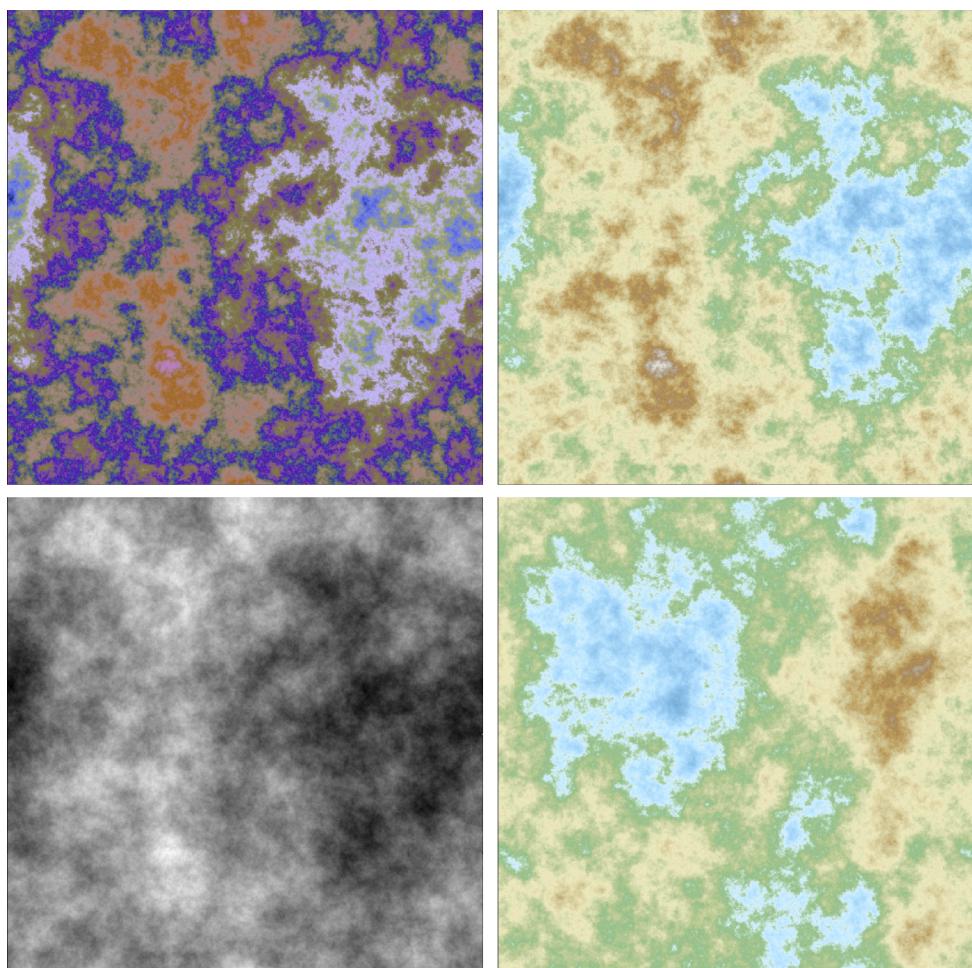

Synthetic Data



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. Chapters 12 and 13 focus on synthetic data applications: fractal-like terrain generation with the diamond-square algorithm, 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 14 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.

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 Python tutorial in itself, and an introduction to scientific computing. Some data animations and chart enhancements are done in R. 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 eager to learn a lot of useful material in a limited amount of time. 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	18
1.3.2.3 Python code	19
1.3.3 Non-periodic sum of periodic time series	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
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	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	49
3.4 Mathematics of auto-regressive time series	50

3.4.1	Simulations: curious fractal time series	50
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 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	Detecting Subtle Departures from Randomness	112
9.1	Introduction	112
9.2	Pseudo-random numbers	113
9.2.1	Strong pseudo-random numbers	113
9.2.1.1	New test of randomness for PRNGs	114
9.2.1.2	Theoretical background: the law of the iterated logarithm	114
9.2.1.3	Connection to the Generalized Riemann Hypothesis	114
9.2.2	Testing well-known sequences	115
9.2.2.1	Reverse-engineering a pseudo-random sequence	116
9.2.2.2	Illustrations	117
9.3	Python code	119
9.3.1	Fixes to the faulty random function in Python	119
9.3.2	Prime test implementation to detect subtle flaws in PRNG's	119
9.3.3	Special formula to compute 10 million digits of $\sqrt{2}$	122
10	Some Unusual Random Walks	126
10.1	Symmetric unbiased constrained random walks	126
10.1.1	Three fundamental properties of pure random walks	126
10.1.2	Random walks with more entropy than pure random signal	127
10.1.2.1	Applications	127
10.1.2.2	Algorithm to generate quasi-random sequences	128
10.1.2.3	Variance of the modified random walk	128
10.1.3	Random walks with less entropy than pure random signal	129
10.2	Related stochastic processes	130
10.2.1	From Brownian motions to clustered Lévy flights	130
10.2.2	Integrated Brownian motions and special auto-regressive processes	131
10.3	Python code	132
10.3.1	Computing probabilities and variances attached to S_n	132
10.3.2	Path simulations	133
11	Empirical Optimization with Divergent Fixed Point Algorithm	135
11.1	Introduction	135
11.1.1	The problem, with illustration	136
11.2	Non-converging fixed-point algorithm	137
11.2.1	Trick leading to intuitive solution	137
11.2.2	Root detection: method and parameters	137
11.2.3	Case study with conclusions	138
11.3	Generalization with synthetic data	138
11.3.1	Example	140
11.3.2	Connection to the Poisson-binomial distribution	141

11.3.2.1	Integer sequences with high density of primes	141
11.3.3	Python code: finding the optimum	141
11.4	Smoothing highly chaotic curves	143
11.4.1	Python code: code smoother	143
12	Synthetic Terrain Generation and AI-generated Art	146
12.1	Introduction	146
12.2	Terrain generation and the evolutionary process	148
12.2.1	Morphing and non-linear palette operations	148
12.2.2	The diamond-square algorithm	148
12.2.3	The evolutionary process	149
12.2.4	Finding optimum parameters	149
12.2.5	Mimicking real terrain: the synthesis step	149
12.3	Python code	150
12.3.1	Producing data videos with four sub-videos in parallel	150
12.3.2	Main program	151
12.4	AI-generated art with 3D contours	155
12.4.1	Python code using Matplotlib	155
12.4.2	Python code using Plotly	157
12.4.3	Tips to quickly solve new problems	158
13	Synthetic Star Clusters Generation with Collision Graphs	159
13.1	Introduction	159
13.2	Model parameters and simulation results	160
13.2.1	Explanation of color codes	160
13.2.2	Detailed description of top parameters	160
13.2.3	Interesting parameter sets	161
13.3	Analysis of star collisions and collision graph	162
13.3.1	Weighted directed graphs: visualization with NetworkX	163
13.3.2	Interesting findings: how the universe got started	163
13.4	Animated data visualizations	164
13.5	Python code and computational issues	165
13.5.1	Simulating the real and synthetic universes	165
13.5.2	Visualizing collision graphs	169
14	Perturbed-Lattice Point Process: Inference, Nearest Neighbor Graph	171
14.1	Perturbed lattices: definition and properties	171
14.1.1	Point counts distribution	172
14.1.2	Periodicity and amplitude of point count expectations	172
14.1.3	Testing the independence of point counts	173
14.1.3.1	Results and Interpretation	174
14.1.3.2	About the Spreadsheet	175
14.2	Cluster processes and nearest neighbor graphs	175
14.2.1	Synthetic, semi-rigid cluster structures	175
14.2.2	Python code to generate cluster processes	177
14.2.3	References on cluster processes	177
14.2.4	Superimposed perturbed lattices: an alternative to mixture models	178
14.2.4.1	Hexagonal lattice, nearest neighbors	179
14.2.4.2	Exercises: nearest neighbor graphs, size of connected components	180
14.2.4.3	Python code to compute connected components	181
14.3	Statistical inference for point processes	183
14.3.1	Estimation of Core Parameters	183
14.3.1.1	Intensity	184
14.3.1.2	Scaling factor	184
14.3.1.3	Alternative estimation method	184
14.3.2	Spatial statistics, nearest neighbors, clustering	185
14.3.2.1	Inference for two-dimensional processes	185
14.3.2.2	Other possible tests	185
14.3.2.3	Rayleigh test	186
14.3.2.4	Exercises	187
14.4	Special topics	188
14.4.1	Minimum contrast estimation and explainable AI	188

14.4.2	Model identifiability, hard-to-detect patterns	189
14.4.2.1	Stochastic residues	189
14.4.3	Hidden model and random permutations	189
14.4.4	Retrieving the F distribution	191
14.4.4.1	Theoretical values obtained by simulations	191
14.4.4.2	Retrieving F from the interarrival times distribution	192
14.4.5	Record distances between an observed point and its vertex	192
14.4.5.1	Distribution of records	193
14.4.5.2	Distribution of arrival times for records	194
15	New Perspective on the Riemann Hypothesis	195
15.1	Introduction	195
15.1.1	Key concepts and terminology	196
15.1.2	Orbits and holes	196
15.1.3	Industrial Applications	196
15.2	Euler products	197
15.2.1	Finite Euler Products	197
15.2.1.1	Generalization using Dirichlet characters	198
15.2.2	Infinite Euler products	199
15.2.2.1	Special products	199
15.2.2.2	Probabilistic properties and conjectures	200
15.3	Finite Dirichlet series and generalizations	201
15.3.1	Finite Dirichlet series	201
15.3.2	Non-trivial cases with infinitely many primes and a hole	203
15.3.2.1	Sums of two cubes, or cuban primes	203
15.3.2.2	Primes associated to elliptic curves	203
15.3.2.3	Analytic continuation, convergence, and functional equation	204
15.3.2.4	Hybrid Dirichlet-Taylor series	204
15.3.3	Riemann Hypothesis with cosines replaced by wavelets	205
15.3.4	Riemann Hypothesis for Beurling primes	206
15.3.5	Stochastic Euler products	207
15.4	Exercises	208
15.5	Python code	211
15.5.1	Computing the orbit of various Dirichlet series	211
15.5.2	Creating videos of the orbit	214
16	Miscellaneous Topics	217
16.1	The sound that data makes	217
16.1.1	From data visualizations to videos to data music	217
16.1.2	References	218
16.1.3	Python code	218
16.2	Data videos and enhanced visualizations in R	219
16.2.1	Cairo library to produce better charts	219
16.2.2	AV library to produce videos	220
16.3	Dual confidence regions	221
16.3.1	Case study	221
16.3.2	Standard confidence region	221
16.3.3	Dual confidence region	222
16.3.4	Simulations	222
16.3.5	Original problem with minimum contrast estimators	223
16.3.6	General shape of confidence regions	224
16.4	Fast feature selection based on predictive power	225
16.4.1	How cross-validation works	226
16.4.2	Measuring the predictive power of a feature	226
16.4.3	Efficient implementation	227
16.5	Natural language processing: taxonomy creation	228
16.5.1	Designing a keyword taxonomy	228
16.5.2	Fast clustering algorithm for keyword data	229
16.5.2.1	Computational complexity	229
16.5.2.2	Smart crawling of the whole Internet and a bit of graph theory	229
16.6	Automated detection of outliers and number of clusters	231
16.6.1	Black-box elbow rule to detect outliers	231

16.7 New developments in explainable AI	232
16.8 Advice to beginners	233
16.8.1 Getting started and learning how to learn	233
16.8.1.1 Getting help	233
16.8.1.2 Beyond Python	234
16.8.2 Automated data cleaning and exploratory analysis	234
16.8.3 Example of simple analysis: marketing attribution	235
16.8.4 Upcoming books and courses	235
Glossary	236
Bibliography	239
Index	244

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 15.2.2.1)	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	Orbit of $L(z, \chi)$ at $\sigma = \frac{1}{2}$, with $0 < t < 200$ and $\chi = \chi_4$ (left) versus pseudo-random χ (right) . .	115
9.2	$L_3^*(n)$ test statistic for four sequences: Python[200] and SQRT[90,91] fail	117
9.3	$ L_3(n) $ test statistic for four sequences: Python[200] and SQRT[90,91] fail	117
10.1	Typical path S_n with $0 \leq n \leq 50,000$ for four types of random walks	127
10.2	$\delta_n = 1 - \text{Var}[S_{n+1}] + \text{Var}[S_n]$ for four types of random walks, with $0 \leq n \leq 5000$	128

10.3	Same as Figure 10.2, using a more aesthetic but less meaningful chart type	129
10.4	Clustered Brownian process	131
10.5	AR models, classified based on the types of roots of the characteristic polynomial	132
11.1	Function $f(b)$ as a better alternative to $g(b)$ in Figure 11.2. Root at $b = 3083$.	136
11.2	Function $g(b) = 2 - \cos(2\pi b) - \cos(2\pi a/b)$, with $a = 3083 \times 7919$.	136
11.3	Transformed function f_3 , amplifying the root at $b = 3083$.	137
11.4	Signal strength ρ_n , first 130 fixed-point iterations; $n = 31$ leads to a root.	140
11.5	(b_n, ρ_n) plot. Yellow and orange dots linked to roots.	140
11.6	Signal strength ρ_n , first 130 fixed-point iterations; $n = 87$ leads to a root.	140
11.7	Random function from section 11.3.1, with root at $b = 5646$.	143
12.1	Six frames from the terrain video, each one containing four images	147
12.2	Contour plot, 3D mixture model, produced with Plotly	155
12.3	Same as Figure 12.2, produced with Matplotlib	156
13.1	Collisions graph for the biggest star eater (star 47) in video 7	163
13.2	Summary statistics for the whole collision structure: the X axis represents the time	164
13.3	Snapshots of universe 4 (left) and universe 7 (right)	165
14.1	Period and amplitude of $\phi_\tau(t)$; here $\tau = 1, \lambda = 1.4, s = 0.3$	173
14.2	A new test of independence (R-squared version)	173
14.3	Radial cluster process ($s = 0.2, \lambda = 1$) with centers in blue; zoom in on the left	176
14.4	Radial cluster process ($s = 2, \lambda = 1$) with centers in blue; zoom in on the left	176
14.5	Manufactured marble lacking true lattice randomness (left)	176
14.6	Four superimposed Poisson-binomial processes: $s = 0$ (left), $s = 5$ (right)	179
14.7	Rayleigh test to assess if a point distribution matches that of a Poisson process	187
14.8	Realization of a 5-interlacing with $s = 0.15$ and $\lambda = 1$: original (left), modulo $2/\lambda$ (right)	190
14.9	Locally random permutation σ ; $\tau(k)$ is the index of X_k 's closest neighbor to the right	190
14.10	Each arrow links a point (blue) to its vertex (red): $s = 0.2$ (left), $s = 1$ (right)	193
14.11	Distance between a point and its vertex ($\lambda = s = 1$)	194
15.1	Three orbits ($\sigma = 0.5, 0.75, 1.25$) with finite Euler product: $P = \{2, 3\}$ (left) vs $\{2, 3, 5\}$ (right)	198
15.2	Distance between orbit and location $(c, 0)$ depending on t on the X-axis	200
15.3	Distance between orbit and location $(c, 0)$ depending on t on the X-axis	200
15.4	Distance between orbit and location $(c, 0)$ depending on t on the X-axis	200
15.5	Four orbits where the “hole” (repulsion basin) is apparent	202
15.6	Three orbits with “hole” closer to the origin, showing impact of $\beta > \frac{1}{2}$ and larger n	202
15.7	Orbit of Dirichlet eta $\eta(z)$ when cosines are replaced by other periodic functions	206
16.1	Data linked to the melody: red curve for note frequencies, blue curve for note durations	218
16.2	R plot before Cairo (left), and after (right)	219
16.3	Intermediate (left) and last frame (right) of the video	220
16.4	Example of 90% dual confidence region for (p, q)	222
16.5	Minimum contrast estimation for (λ, s) using (p, q) as proxy stats	223
16.6	Non-elliptic confidence regions with various confidence levels	224
16.7	Elbow rule (right) finds $m = 3$ clusters in Brownian motion (left)	232

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
9.1	$L_3^*(n)$, for various sequences ($n = 20,000$); “Fail” means failing the prime test	118
11.1	High ρ_n at iterations $n = 31$ and $n = 127$ points to roots 3083 and 7919	139
13.1	Description of top parameters used in the star cluster simulator	161
13.2	Eight selected parameter sets covering various situations	162
14.1	Variance attached to F_s , as a function of s	172
14.2	Poisson process ($s = \infty$) versus $s = 39.85$	192
16.1	Extract of the mapping table used to recover (λ, s) from (p, q)	224
16.2	Eight bins: 2 features (A, B) times 2 outcomes (Good/Bad)	226
16.3	Amount of data collected at each level, when crawling the Internet	230

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, 131
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, 226
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, 236
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, 187, 236
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, 184, 185, 221, 224
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, 162, 226, 236
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, 236, 237
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, 237, 238
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, 113, 171, 174, 180, 186, 191, 193, 205
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, 236
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, 69, 74, 83, 90, 135, 150, 188, 232

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, 217, 225, 236, 238
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, 226, 236, 238
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, 163, 229, 230
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, 149, 237
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, 237
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, 236, 237
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, 236, 237
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 16.5 deals with creating a keyword taxonomy based on parsing Google search results pages. See pages 36, 228
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, 236, 238

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, 225, 227, 232, 237
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, 56, 90, 93, 95, 97, 104, 237, 238
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, 112, 119, 231
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, 49, 52, 56, 89, 95, 101, 108, 237, 238
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 238
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, 48, 55, 69, 70, 75, 88, 94, 105, 127, 135, 148, 155, 162, 222, 231, 236
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, 162, 226, 236, 237, 238
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, 162, 226, 236, 237, 238

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. [\[Link\]](#). 180
- [4] José M. Amigó, Roberto Dale, and Piergiulio Tempesta. A generalized permutation entropy for random processes. *Preprint*, pages 1–9, 2012. arXiv:2003.13728 [\[Link\]](#). 191
- [5] Luc Anselin. *Point Pattern Analysis: Nearest Neighbor Statistics*. The Center for Spatial Data Science, University of Chicago, 2016. Slide presentation [\[Link\]](#). 178
- [6] Hui Guo at al. Eyes tell all: Irregular pupil shapes reveal gan-generated faces. *Preprint*, pages 1–7, 2021. arXiv:2109.00162 [\[Link\]](#). 233
- [7] Adrian Baddeley. Spatial point processes and their applications. In Weil W., editor, *Stochastic Geometry. Lecture Notes in Mathematics*, pages 1–75. Springer, Berlin, 2007. [\[Link\]](#). 177
- [8] N. Balakrishnan and C.R. Rao (Editors). *Order Statistics: Theory and Methods*. North-Holland, 1998. 180, 194
- [9] Christopher Beckham and Christopher Pal. A step towards procedural terrain generation with GANs. *Preprint*, pages 1–5, 2017. arXiv:1707.03383 [\[Link\]](#). 147
- [10] Rabi Bhattacharya and Edward Waymire. *Random Walk, Brownian Motion, and Martingales*. Springer, 2021. 126
- [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. [\[Link\]](#). 181
- [13] Miklos Bona. *Combinatorics of Permutations*. Routledge, second edition, 2012. 191
- [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. [\[Link\]](#). 199
- [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. [\[Link\]](#). 201
- [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. [\[Link\]](#). 211
- [18] Bartłomiej Błaszczyzyn and Dhandapani Yogeshwaran. Clustering and percolation of point processes. *Preprint*, pages 1–20, 2013. Project Euclid [\[Link\]](#). 177
- [19] Bartłomiej Błaszczyzyn and Dhandapani Yogeshwaran. On comparison of clustering properties of point processes. *Preprint*, pages 1–26, 2013. arXiv:1111.6017 [\[Link\]](#). 177
- [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 [\[Link\]](#). 177
- [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. 232
- [22] Keith Conrad. *L-functions and the Riemann Hypothesis*. 2018. 2018 CTNT Summer School [\[Link\]](#). 115, 196, 199, 204
- [23] Noel Cressie. *Statistic for Spatial Data*. Wiley, revised edition, 2015. 177

- [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. [130](#)
- [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. [130](#)
- [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. [223](#)
- [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 [\[Link\]](#). [116](#), [207](#)
- [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. [178](#)
- [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. [178](#)
- [31] David Coupier (Editor). *Stochastic Geometry: Modern Research Frontiers*. Wiley, 2019. [188](#)
- [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. [179](#)
- [33] 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. [\[Link\]](#). [181](#)
- [34] Achim Zeileis et al. Colorsphere: A toolbox for manipulating and assessing colors and palettes. *Preprint*, pages 1–45, 2019. arXiv:1903.06490 [\[Link\]](#) [\[R Library\]](#). [147](#)
- [35] Arash Farahmand. *Math 55 Lecture Notes*. 2021. University of Berkeley [\[Link\]](#). [47](#), [53](#)
- [36] W. Feller. On the Kolmogorov-Smirnov limit theorems for empirical distributions. *Annals of Mathematical Statistics*, 19(2):177–189, 1948. [\[Link\]](#). [180](#), [187](#)
- [37] Nikos Frantzikinakis. Ergodicity of the Liouville system implies the Chowla conjecture. *Preprint*, pages 1–41, 2016. arXiv [\[Link\]](#). [201](#)
- [38] P. M. Gauthier. Approximating the Riemann zeta-function by polynomials with restricted zeros. *Canadian Mathematical Bulletin*, 62(3):475–478, 2018. [\[Link\]](#). [211](#)
- [39] P. A. Van Der Geest. The binomial distribution with dependent Bernoulli trials. *Journal of Statistical Computation and Simulation*, pages 141–154, 2004. [\[Link\]](#). [127](#)
- [40] Stamatia Giannarou and Tania Stathaki. Shape signature matching for object identification invariant to image transformations and occlusion. 2007. ResearchGate [\[Link\]](#). [84](#)
- [41] 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 [\[Link\]](#). [181](#)
- [42] B.V. Gnedenko and A. N. Kolmogorov. *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, 1954. [131](#)
- [43] Manuel González-Navarrete and Rodrigo Lambert. Non-markovian random walks with memory lapses. *Preprint*, pages 1–14, 2018. arXiv [\[Link\]](#). [126](#)
- [44] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. [\[Link\]](#). [53](#)
- [45] Vincent Granville. Estimation of the intensity of a Poisson point process by means of nearest neighbor distances. *Statistica Neerlandica*, 52(2):112–124, 1998. [\[Link\]](#). [178](#)
- [46] Vincent Granville. *Applied Stochastic Processes, Chaos Modeling, and Probabilistic Properties of Numeration Systems*. MLTechniques.com, 2018. [\[Link\]](#). [116](#)
- [47] Vincent Granville. *Stochastic Processes and Simulations: A Machine Learning Perspective*. MLTechniques.com, 2022. [\[Link\]](#). [51](#), [59](#), [131](#), [141](#), [171](#), [172](#), [173](#), [175](#), [179](#), [181](#), [207](#), [211](#), [224](#), [235](#)
- [48] 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](#)
- [49] Vincent Granville and Richard L Smith. Disaggregation of rainfall time series via Gibbs sampling. *NISS Technical Report*, pages 1–21, 1996. [\[Link\]](#). [107](#)
- [50] Kristen Grauman. Shape matching. 2008. University of Texas, Austin [\[Link\]](#). [87](#)
- [51] Radim Halir and Jan Flusser. Numerically stable direct least squares fitting of ellipses. *Preprint*, pages 1–8, 1998. [\[Link\]](#). [17](#), [19](#)
- [52] Peter Hall. *Introduction to the theory of coverage processes*. Wiley, 1988. [188](#)

- [53] Adam J. Harper. Moments of random multiplicative functions, II: High moments. *Algebra and Number Theory*, 13(10):2277–2321, 2019. [[Link](#)]. 112, 207
- [54] 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](#)]. 112, 114, 207
- [55] Adam J. Harper. Almost sure large fluctuations of random multiplicative functions. *Preprint*, pages 1–38, 2021. arXiv [[Link](#)]. 114, 201, 207
- [56] K. Hartmann, J. Krois, and B. Waske. *Statistics and Geospatial Data Analysis*. Freie Universität Berlin, 2018. E-Learning Project SOGA [[Link](#)]. 174
- [57] D. R. Heath-Brown. Primes represented by $x^3 + 2y^3$. *Acta Mathematica*, 186:1–84, 2001. [[Link](#)]. 203
- [58] T. W. Hilberdink and M. L. Lapidus. Beurling Zeta functions, generalised primes, and fractal membranes. *Preprint*, pages 1–31, 2004. arXiv [[Link](#)]. 115, 116, 207
- [59] Christian Hill. *Learning Scientific Programming with Python*. Cambridge University Press, 2016. [[Link](#)]. 19
- [60] Robert V. Hogg, Joseph W. McKean, and Allen T. Craig. *Introduction to Mathematical Statistics*. Pearson, eighth edition, 2016. [[Link](#)]. 53
- [61] 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](#)]. 222
- [62] Peter Humphries. The distribution of weighted sums of the Liouville function and Pólya’s conjecture. *Preprint*, pages 1–33, 2011. arXiv [[Link](#)]. 208
- [63] Timothy D. Johnson. Introduction to spatial point processes. *Preprint*, page 2008. NeuroImaging Statistics Oxford (NISOx) group [[Link](#)][[Mirror](#)]. 178
- [64] Chigozie Kelechi. Towards efficiency in the residual and parametric bootstrap techniques. *American Journal of Theoretical and Applied Statistics*, 5(5), 2016. [[Link](#)]. 97
- [65] Denis Kojevnikov, Vadim Marmer, and Kyungchul Song. Limit theorems for network dependent random variables. *Journal of Econometrics*, 222(2):419–427, 2021. [[Link](#)]. 178
- [66] Samuel Kotz, Tomasz Kozubowski, and Krzysztof Podgorski. *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Springer, 2001. 192
- [67] Faraj Lagum. *Stochastic Geometry-Based Tools for Spatial Modeling and Planning of Future Cellular Networks*. PhD thesis, Carleton University, 2018. [[Link](#)]. 177
- [68] Günther Last and Mathew Penrose. *Lectures on the Poisson Process*. Cambridge University Press, 2017. 177
- [69] 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](#)]. 112, 114
- [70] Jing Lei et al. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113:1094–1111, 2018. [[Link](#)]. 97
- [71] G. Last M.A. Klatt and D. Yogeshwaran. Hyperuniform and rigid stable matchings. *Random Structures and Algorithms*, 2:439–473, 2020. [[Link](#)][[PowerPoint](#)]. 177
- [72] Jorge Mateu, Frederic P Schoenberg, and David M Diez. On distances between point patterns and their applications. *Preprint*, pages 1–29, 2010. [[Link](#)]. 178
- [73] 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](#)]. 181
- [74] Masahiro Mine. Probability density functions attached to random Euler products for automorphic L-functions. *Preprint*, pages 1–38, 2020. arXiv [[Link](#)]. 207, 208
- [75] Christoph Molnar. *Interpretable Machine Learning*. ChristophMolnar.com, 2022. [[Link](#)]. 97, 232
- [76] Marc-Andreas Muendler. Linear difference equations and autoregressive processes. 2000. University of Berkeley [[Link](#)]. 49
- [77] V. Kumar Murty. Seminar on Fermat’s last theorem. In *Canadian Mathematical Society – Conference Proceedings*, volume 17, Toronto, Canada, 1995. [[Link](#)]. 204
- [78] Peter Mörters and Yuval Peres. *Brownian Motion*. Cambridge University Press, 2010. Cambridge Series in Statistical and Probabilistic Mathematics, Volume 30 [[Link](#)]. 126, 130

- [79] 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](#)]. 178, 191, 223
- [80] Jesper Møller and Rasmus P. Waagepetersen. *An Introduction to Simulation-Based Inference for Spatial Point Processes*. Springer, 2003. 178
- [81] Jesper Møller and Rasmus P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. CRC Press, 2007. 178
- [82] S. Ghosh N., Miyoshi, and T. Shirai. Disordered complex networks: energy optimal lattices and persistent homology. *Preprint*, pages 1–44, 2020. arXiv:2009.08811. 171
- [83] Saralees Nadarajah. A modified Bessel distribution of the second kind. *Statistica*, 67(4):405–413, 2007. [[Link](#)]. 192
- [84] 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](#)]. 188
- [85] Guillermo Navas-Palencia. Optimal binning: mathematical programming formulation. *Preprint*, pages 1–21, 2020. arXiv:2001.08025 [[Link](#)]. 37
- [86] Nathan Ng. Large gaps between the zeros of the Riemann zeta function. *Journal of Number Theory*, 128(3):509–556, 2007. [[Link](#)]. 211
- [87] 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. 177
- [88] Fred Park. Shape descriptor / feature extraction techniques. 2011. UCI iCAMP 2011 [[Link](#)]. 84
- [89] Yuval Peres and Allan Sly. Rigidity and tolerance for perturbed lattices. *Preprint*, pages 1–20, 2020. arXiv:1409.4490 [[Link](#)]. 171, 177
- [90] Carl Rasmussen and Christopher Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006. [[Link](#)]. 52
- [91] Kamron Saniee. A simple expression for multivariate Lagrange interpolation. *SIAM Undergraduate Research Online*, 2007. SIURO [[Link](#)]. 103
- [92] Karl Sigman. Notes on the Poisson process. New York NY, 2009. IEOR 6711: Columbia University course [[Link](#)]. 177
- [93] Luuk Spreeuwiers. *Image Filtering with Neural Networks: Applications and Performance Evaluation*. PhD thesis, University of Twente, 1992. 73
- [94] J. Michael Steele. Le Cam’s inequality and Poisson approximations. *The American Mathematical Monthly*, 101(1):48–54, 1994. [[link](#)]. 141
- [95] Dietrich Stoyan, Wilfrid S. Kendall, Sung Nok Chiu, and Joseph Mecke. *Stochastic Geometry and Its Applications*. Wiley, 2013. 188
- [96] E.C. Titchmarsh and D.R. Heath-Brown. *The Theory of the Riemann Zeta-Function*. Oxford Science Publications, second edition, 1987. 58, 115, 196
- [97] Chris Tofallis. Fitting equations to data with the perfect correlation relationship. *Preprint*, pages 1–11, 2015. Hertfordshire Business School Working Paper [[Link](#)]. 13
- [98] 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
- [99] 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
- [100] Remco van der Hofstad. *Random Graphs and Complex Networks*. Cambridge University Press, 2016. [[Link](#)]. 180
- [101] Yu Vizilter and Sergey Zheltov. Geometrical correlation and matching of 2D image shapes. 2012. ResearchGate [[Link](#)]. 86
- [102] 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](#)]. 147
- [103] Lan Wu, Yongcheng Qi, and Jingping Yang. Asymptotics for dependent Bernoulli random variables. *Statistics and Probability Letters*, pages 455–463, 2012. [[Link](#)]. 126
- [104] Oren Yakir. Recovering the lattice from its random perturbations. *Preprint*, pages 1–18, 2020. arXiv:2002.01508 [[Link](#)]. 177

- [105] 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. 191
- [106] 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
- [107] D. Yogeshwaran. Geometry and topology of the boolean model on a stationary point processes : A brief survey. *Preprint*, pages 1–13, 2018. Researchgate [[Link](#)]. 178
- [108] 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. 173

Index

- α -compositing, 57
- m -interlacing, 71, 178, 186, 189
- A/B testing, 235
- AdaBoost, 36, 236
- additive number theory, 203, 210
- adversarial learning, 233, 238
- AI-art, 155
- algebraic number, 117
- analytic continuation, 197, 204
- analytic function, 115
- anisotropy, 164, 185, 191
- anti-aliasing, 55, 59, 214, 219
- association rule, 227
- attraction (point process), 176
- attraction basin, 58
- attractor distribution, 131, 180, 186, 193
- augmented data, 88, 148, 238
- auto-correlation, 49, 116
- auto-regressive process, 49, 131, 236
- Bailey–Borwein–Plouffe formulas, 117
- Bayesian classification, 74
- Bayesian inference, 44
 - hierarchical models, 237
 - naive Bayes, 227
- Bernoulli trials, 221
- Berry-Esseen inequality, 113
- Bessel function, 192
- Beurling primes, 116, 207
- binning, 226
 - optimum binning, 37, 236
- binomial distribution, 186
- bisection method (root finding), 138
- bootstrapping, 15, 96, 187
 - percentile method, 101
- boundary effect, 173, 174, 179, 180, 184–186, 190
- Brownian motion, 51, 126, 130, 146, 147, 162, 231, 236
 - Lévy flight, 131
- Brun's theorem, 203
- Cauchy distribution, 131
- Cauchy-Riemann equations, 115
- causality, 237
- Cayley-Hamilton theorem, 47
- CDF regression, 17
- censored data, 179
- central limit theorem, 131, 171, 221
- chaotic dynamical system, 146, 162
- character
 - principal, 199
- characteristic function, 192
- characteristic polynomial, 47, 49, 50, 132
- Chebyshev's bias (prime numbers), 115, 199
- checksum, 234
- Chi-squared test, 173
- Chowla conjecture, 201
- classification, 238
- clique (graph theory), 181
- cluster process, 175, 179, 186
- clustering, 238
- Collatz conjecture, 123
- collision graph, 162
- color model
 - RGB, 55, 220
 - RGBA, 55, 56, 76, 220
- color opacity, 155
- color transparency, 20, 148, 220
- complex random variable, 112, 207
- computational complexity, 228
- computer vision, 12, 83
- confidence band, 187
- confidence interval, 44, 184, 236
- confidence level, 221, 224
- confidence region, 15, 29, 184, 185, 221
 - dual region, 44, 222, 236
- conformal map, 14
- confusion matrix, 226, 238
- connected components, 163, 179–181, 185, 186, 229
- contour level, 155, 224
- contour plot, 224
- convergence
 - abscissa, 199, 204
 - absolute, 196, 197
 - alternating series, 197
 - conditional, 114, 199
 - Dirichlet test, 197
- convergence acceleration, 59
- convex linear combination, 108
- convolution of distributions, 192
- counting measure, 172
- covariance matrix, 90, 221
- covering (stochastic), 188
- covering problem, 187
- credible interval, 44
- credible region (Bayesian), 222, 236
- cross-validation, 15, 162, 226
- cuban primes, 203
- curve fitting, 26
- data video, 146, 155

decision tree, 36
 Dedekind zeta function, 115, 207
 deep neural network, 73, 237
 dense set (topology), 201
 density estimation, 178
 diamond-square algorithm, 147
 Diehard tests of randomness, 114
 dimensionality reduction, 16
 Dirichlet character, 115, 116, 198, 203
 modulo 4, 199, 204, 205
 Dirichlet eta function, 211
 Dirichlet functional equation, 115, 204, 205
 Dirichlet series, 112
 Dirichlet's theorem, 115, 199, 201, 205
 Dirichlet-*L* function, 115, 198, 205
 dissimilarity metric, 228
 distributed architecture, 225
 distribution
 Cauchy, 131
 Fréchet, 51, 131
 Gaussian, 221
 generalized logistic, 90, 175
 Hotelling, 222
 Laplace, 192
 logistic, 16
 Lévy, 131
 modified Bessel, 192
 Poisson-binomial, 141, 194
 Poisson-exponential, 171
 Rademacher, 112, 113
 Rayleigh, 186, 187, 193
 Weibull, 51, 131, 186
 domain of attraction, 180
 dot product, 14
 dummy variable, 36
 dyadic map, 116
 dynamical systems, 116, 180
 chaotic systems, 146, 162
 dyadic map, 116
 ergodicity, 116
 logistic map, 116
 shift map, 116
 stochastic, 147
 edge effect (statistics), 179
 eigenvalue, 13, 52, 90, 238
 power iteration, 92
 elbow rule, 135, 179, 186, 231
 elliptic curve, 203
 empirical distribution, 16, 96, 171, 174, 180, 186, 191, 193, 205
 multivariate, 113
 empirical quantiles, 101
 ensemble methods, 36, 83
 entropy, 165, 191, 227
 equidistribution modulo 1, 119
 equilibrium distribution, 147
 Erdős-Rényi model, 181
 ergodicity, 116, 147, 184, 186, 192
 Euler product, 112, 197, 204
 random, 207
 Euler's transform, 211
 evolutionary process, 146
 experimental design, 235
 experimental math, 56, 195
 explainable AI, 13, 74, 83, 90, 135, 150, 188, 232
 exploratory analysis, 234
 exponential decay, 40
 exponential sums, 204
 extrapolation, 108
 extreme value theory, 131, 193
 feature attribution, 232
 feature importance, 232
 feature selection, 15, 97, 225
 Fermat's last theorem, 204
 fixed-point algorithm, 59, 89, 135, 237
 flag vector, 227, 235
 Fourier transform, 192
 fractal dimension, 51
 fractional part function, 118
 Frobenius norm, 90
 Fruchterman and Rheingold algorithm, 229
 Fréchet distribution, 51, 131
 fuzzy classification, 56
 Gamma function, 51, 131
 Gaussian circle problem, 210
 Gaussian distribution, 221
 Gaussian mixture, 70
 Gaussian primes, 115, 206
 Gaussian process, 49, 236
 general linear model, 13
 generalized linear model, 13, 48
 generalized logistic distribution, 90, 185
 generative adversarial networks, 150, 233
 generative model, 155, 238
 geostatistics, 102
 Glivenko-Cantelli theorem, 205
 Goldbach's conjecture, 203
 goodness-of-fit, 56, 226
 GPU-based clustering, 71
 gradient (optimization), 135
 gradient boosting, 236
 gradient operator, 15
 graph, 179
 collision graph, 162
 connected components, 163, 185, 229
 directed, 163
 edge, 179
 Fruchterman-Reingold, 163
 nearest neighbor graph, 181, 185
 node, 179, 181
 random graph, 180
 random nearest neighbor graph, 180
 tree, 163
 undirected, 179–181, 186
 vertex, 179
 graph database, 229
 graph theory, 179
 GraphViz, 163
 greedy algorithm, 210

grid search, 135, 149
 half-tone (music), 217
 Hartman–Wintner theorem, 126
 hash table, 165, 190, 227, 228
 sparse, 227
 Hausdorff distance, 87
 hexagonal lattice, 179
 hidden decision trees, 36, 37, 236
 hidden layer, 73
 hidden process, 171, 189, 193
 hierarchical clustering, 73, 228
 Hilbert primes, 206
 histogram equalization, 71, 73
 Hoeffding inequality, 129
 homogeneity (point process), 178
 Hotelling distribution, 222
 Hurst exponent, 51
 hyperparameter, 29, 56, 103, 149
 identifiability, 189, 191
 ill-conditioned problem, 26, 52, 92, 237
 image segmentation, 73
 index
 index discrepancy, 191
 intensity (stochastic process), 171, 178, 185
 interarrival times, 130, 171, 180, 184, 191
 standardized, 192
 interlaced processes, 178
 Internet of Things, 171
 inverse distance weighting, 104
 inverse square law, 159
 iterated logarithm, 113, 114, 126
 Itô integral, 52
 K-means clustering, 31, 32
 key-value pair, 37, 227
 Kolmogorov-Smirnov test, 113, 173, 180
 Kronecker's theorem, 201, 209
 Lagrange interpolation, 52
 Lagrange multiplier, 15, 237
 Laplace distribution, 192
 Lasso regression, 15, 238
 lattice, 177
 perturbed lattice, 171
 shifted, 179
 stretched, 179
 law of the iterated logarithm, 113, 114, 126, 201, 207
 Le Cam's theorem, 141, 172
 least absolute residuals, 101
 link function, 13, 16
 Liouville function, 198, 209
 log-polar map, 14
 logistic distribution, 16, 178
 logistic map, 116
 logistic regression, 16
 unsupervised, 33
 logit function, 237
 Lévy distribution, 131
 Lévy flight, 131
 Map-reduce, 225
 marketing attribution, 235
 Markov chain, 49
 MCMC, 112
 Mathematica, 224
 MaxCliqueDyn algorithm, 181
 maximum likelihood estimation, 223, 237, 238
 mean squared error, 15, 30
 medoid, 31
 Mersenne twister, 29, 116, 119, 128
 Mertens function, 198
 minimum contrast estimation, 149, 188, 191, 223
 mixture model, 29, 45, 147, 155, 171, 178, 179, 186, 224, 238
 blending, 147
 model fitting, 56, 237
 model identifiability, 15
 modulus (complex number), 132, 197
 Monte Carlo simulations, 112, 237
 morphing (computer vision), 146
 moving average, 137
 multiplicative function
 completely multiplicative, 112, 114, 198, 199, 210
 Rademacher, 112
 Möbius function, 198
 N-body problem, 159
 n-gram (NLP), 228
 naive Bayes, 227, 236
 natural language processing, 36, 228
 nearest neighbor interpolation, 101, 104
 nearest neighbors, 171, 181, 187, 237
 nearest neighbor distances, 185–187, 189, 193
 nearest neighbor graph, 185
 NetworkX, 163
 neural network, 73
 hidden layer, 73
 hyperparameter, 75
 neuron, 73, 237
 sparse, 69
 very deep, 73
 Newton's method, 135
 node (decision tree), 37, 236
 perfect node, 44
 usable node, 38
 normal number, 113, 201, 205
 strongly normal, 114
 numerical stability, 47
 Omega function, 198, 204
 order statistics, 193
 ordinary least squares, 49, 101
 outliers, 193, 231
 overfitting, 15, 191, 236
 palette, 146, 220
 parametric bootstrap, 20, 29, 97, 187, 236
 partial least squares, 13
 path (graph theory), 179
 percentile bootstrap, 101
 permutation

entropy, 191
 random permutation, 190
 perturbed lattices, 171
 Plotly, 155
 point count distribution, 172, 185, 188
 point process
 attractive, 186
 cluster process
 Matérn, 177
 Neyman-Scott, 177
 non-homogeneous, 178
 perturbed lattice process, 177
 radial, 178
 renewal process, 177
 repulsive, 176
 Poisson point process, 130, 171, 185
 Poisson-binomial distribution, 141, 171, 194
 Poisson-exponential distribution, 171
 positive semidefinite (matrix), 48, 91
 power iteration, 92
 preconditioning, 92
 prediction interval, 15, 96, 101
 predictive power, 37, 44, 226, 227, 232
 prime test (of randomness), 114, 127
 principal component analysis, 47, 232, 236
 probability generating function, 127
 proxy space, 224
 pseudo-inverse matrix, 48
 pseudo-random numbers, 128, 231
 congruential generator, 119
 Diehard tests, 114
 Mersenne twister, 119, 128, 148
 prime test, 114, 127
 strongly random, 114, 117
 TestU01, 114
 Pólya conjecture, 200
 quadratic irrational, 116, 119
 quantile, 222, 237
 empirical, 101
 weighted, 101
 quantile function, 171, 175, 186
 quantile regression, 15
 R-squared, 15, 149
 Rademacher distribution, 113
 Rademacher function, 112, 201, 207
 random, 114
 random graph, 180, 181
 random multiplicative function, 112
 Rademacher, 114
 random permutation, 190
 random variable
 complex, 112
 random walk, 126, 149, 236
 first hitting time, 127, 130
 zero crossing, 126
 Rayleigh distribution, 186, 187, 193
 Rayleigh test, 186
 records, 193
 regression splines, 13
 regular expression, 228, 234
 reinforcement learning, 238
 renewal process, 177
 repulsion (point process), 176, 187
 repulsion basin, 197
 resampling, 96, 187
 Riemann Hypothesis, 107
 Generalized, 114, 199, 203, 205
 Riemann zeta function, 112, 115, 205, 207
 root mean squared error, 56
 scaling factor, 185, 193
 seed (random number generator), 148
 semi-supervised learning, 238
 shape signature, 84
 Shapley value, 232
 Shepard's method, 104
 shift map, 116
 sigmoid function, 237
 simplex, 208
 singular value decomposition, 13, 238
 singularity, 165
 six degrees of separation, 230
 smoothing parameter, 103
 spatial statistics, 102, 177
 spectral domain, 147
 square root (matrix), 48, 91
 square-free integer, 113, 201
 stable distribution, 131, 148, 192
 state space, 147
 stationary distribution, 52
 stationary process, 49, 131, 147, 162, 173, 178, 185
 stepwise regression, 98
 stochastic convergence, 147
 stochastic function, 51
 stochastic geometry, 188
 stochastic process, 171
 stochastic residues, 189
 stop word (NLP), 228
 stretching (point process), 179
 superimposition (point processes), 178
 supervised classification, 71
 surface plot, 155
 swarm optimization, 27, 237
 synthetic data, 13, 27, 29, 88, 90, 127, 135, 148, 155, 162, 195, 222
 synthetic metric, 227
 Tarjan's algorithm, 229
 tensor, 74
 text normalization, 228
 Theil-Sen estimator, 101
 time series, 50
 auto-regressive, 50, 131
 disaggregation, 107
 Hurst exponent, 51
 non-periodic, 25
 total least squares, 13
 training set, 101, 162, 226
 transcendental number, 117
 transformer, 73

tree (graph theory), 163
twin primes, 203

universality property, 197, 201, 203
unsupervised clustering, 71
unsupervised learning, 33, 238

validation set, 15, 56, 101, 162, 226
Vandermonde matrix, 47, 52
vertex, 171, 179, 180, 193
video compression
 FFmpeg, 55, 59

Waring's problem, 203
Watts and Strogatz model, 230
Weibull distribution, 51, 131, 186, 193
weighted least squares, 13
weighted quantiles, 101
weighted regression, 16
white noise, 27, 49, 131, 236
wide data, 237

XOR operator, 119