259
- AdaBoost, 36, 260
- additive number theory, 224, 231
- adversarial learning, 254, 262
- agent-based modeling, 167, 180
- AI art, 176, 254
- algebraic number, 133
- algorithmic bias, 254
- analytic continuation, 218, 225
- analytic function, 131
- anisotropy, 185, 206, 212
- anti-aliasing, 55, 59, 235, 240
- association rule, 248
- attraction (point process), 197
- attraction basin, 58
- attractor distribution, 151, 201, 207, 214
- augmented data, 35, 88, 167, 169, 262
- auto-correlation, 49, 132
- auto-regressive process, 49, 151, 260
- Bailey–Borwein–Plouffe formulas, 133
- Bayesian classification, 74
- Bayesian inference, 44
 - hierarchical models, 261
 - naive Bayes, 248
- Bernoulli trials, 242
- Berry-Esseen inequality, 129
- Bessel function, 213
- Beurling primes, 132, 228
- binning, 247
 - optimum binning, 37, 260
- binomial distribution, 207
- bisection method (root finding), 158
- boosted trees, 255
- bootstrapping, 15, 96, 208, 254
 - percentile method, 101
- boundary effect, 194, 195, 200, 201, 205–207, 211
- Brownian motion, 51, 146, 150, 167, 168, 183, 252, 260
 - Lévy flight, 151
- Brun's theorem, 224
- Cauchy distribution, 151
- Cauchy-Riemann equations, 131
- causality, 261
- Cayley-Hamilton theorem, 47
- CDF regression, 17
- censored data, 200
- central limit theorem, 151, 192, 242
- chaotic dynamical system, 167, 183
- character
 - principal, 220
- characteristic function, 213
- characteristic polynomial, 47, 49–51, 152
- Chebyshev's bias (prime numbers), 131, 220
- checksum, 258
- Chi-squared test, 194
- Chowla conjecture, 222
- classification, 262
- clique (graph theory), 202
- cluster process, 196, 200, 207
- clustering, 262
- Collatz conjecture, 139
- collision graph, 183
- color model
 - RGB, 55, 241
 - RGBA, 55, 56, 76, 241
- color opacity, 177
- color transparency, 20, 169, 241
- complex random variable, 128, 228
- computational complexity, 141, 249
- computer vision, 12, 83
- confidence band, 208
- confidence interval, 44, 205, 260
- confidence level, 243, 245
- confidence region, 15, 29, 205, 206, 242, 254
 - dual region, 44, 243, 260
- conformal map, 14
- confusion matrix, 247, 262
- connected components, 184, 200–202, 206, 207, 250
- contour level, 176, 245
- contour plot, 245
- convergence
 - abscissa, 220, 225
 - absolute, 112, 217, 218
 - alternating series, 218
 - conditional, 112, 130, 220
 - Dirichlet test, 218
- convergence acceleration, 59
- convex linear combination, 108
- convolution of distributions, 213
- copula, 99, 238, 254
- counting measure, 193
- covariance matrix, 90, 242
- covering (stochastic), 209
- covering problem, 208
- credible interval, 44
- credible region (Bayesian), 243, 260
- critical line (number theory), 113

cross-validation, 15, 183, 247
 cuban primes, 224
 curse of dimensionality, 115
 curve fitting, 26
 data video, 167, 176
 decision tree, 36
 Dedekind zeta function, 131, 228
 deep neural network, 73, 261
 dense set (topology), 222
 density estimation, 199
 diamond-square algorithm, 168
 Diehard tests of randomness, 130
 dimensionality reduction, 16
 Dirichlet character, 131, 132, 219, 224
 modulo 4, 220, 225, 226
 Dirichlet eta function, 232
 Dirichlet functional equation, 131, 225, 226
 Dirichlet series, 128
 Dirichlet's theorem, 131, 220, 222, 226
 Dirichlet-*L* function, 131, 219, 226
 disaggregation, 114
 discrete Fourier series, 119
 discrete orthogonal functions, 119
 dissimilarity metric, 249
 distributed architecture, 246
 distribution
 Cauchy, 151
 Fréchet, 51, 151
 Gaussian, 242
 generalized logistic, 90, 196
 Hotelling, 243
 Laplace, 213
 logistic, 16
 Lévy, 151
 modified Bessel, 213
 Poisson-binomial, 161, 215
 Poisson-exponential, 192
 Rademacher, 128, 129
 Rayleigh, 207, 208, 214
 Weibull, 51, 151, 207
 domain of attraction, 201
 dot product, 14
 dummy variable, 36
 dyadic map, 132
 dynamical systems, 132, 201
 chaotic systems, 167, 183
 dyadic map, 132
 ergodicity, 132
 logistic map, 132
 shift map, 132
 stochastic, 168
 edge effect (statistics), 200
 eigenvalue, 13, 52, 90, 262
 power iteration, 92
 elbow rule, 155, 200, 207, 252
 elliptic curve, 224
 EM algorithm, 35, 261
 empirical distribution, 16, 96, 120, 192, 195, 201, 207,
 212, 214, 226, 254
 multivariate, 129
 empirical quantiles, 101
 ensemble methods, 36, 83, 255
 entropy, 186, 212, 248
 equidistribution modulo 1, 135
 equilibrium distribution, 168
 Erdős-Rényi model, 202
 ergodicity, 132, 168, 205, 207, 213
 Euler product, 128, 218, 225
 random, 228
 Euler's transform, 232
 evolutionary process, 168
 experimental design, 259
 experimental math, 56, 216
 explainable AI, 13, 35, 74, 83, 90, 155, 171, 209, 253
 exploratory analysis, 258
 exponential decay, 40
 exponential sums, 225
 extrapolation, 108
 extreme value theory, 151, 214
 feature attribution, 253
 feature importance, 253
 feature selection, 15, 97, 246
 Fermat's last theorem, 225
 fixed-point algorithm, 59, 89, 155, 261
 flag vector, 248, 259
 Fourier series, 119
 Fourier transform, 213
 fractal dimension, 51
 fractional part function, 134
 Frobenius norm, 90
 Fruchterman and Rheingold algorithm, 250
 Fréchet distribution, 51, 151
 fuzzy classification, 56
 Gamma function, 51, 151
 GAN (generative adversarial networks), 35, 261
 Gaussian circle problem, 231
 Gaussian distribution, 242
 Gaussian mixture, 35, 70
 Gaussian primes, 131, 227
 Gaussian process, 49, 260
 general linear model, 13
 generalized linear model, 13, 48
 generalized logistic distribution, 90, 206
 generative adversarial networks, 35, 171, 254, 261
 generative AI, 167, 180
 generative model, 35, 52, 99, 166, 167, 169, 176, 183,
 262
 geostatistics, 102
 GIS, 116
 Glivenko-Cantelli theorem, 226
 GMM (Gaussian mixture model), 35, 69, 261
 Goldbach's conjecture, 224
 goodness-of-fit, 56, 247
 GPU-based clustering, 71
 gradient (optimization), 155
 gradient boosting, 260
 gradient operator, 15
 Gram-Schmidt orthogonalization, 119

graph, 200
 collision graph, 183
 connected components, 184, 206, 250
 directed, 184
 edge, 200
 Fruchterman-Reingold, 184
 nearest neighbor graph, 202, 206
 node, 200, 202
 random graph, 201
 random nearest neighbor graph, 201
 tree, 184
 undirected, 200–202, 207
 vertex, 200
graph database, 250
graph theory, 200
GraphViz, 184
greedy algorithm, 112, 231
grid search, 155, 170
half-tone (music), 238
Hartman–Wintner theorem, 146
hash table, 142, 186, 211, 248, 249
 sparse, 249
Hausdorff distance, 87
Hellinger distance, 255
Hermite polynomials, 119
hexagonal lattice, 200
hidden decision trees, 36, 37, 260
hidden layer, 73
hidden process, 192, 210, 214
hierarchical clustering, 73, 249
Hilbert primes, 227
histogram equalization, 71, 73
Hoeffding inequality, 149
homogeneity (point process), 161, 199
Hotelling distribution, 243
Hurst exponent, 51
hyperparameter, 29, 56, 103, 170
identifiability, 210, 212
ill-conditioned problem, 26, 52, 92, 261
image segmentation, 73
imputation (missing values), 254
index
 index discrepancy, 212
intensity (stochastic process), 192, 199, 206
interarrival times, 150, 192, 201, 205, 212
 standardized, 213
interlaced processes, 199
Internet of Things, 192
inverse distance weighting, 104
inverse square law, 180
iterated logarithm, 129, 130, 146
Itô integral, 52
K-means clustering, 31, 32
key-value pair, 37, 248
Kolmogorov-Smirnov test, 129, 194, 201
kriging, 112
Kronecker's theorem, 222, 230
Lagrange interpolation, 52
Lagrange multiplier, 15, 261
Laplace distribution, 213
Lasso regression, 15, 262
lattice, 198
 perturbed lattice, 192
 shifted, 200
 stretched, 200
law of the iterated logarithm, 129, 130, 146, 222, 228
Le Cam's theorem, 161, 193
least absolute residuals, 101
link function, 13, 16
Liouville function, 219, 230
log-polar map, 14
logistic distribution, 16, 199
logistic map, 132
logistic regression, 16
 unsupervised, 33
logit function, 261
Lévy distribution, 151
Lévy flight, 151
Map-reduce, 246
marketing attribution, 259
Markov chain, 49
 MCMC, 128
Mathematica, 245
MaxCliqueDyn algorithm, 202
maximum likelihood estimation, 244, 261, 262
mean squared error, 15, 30
medoid, 31
Mersenne twister, 29, 132, 135, 148
Mertens function, 219
minimum contrast estimation, 170, 209, 212, 244
mixture model, 29, 45, 168, 176, 192, 199, 200, 207,
 245, 262
 blending, 168
model fitting, 56, 261
model identifiability, 15
modulus (complex number), 152, 218
Monte Carlo simulations, 128, 261
morphing (computer vision), 167
moving average, 157
multidimensional Fourier series, 120
multiple root, 113
multiplicative function
 completely multiplicative, 128, 130, 219, 220, 231
 Rademacher, 128
Möbius function, 219
N-body problem, 180
n-gram (NLP), 249
naive Bayes, 248, 260
natural language processing, 36, 249
nearest neighbor interpolation, 101, 104
nearest neighbors, 192, 202, 208, 261
 nearest neighbor distances, 206–208, 210, 214
 nearest neighbor graph, 206
NetworkX, 184
neural network, 73
 hidden layer, 73
 hyperparameter, 75

neuron, 73, 261
 seq2seq, 166
 sparse, 69
 very deep, 73
 Newton's method, 155
 node (decision tree), 37, 255, 260
 perfect node, 44
 usable node, 38
 node (interpolation), 113
 normal number, 129, 222, 226
 strongly normal, 130
 numerical stability, 47

 Omega function, 219, 225
 order statistics, 214
 ordinary least squares, 50, 101, 119
 orthogonal function, 119
 outliers, 214, 252
 overfitting, 15, 212, 254, 255, 260

 palette, 167, 241
 parametric bootstrap, 20, 29, 35, 97, 208, 254, 260, 261
 partial derivative, 113
 partial least squares, 13
 path (graph theory), 200
 percentile bootstrap, 101
 permutation
 entropy, 212
 random permutation, 211
 perturbed lattices, 192
 Plotly, 176
 point count distribution, 193, 206, 209
 point process
 attractive, 207
 cluster process
 Matérn, 198
 Neyman-Scott, 198
 non-homogeneous, 161, 199
 perturbed lattice process, 198
 radial, 199
 renewal process, 198
 repulsive, 197
 Poisson point process, 150, 161, 192, 206
 Poisson-binomial distribution, 161, 192, 215
 Poisson-exponential distribution, 192
 positive semidefinite (matrix), 48, 91
 power iteration, 92
 preconditioning, 92
 prediction interval, 15, 96, 101
 predictive power, 37, 44, 247, 248, 253
 prime test (of randomness), 130, 141, 147
 principal component analysis, 48, 253, 260
 probability generating function, 147
 proxy space, 245
 pseudo-inverse matrix, 48
 pseudo-random numbers, 148, 252
 combined generators, 144
 congruential generator, 135
 Diehard tests, 130, 142
 Mersenne twister, 135, 148, 169
 prime test, 130, 147
 strongly random, 130, 133
 TestU01, 130
 Pólya conjecture, 221

 quadratic irrational, 132, 135, 141
 quantile, 243, 261
 empirical, 101, 254
 weighted, 101
 quantile function, 99, 120, 192, 196, 207
 quantile regression, 15

 R-squared, 15, 35, 170
 Rademacher distribution, 129
 Rademacher function, 128, 222, 228
 random, 130
 random function, 160
 random graph, 201, 202
 random multiplicative function, 128
 Rademacher, 130
 random permutation, 211
 random variable
 complex, 128
 random walk, 146, 170, 260
 first hitting time, 147, 150
 zero crossing, 146
 Rayleigh distribution, 207, 208, 214
 Rayleigh test, 207
 records, 214
 regression splines, 13
 regular expression, 249, 258
 reinforcement learning, 262
 rejection sampling, 255
 renewal process, 198
 repulsion (point process), 197, 208
 repulsion basin, 218
 resampling, 96, 208
 Riemann Hypothesis, 107, 113
 Generalized, 130, 220, 224, 226
 Riemann zeta function, 113, 128, 131, 226, 228
 root mean squared error, 56

 scaling factor, 206, 214
 seed (random number generator), 142, 169, 255
 semi-supervised learning, 262
 shape signature, 84
 Shapley value, 253
 Shepard's method, 104
 shift map, 132
 sigmoid function, 261
 simplex, 229
 singular value decomposition, 13, 262
 singularity, 186
 six degrees of separation, 251
 Sklar's theorem, 254
 smoothing parameter, 103
 spatial statistics, 102, 198
 spectral domain, 168
 spline regression, 120
 square root (matrix), 48, 91
 square-free integer, 129, 142, 222
 stable distribution, 151, 169, 213

state space, 168
stationary distribution, 52
stationary process, 49, 151, 168, 183, 194, 199, 206
stepwise regression, 98
stochastic convergence, 168
stochastic function, 51
stochastic geometry, 209
stochastic process, 192
stochastic residues, 210
stop word (NLP), 249
stretching (point process), 200
Sturm-Liouville theory, 119
superimposition (point processes), 199
supervised classification, 71
surface plot, 176
swarm optimization, 27, 261
synthetic data, 13, 27, 29, 52, 88, 90, 112, 118, 141, 147, 155, 169, 176, 183, 216, 243, 254
synthetic metric, 248

Tarjan's algorithm, 250
tensor, 74
text normalization, 249
Theil-Sen estimator, 101
time series, 50

- auto-regressive, 51, 151
- disaggregation, 107
- Hurst exponent, 51
- non-periodic, 25

total least squares, 13
training set, 101, 183, 247
transcendental number, 133
transformer, 73, 166
tree (graph theory), 184
twin primes, 224

universality property, 218, 222, 224
unsupervised clustering, 71
unsupervised learning, 33, 262

validation set, 15, 56, 101, 183, 247, 254, 255
Vandermonde matrix, 47, 52
vertex, 192, 200, 201, 214
video compression

- FFmpeg, 55, 59

Waring's problem, 224
Watts and Strogatz model, 251
Weibull distribution, 51, 151, 207, 214
weighted least squares, 13
weighted quantiles, 101
weighted regression, 16
white noise, 27, 49, 151, 260
wide data, 120, 262

XOR operator, 135