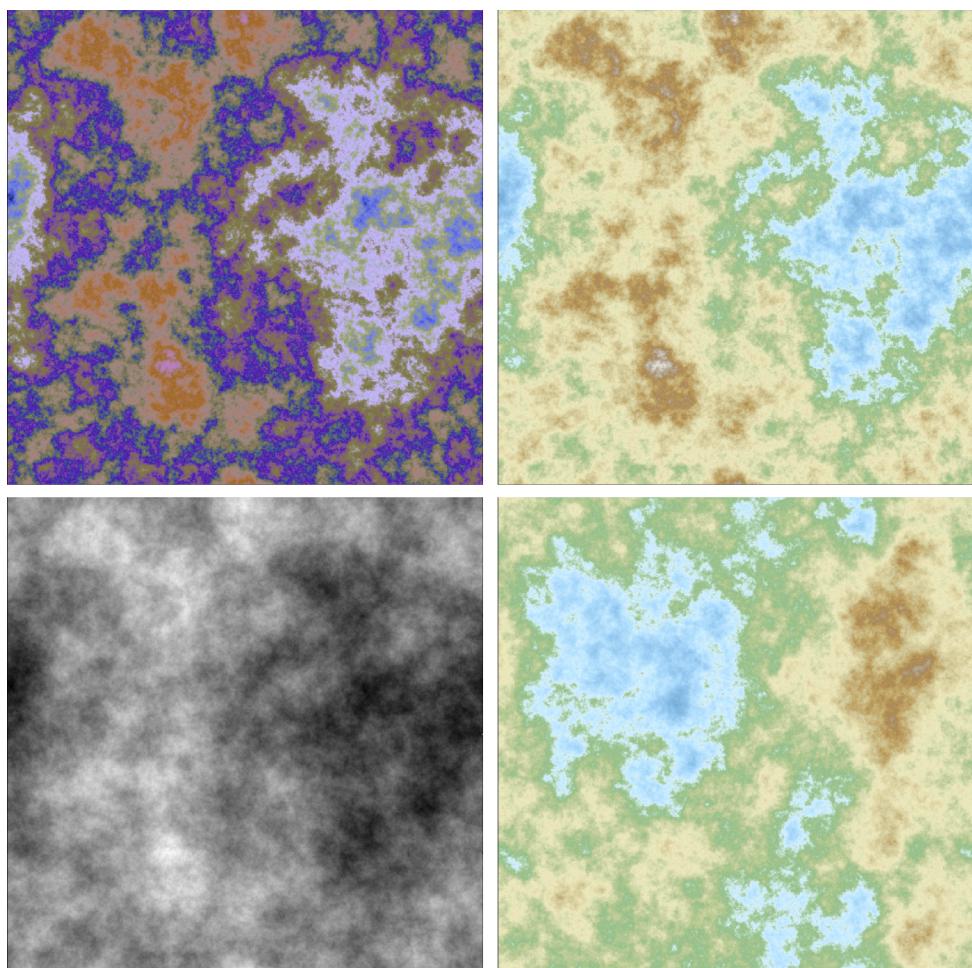

Synthetic Data



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 11 and 12 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 13 and 15 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	8
List of Tables	10
1 Machine Learning Cloud Regression and Optimization	11
1.1 Introduction: circle fitting	11
1.1.1 Previous versions of my method	12
1.2 Methodology, implementation details and caveats	13
1.2.1 Solution, R-squared and backward compatibility	13
1.2.2 Upgrades to the model	14
1.3 Case studies	15
1.3.1 Logistic regression, two ways	15
1.3.2 Ellipsoid and hyperplane fitting	16
1.3.2.1 Curve fitting: 250 examples in one video	16
1.3.2.2 Confidence region for the fitted ellipse	17
1.3.2.3 Python code	18
1.3.3 Non-periodic sum of periodic time series	24
1.3.3.1 Numerical instability and how to fix it	25
1.3.3.2 Python code	26
1.3.4 Fitting a line in 3D, unsupervised clustering, and other generalizations	27
1.3.4.1 Example: confidence region for the cluster centers	28
1.3.4.2 Exact solution and caveats	29
1.3.4.3 Comparison with K-means clustering	30
1.3.4.4 Python code	32
2 A Simple, Robust and Efficient Ensemble Method	35
2.1 Introduction	35
2.2 Methodology	36
2.2.1 How hidden decision trees (HDT) work	36
2.2.2 NLP case study: summary and findings	37
2.2.3 Parameters	38
2.2.4 Improving the methodology	38
2.3 Implementation details	38
2.3.1 Correcting for bias	38
2.3.1.1 Time-adjusted scores	39
2.3.2 Excel spreadsheet	39
2.3.3 Python code and dataset	39
2.4 Model-free confidence intervals and perfect nodes	43
2.4.1 Interesting asymptotic properties of confidence intervals	43
3 Gentle Introduction to Linear Algebra	45
3.1 Power of a matrix	45
3.2 Examples, generalization, and matrix inversion	46
3.2.1 Example with a non-invertible matrix	47
3.2.2 Fast computations	47
3.2.3 Square root of a matrix	47
3.3 Application to machine learning problems	48
3.3.1 Markov chains	48
3.3.2 Time series: auto-regressive processes	48
3.3.3 Linear regression	48
3.4 Mathematics of auto-regressive time series	49

3.4.1	Simulations: curious fractal time series	49
3.4.1.1	White noise: Fréchet, Weibull and exponential cases	50
3.4.1.2	Illustration	50
3.4.2	Solving Vandermonde systems: a numerically stable method	51
3.5	Math for Machine Learning: Must-Read Books	52
4	The Art of Visualizing High Dimensional Data	53
4.1	Introduction	53
4.2	Applications	54
4.2.1	Spatial time series	54
4.2.2	Prediction intervals in any dimensions	54
4.2.3	Supervised classification of an infinite dataset	55
4.2.3.1	Machine learning perspective	55
4.2.3.2	Six challenging problems	56
4.2.3.3	Mathematical background: the Riemann Hypothesis	56
4.2.3.4	Partial solutions to the six challenging problems	57
4.2.4	Algorithms with chaotic convergence	58
4.3	Python code	58
4.3.1	Path simulation	58
4.3.2	Visual convergence analysis in 2D	61
4.3.3	Supervised classification	62
4.4	Visualizations	65
5	Fast Classification and Clustering via Image Convolution Filters	68
5.1	Introduction	68
5.2	Generating the synthetic data	69
5.2.1	Simulations with logistic distribution	69
5.2.2	Mapping the raw observations onto an image bitmap	70
5.3	Classification and unsupervised clustering	70
5.3.1	Supervised classification based on convolution filters	71
5.3.2	Clustering based on histogram equalization	71
5.3.3	Fractal classification: deep neural network analogy	72
5.3.4	Generalization to higher dimensions	73
5.3.5	Towards a very fast implementation	73
5.4	Python code	74
5.4.1	Fractal classification	75
5.4.2	GPU classification and clustering	77
5.4.3	Home-made graphic library	79
6	Shape Classification via Explainable AI	82
6.1	Introduction	82
6.2	Mathematical foundations	82
6.3	Shape signature	83
6.3.1	Weighted centroid	83
6.3.2	Computing the signature	84
6.3.3	Example	85
6.4	Shape comparison	85
6.4.1	Shape classification	86
6.5	Application	86
6.6	Exercises	87
7	Synthetic Data, Interpretable Regression, and Submodels	88
7.1	Introduction	88
7.2	Synthetic data sets and the spreadsheet	89
7.2.1	Correlation structure	89
7.2.2	Standardized regression	90
7.2.3	Initial conditions	90
7.2.4	Simulations and Excel spreadsheet	90
7.3	Damping schedule and convergence acceleration	91
7.3.1	Spreadsheet implementation	91
7.3.2	Interpretable regression with no overfitting	92
7.3.3	Adaptive damping	92
7.4	Performance assessment on synthetic data	92

7.4.1	Results	93
7.4.2	Distribution-free confidence intervals	95
7.4.2.1	Parametric bootstrap	96
7.5	Feature selection	96
7.5.1	Combinatorial approach	96
7.5.2	Stepwise approach	97
7.6	Conclusion	98
8	From Interpolation to Fuzzy Regression	100
8.1	Introduction	100
8.2	Original version	101
8.3	Full, non-linear model in higher dimensions	101
8.3.1	Geometric proximity, weights, and numerical stability	102
8.3.2	Predicted values and prediction intervals	102
8.3.3	Illustration, with spreadsheet	103
8.3.3.1	Output fields	104
8.4	Results	104
8.4.1	Performance assessment	104
8.4.2	Visualization	105
8.4.3	Amplitude restoration	105
8.5	Exercises	106
8.6	Python source code and datasets	107
9	Detecting Subtle Departures from Randomness	111
9.1	Introduction	111
9.2	Pseudo-random numbers	112
9.2.1	Strong pseudo-random numbers	112
9.2.1.1	New test of randomness for PRNGs	113
9.2.1.2	Theoretical background: the law of the iterated logarithm	113
9.2.1.3	Connection to the Generalized Riemann Hypothesis	113
9.2.2	Testing well-known sequences	114
9.2.2.1	Reverse-engineering a pseudo-random sequence	115
9.2.2.2	Illustrations	116
9.3	Python code	118
9.3.1	Fixes to the faulty random function in Python	118
9.3.2	Prime test implementation to detect subtle flaws in PRNG's	118
9.3.3	Special formula to compute 10 million digits of $\sqrt{2}$	121
10	Some Unusual Random Walks	125
10.1	Symmetric unbiased constrained random walks	125
10.1.1	Three fundamental properties of pure random walks	125
10.1.2	Random walks with more entropy than pure random signal	126
10.1.2.1	Applications	126
10.1.2.2	Algorithm to generate quasi-random sequences	127
10.1.2.3	Variance of the modified random walk	127
10.1.3	Random walks with less entropy than pure random signal	128
10.2	Related stochastic processes	129
10.2.1	From Brownian motions to clustered Lévy flights	129
10.2.2	Integrated Brownian motions and special auto-regressive processes	130
10.3	Python code	131
10.3.1	Computing probabilities and variances attached to S_n	131
10.3.2	Path simulations	132
11	Synthetic Terrain Generation and AI-generated Art	134
11.1	Introduction	134
11.2	Terrain generation and the evolutionary process	136
11.2.1	Morphing and non-linear palette operations	136
11.2.2	The diamond-square algorithm	136
11.2.3	The evolutionary process	137
11.2.4	Finding optimum parameters	137
11.2.5	Mimicking real terrain: the synthesis step	137
11.3	Python code	138
11.3.1	Producing data videos with four sub-videos in parallel	138

11.3.2	Main program	139
11.4	AI-generated art with 3D contours	143
11.4.1	Python code using Matplotlib	143
11.4.2	Python code using Plotly	145
11.4.3	Tips to quickly solve new problems	146
12	Synthetic Star Clusters Generation with Collision Graphs	147
12.1	Introduction	147
12.2	Model parameters and simulation results	148
12.2.1	Explanation of color codes	148
12.2.2	Detailed description of top parameters	148
12.2.3	Interesting parameter sets	149
12.3	Analysis of star collisions and collision graph	150
12.3.1	Weighted directed graphs: visualization with NetworkX	151
12.3.2	Interesting findings: how the universe got started	152
12.4	Animated data visualizations	152
12.5	Python code and computational issues	153
12.5.1	Simulating the real and synthetic universes	153
12.5.2	Visualizing collision graphs	157
13	Perturbed-Lattice Point Process: Inference, Nearest Neighbor Graph	159
13.1	Perturbed lattices: definition and properties	159
13.1.1	Point counts distribution	160
13.1.2	Periodicity and amplitude of point count expectations	160
13.1.3	Testing the independence of point counts	161
13.1.3.1	Results and Interpretation	162
13.1.3.2	About the Spreadsheet	163
13.2	Cluster processes and nearest neighbor graphs	163
13.2.1	Synthetic, semi-rigid cluster structures	163
13.2.2	Python code to generate cluster processes	165
13.2.3	References on cluster processes	165
13.2.4	Superimposed perturbed lattices: an alternative to mixture models	166
13.2.4.1	Hexagonal lattice, nearest neighbors	167
13.2.4.2	Exercises: nearest neighbor graphs, size of connected components	168
13.2.4.3	Python code to compute connected components	169
13.3	Statistical inference for point processes	171
13.3.1	Estimation of Core Parameters	171
13.3.1.1	Intensity	172
13.3.1.2	Scaling factor	172
13.3.1.3	Alternative estimation method	172
13.3.2	Spatial statistics, nearest neighbors, clustering	173
13.3.2.1	Inference for two-dimensional processes	173
13.3.2.2	Other possible tests	173
13.3.2.3	Rayleigh test	174
13.3.2.4	Exercises	175
13.4	Special topics	176
13.4.1	Minimum contrast estimation and explainable AI	176
13.4.2	Model identifiability, hard-to-detect patterns	177
13.4.2.1	Stochastic residues	177
13.4.3	Hidden model and random permutations	177
13.4.4	Retrieving the F distribution	179
13.4.4.1	Theoretical values obtained by simulations	179
13.4.4.2	Retrieving F from the interarrival times distribution	180
13.4.5	Record distances between an observed point and its vertex	180
13.4.5.1	Distribution of records	181
13.4.5.2	Distribution of arrival times for records	182
14	New Perspective on the Riemann Hypothesis	183
14.1	Introduction	183
14.1.1	Key concepts and terminology	184
14.1.2	Orbits and holes	184
14.1.3	Industrial Applications	184

14.2	Euler products	185
14.2.1	Finite Euler Products	185
14.2.1.1	Generalization using Dirichlet characters	186
14.2.2	Infinite Euler products	187
14.2.2.1	Special products	187
14.2.2.2	Probabilistic properties and conjectures	188
14.3	Finite Dirichlet series and generalizations	189
14.3.1	Finite Dirichlet series	189
14.3.2	Non-trivial cases with infinitely many primes and a hole	191
14.3.2.1	Sums of two cubes, or cuban primes	191
14.3.2.2	Primes associated to elliptic curves	191
14.3.2.3	Analytic continuation, convergence, and functional equation	192
14.3.2.4	Hybrid Dirichlet-Taylor series	192
14.3.3	Riemann Hypothesis with cosines replaced by wavelets	193
14.3.4	Riemann Hypothesis for Beurling primes	194
14.3.5	Stochastic Euler products	195
14.4	Exercises	196
14.5	Python code	199
14.5.1	Computing the orbit of various Dirichlet series	199
14.5.2	Creating videos of the orbit	202
15	Miscellaneous Topics	205
15.1	The sound that data makes	205
15.1.1	From data visualizations to videos to data music	205
15.1.2	References	206
15.1.3	Python code	206
15.2	Data videos and enhanced visualizations in R	207
15.2.1	Cairo library to produce better charts	207
15.2.2	AV library to produce videos	208
15.3	Dual confidence regions	209
15.3.1	Case study	209
15.3.2	Standard confidence region	209
15.3.3	Dual confidence region	210
15.3.4	Simulations	210
15.3.5	Original problem with minimum contrast estimators	211
15.3.6	General shape of confidence regions	212
15.4	Fast feature selection based on predictive power	213
15.4.1	How cross-validation works	214
15.4.2	Measuring the predictive power of a feature	214
15.4.3	Efficient implementation	215
15.5	Natural language processing: taxonomy creation	216
15.5.1	Designing a keyword taxonomy	216
15.5.2	Fast clustering algorithm for keyword data	217
15.5.2.1	Computational complexity	217
15.5.2.2	Smart crawling of the whole Internet and a bit of graph theory	217
15.6	Automated detection of outliers and number of clusters	219
15.6.1	Black-box elbow rule to detect outliers	219
15.7	New developments in explainable AI	220
15.8	Advice to beginners	221
15.8.1	Getting started and learning how to learn	221
15.8.1.1	Getting help	221
15.8.1.2	Beyond Python	222
15.8.2	Automated data cleaning and exploratory analysis	222
15.8.3	Example of simple analysis: marketing attribution	223
15.8.4	Upcoming books and courses	223
Glossary		224
Bibliography		227
Index		232

List of Figures

1.1	Fitted ellipse (blue), given the training set (red) distributed around a partial arc	17
1.2	Confidence region in blue, $n = 30$ training set points; 50 training sets (left) vs 150 (right)	18
1.3	Three non-periodic time series made of periodic terms (see section 14.2.2.1)	24
1.4	Training set (red), validation set (orange), fitted curve (blue) and model (gray)	25
1.6	Biased confidence region for (θ_A^*, θ_B^*) ; same example as in Figure 1.5; true value is $(0.5, 1.0)$	28
1.5	Finding the two centers θ_A^*, θ_B^* in sample 39; $n = 1000$	29
1.7	Challenging mixture, requiring $p_A = 3, p_B = 1$ to identify the two cluster centers	30
2.1	Output from the Excel version of HDT	40
3.1	AR models, classified based on the types of roots of the characteristic polynomial	51
4.1	Scatterplot observations vs. predicted values, with prediction intervals (in any dimension)	65
4.2	Comets orbiting the sun: simulation	65
4.3	Comets orbiting the sun: snapshot in time	66
4.4	Three orbits of $\eta(\sigma + it)$: $\sigma = 0.5$ (red), 0.75 (blue) and 1.25 (yellow)	66
4.5	Sample orbit points of $\eta(\sigma + it)$: $\sigma = 0.5$ (red), 0.75 (blue) and 1.25 (yellow)	66
4.6	Sample orbit points of $\eta(\sigma + it)$: $\sigma = 0.5$ (red), 0.75 (blue) and 1.25 (yellow)	67
4.7	Raw orbit points of $\eta(\sigma + it)$: $\sigma = 0.5$ (red), 0.75 (blue) and 1.25 (yellow)	67
4.8	Convergence of partial sums of $\eta(z)$, for six $z = \sigma + it$ in the complex plane	67
5.1	Special interlacing of 4 lattice processes with $s = 0$	70
5.2	Classification of left dataset; $s = 0.15, w = 10$. One loop (middle) vs 3 (right).	71
5.3	Clustering of left dataset; $s = 0.15$, 3 loops, $w = 10$ (middle) vs 20 (right).	72
5.4	Classification ($w = 10$) and clustering ($w = 20$); $s = 0.05$, three loops.	72
5.5	Fractal classification, $s = 0.15$. Loop 6, 250 and 400.	73
5.6	Fractal classification, $s = 0.05$.Loop: 6 and 60.	73
5.7	Fast (left) vs standard method (right), 3 loops, $s = 0.15, w = 10$	74
5.8	Fast method, $s = 0.05, w = 20$. Three loops (middle), one loop (right).	74
6.1	Comparing two shapes	83
6.2	Weighted centroid, shape signature	84
6.3	Weight function used in Figure 6.2	85
6.4	Another interesting shape	86
7.1	Regression coefficients oscillating when using adaptive damping	93
7.2	Convergence of regression coefficients (left) and distribution of residual error (right)	94
7.3	Goodness-of-fit: training set (right) versus validation set (left)	94
8.1	Fuzzy regression with prediction intervals, original version, 1D	101
8.2	Fuzzy regression with prediction intervals, full model, 2D	103
8.3	Scatterplots: median vs weighted method, on validation (left) vs training set (right)	105
8.4	Dirichlet eta function (real part, bottom) and interpolation error (top)	107
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) . .	114
9.2	$L_3^*(n)$ test statistic for four sequences: Python[200] and SQRT[90,91] fail	116
9.3	$ L_3(n) $ test statistic for four sequences: Python[200] and SQRT[90,91] fail	116
10.1	Typical path S_n with $0 \leq n \leq 50,000$ for four types of random walks	126
10.2	$\delta_n = 1 - \text{Var}[S_{n+1}] + \text{Var}[S_n]$ for four types of random walks, with $0 \leq n \leq 5000$	127

10.3	Same as Figure 10.2, using a more aesthetic but less meaningful chart type	128
10.4	Clustered Brownian process	130
10.5	AR models, classified based on the types of roots of the characteristic polynomial	131
11.1	Six frames from the terrain video, each one containing four images	135
11.2	Contour plot, 3D mixture model, produced with Plotly	143
11.3	Same as Figure 11.2, produced with Matplotlib	144
12.1	Collisions graph for the biggest star eater (star 47) in video 7	151
12.2	Summary statistics for the whole collision structure: the X axis represents the time	152
12.3	Snapshots of universe 4 (left) and universe 7 (right)	153
13.1	Period and amplitude of $\phi_\tau(t)$; here $\tau = 1, \lambda = 1.4, s = 0.3$	161
13.2	A new test of independence (R-squared version)	161
13.3	Radial cluster process ($s = 0.2, \lambda = 1$) with centers in blue; zoom in on the left	164
13.4	Radial cluster process ($s = 2, \lambda = 1$) with centers in blue; zoom in on the left	164
13.5	Manufactured marble lacking true lattice randomness (left)	164
13.6	Four superimposed Poisson-binomial processes: $s = 0$ (left), $s = 5$ (right)	167
13.7	Rayleigh test to assess if a point distribution matches that of a Poisson process	175
13.8	Realization of a 5-interlacing with $s = 0.15$ and $\lambda = 1$: original (left), modulo $2/\lambda$ (right)	178
13.9	Locally random permutation σ ; $\tau(k)$ is the index of X_k 's closest neighbor to the right	178
13.10	Each arrow links a point (blue) to its vertex (red): $s = 0.2$ (left), $s = 1$ (right)	181
13.11	Distance between a point and its vertex ($\lambda = s = 1$)	182
14.1	Three orbits ($\sigma = 0.5, 0.75, 1.25$) with finite Euler product: $P = \{2, 3\}$ (left) vs $\{2, 3, 5\}$ (right)	186
14.2	Distance between orbit and location $(c, 0)$ depending on t on the X-axis	188
14.3	Distance between orbit and location $(c, 0)$ depending on t on the X-axis	188
14.4	Distance between orbit and location $(c, 0)$ depending on t on the X-axis	188
14.5	Four orbits where the “hole” (repulsion basin) is apparent	190
14.6	Three orbits with “hole” closer to the origin, showing impact of $\beta > \frac{1}{2}$ and larger n	190
14.7	Orbit of Dirichlet eta $\eta(z)$ when cosines are replaced by other periodic functions	194
15.1	Data linked to the melody: red curve for note frequencies, blue curve for note durations	206
15.2	R plot before Cairo (left), and after (right)	207
15.3	Intermediate (left) and last frame (right) of the video	208
15.4	Example of 90% dual confidence region for (p, q)	210
15.5	Minimum contrast estimation for (λ, s) using (p, q) as proxy stats	211
15.6	Non-elliptic confidence regions with various confidence levels	212
15.7	Elbow rule (right) finds $m = 3$ clusters in Brownian motion (left)	220

List of Tables

1.1	Estimated ellipse parameters vs true values ($n = 30$), for shape in Figure 1.2	18
1.2	First and last step of <code>curve_fitting</code> , approaching the model.	26
1.3	MSE for different methods and θ s, same data set as in Figure 1.5	31
1.4	MSE for different methods and θ s, same data set as in Figure 1.7	31
2.1	List of potential features to use in the model	36
2.2	Statistics for selected HDT nodes (Excel version)	39
2.3	Order of magnitude for the expectation and standard deviation of the range R_n	43
3.1	Characteristic polynomials used in the simulations	50
7.1	Regression coefficients and performance metrics r, s based on methodology	95
7.2	Correlation matrix	95
7.3	Best performance given m (number of features)	96
7.4	Feature comparison table (top 32 feature combinations)	98
7.5	Feature comparison table (bottom 31 feature combinations)	99
8.1	R-squared ρ^2 and slope β , on training and validation sets, median vs weighted	105
9.1	$L_3^*(n)$, for various sequences ($n = 20,000$); “Fail” means failing the prime test	117
12.1	Description of top parameters used in the star cluster simulator	149
12.2	Eight selected parameter sets covering various situations	150
13.1	Variance attached to F_s , as a function of s	160
13.2	Poisson process ($s = \infty$) versus $s = 39.85$	180
15.1	Extract of the mapping table used to recover (λ, s) from (p, q)	212
15.2	Eight bins: 2 features (A, B) times 2 outcomes (Good/Bad)	214
15.3	Amount of data collected at each level, when crawling the Internet	218

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 48, 96, 130
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 36, 72, 214
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 35, 224
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 14, 95, 175, 224
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 11, 14, 17, 19, 28, 172, 173, 209, 212
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 14, 36, 92, 98, 150, 214, 224
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 35, 36, 38, 40, 224, 225
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 11, 15, 225, 226
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 15, 95, 112, 159, 162, 168, 174, 179, 181, 193
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 35, 82, 224
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 12, 68, 73, 82, 89, 138, 176, 220

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 11, 14, 36, 93, 96, 205, 213, 224, 226
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 14, 55, 92, 93, 214, 224, 226
Gradient methods	Iterative optimization techniques to find the minimum or 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 14, 30, 54, 88
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 69, 73, 152, 217, 218
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 28, 55, 69, 74, 100, 137, 225
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 12, 15, 225
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 15, 32, 35, 39, 224, 225
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 68, 72, 74, 82, 100, 224, 225
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 15.5 deals with creating a keyword taxonomy based on parsing Google search results pages. See pages 35, 216
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 11, 13, 58
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 14, 91, 100, 224, 226

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 37, 39, 43, 213, 215, 220, 225
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 11, 14, 55, 89, 92, 94, 96, 103, 225, 226
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 28, 111, 118, 219
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 11, 12, 14, 15, 18, 26, 35, 39, 45, 48, 51, 55, 88, 94, 100, 107, 225, 226
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 226
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 11, 12, 14, 16, 26, 28, 32, 47, 54, 68, 69, 74, 87, 93, 104, 126, 136, 143, 150, 210, 219, 224
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 68, 73
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 12, 14, 16, 19, 28, 35, 39, 55, 71, 87, 94, 100, 104, 150, 214, 224, 225, 226
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 14, 26, 40, 55, 92, 100, 150, 214, 224, 225, 226

Bibliography

- [1] Weighted percentiles using numpy. *Forum discussion*, 2020. StackOverflow [\[Link\]](#). 100
- [2] Jan Ackmann et al. Machine-learned preconditioners for linear solvers in geophysical fluid flows. *Preprint*, pages 1–19, 2020. arXiv:2010.02866 [\[Link\]](#). 92
- [3] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. Wiley, fourth edition, 2016. [\[Link\]](#). 168
- [4] José M. Amigó, Roberto Dale, and Piergiulio Tempesta. A generalized permutation entropy for random processes. *Preprint*, pages 1–9, 2012. arXiv:2003.13728 [\[Link\]](#). 179
- [5] Luc Anselin. *Point Pattern Analysis: Nearest Neighbor Statistics*. The Center for Spatial Data Science, University of Chicago, 2016. Slide presentation [\[Link\]](#). 166
- [6] Hui Guo at al. Eyes tell all: Irregular pupil shapes reveal gan-generated faces. *Preprint*, pages 1–7, 2021. arXiv:2109.00162 [\[Link\]](#). 221
- [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. [\[Link\]](#). 165
- [8] N. Balakrishnan and C.R. Rao (Editors). *Order Statistics: Theory and Methods*. North-Holland, 1998. 168, 182
- [9] Christopher Beckham and Christopher Pal. A step towards procedural terrain generation with GANs. *Preprint*, pages 1–5, 2017. arXiv:1707.03383 [\[Link\]](#). 135
- [10] Rabi Bhattacharya and Edward Waymire. *Random Walk, Brownian Motion, and Martingales*. Springer, 2021. 125
- [11] Barbara Bogacka. *Lecture Notes on Time Series*. 2008. Queen Mary University of London [\[Link\]](#). 48
- [12] B. Bollobas and P. Erdös. Cliques in random graphs. *Mathematical Proceedings of the Cambridge Philosophical Society*, 80(3):419–427, 1976. [\[Link\]](#). 169
- [13] Miklos Bona. *Combinatorics of Permutations*. Routledge, second edition, 2012. 179
- [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. [\[Link\]](#). 187
- [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. [\[Link\]](#). 189
- [16] Oliver Bröker and Marcus J. Groteb. Sparse approximate inverse smoothers for geometric and algebraic multigrid. *Applied Numerical Mathematics*, 41(1):61–80, 2002. 89
- [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. [\[Link\]](#). 199
- [18] Bartłomiej Błaszczyzyn and Dhandapani Yogeshwaran. Clustering and percolation of point processes. *Preprint*, pages 1–20, 2013. Project Euclid [\[Link\]](#). 165
- [19] Bartłomiej Błaszczyzyn and Dhandapani Yogeshwaran. On comparison of clustering properties of point processes. *Preprint*, pages 1–26, 2013. arXiv:1111.6017 [\[Link\]](#). 165
- [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 [\[Link\]](#). 165
- [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. 220
- [22] Keith Conrad. *L-functions and the Riemann Hypothesis*. 2018. 2018 CTNT Summer School [\[Link\]](#). 114, 184, 187, 192
- [23] Noel Cressie. *Statistic for Spatial Data*. Wiley, revised edition, 2015. 165

- [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. [129](#)
- [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. [129](#)
- [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. [211](#)
- [27] Marc Deisenroth, A. Faisal, and Cheng Soon Ong. *Mathematics for Machine Learning*. Cambridge University Press, 2020. [\[Link\]](#). [52](#)
- [28] Harold G. Diamond and Wen-Bin Zhang. *Beurling Generalized Numbers*. American Mathematical Society, 2016. Mathematical Surveys and Monographs, Volume 213 [\[Link\]](#). [115](#), [195](#)
- [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. [166](#)
- [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. [166](#)
- [31] David Coupier (Editor). *Stochastic Geometry: Modern Research Frontiers*. Wiley, 2019. [176](#)
- [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. [167](#)
- [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. [\[Link\]](#). [169](#)
- [34] Achim Zeileis et al. Colorsphere: A toolbox for manipulating and assessing colors and palettes. *Preprint*, pages 1–45, 2019. arXiv:1903.06490 [\[Link\]](#) [\[R Library\]](#). [135](#)
- [35] Arash Farahmand. *Math 55 Lecture Notes*. 2021. University of Berkeley [\[Link\]](#). [46](#), [52](#)
- [36] W. Feller. On the Kolmogorov-Smirnov limit theorems for empirical distributions. *Annals of Mathematical Statistics*, 19(2):177–189, 1948. [\[Link\]](#). [168](#), [175](#)
- [37] Nikos Frantzikinakis. Ergodicity of the Liouville system implies the Chowla conjecture. *Preprint*, pages 1–41, 2016. arXiv [\[Link\]](#). [189](#)
- [38] P. M. Gauthier. Approximating the Riemann zeta-function by polynomials with restricted zeros. *Canadian Mathematical Bulletin*, 62(3):475–478, 2018. [\[Link\]](#). [199](#)
- [39] P. A. Van Der Geest. The binomial distribution with dependent Bernoulli trials. *Journal of Statistical Computation and Simulation*, pages 141–154, 2004. [\[Link\]](#). [126](#)
- [40] Stamatia Giannarou and Tania Stathaki. Shape signature matching for object identification invariant to image transformations and occlusion. 2007. ResearchGate [\[Link\]](#). [83](#)
- [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 [\[Link\]](#). [169](#)
- [42] B.V. Gnedenko and A. N. Kolmogorov. *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, 1954. [130](#)
- [43] Manuel González-Navarrete and Rodrigo Lambert. Non-markovian random walks with memory lapses. *Preprint*, pages 1–14, 2018. arXiv [\[Link\]](#). [125](#)
- [44] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. [\[Link\]](#). [52](#)
- [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. [\[Link\]](#). [166](#)
- [46] Vincent Granville. *Applied Stochastic Processes, Chaos Modeling, and Probabilistic Properties of Numeration Systems*. MLTechniques.com, 2018. [\[Link\]](#). [115](#)
- [47] Vincent Granville. *Stochastic Processes and Simulations: A Machine Learning Perspective*. MLTechniques.com, 2022. [\[Link\]](#). [50](#), [58](#), [130](#), [159](#), [160](#), [161](#), [163](#), [167](#), [169](#), [195](#), [199](#), [212](#), [223](#)
- [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. [71](#)
- [49] Vincent Granville and Richard L Smith. Disaggregation of rainfall time series via Gibbs sampling. *NISS Technical Report*, pages 1–21, 1996. [\[Link\]](#). [106](#)
- [50] Kristen Grauman. Shape matching. 2008. University of Texas, Austin [\[Link\]](#). [86](#)
- [51] Radim Halir and Jan Flusser. Numerically stable direct least squares fitting of ellipses. *Preprint*, pages 1–8, 1998. [\[Link\]](#). [16](#), [18](#)
- [52] Peter Hall. *Introduction to the theory of coverage processes*. Wiley, 1988. [176](#)

- [53] Adam J. Harper. Moments of random multiplicative functions, II: High moments. *Algebra and Number Theory*, 13(10):2277–2321, 2019. [\[Link\]](#). 111, 195
- [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. [\[Link\]](#). 111, 113, 195
- [55] Adam J. Harper. Almost sure large fluctuations of random multiplicative functions. *Preprint*, pages 1–38, 2021. arXiv [\[Link\]](#). 113, 189, 195
- [56] K. Hartmann, J. Krois, and B. Waske. *Statistics and Geospatial Data Analysis*. Freie Universität Berlin, 2018. E-Learning Project SOGA [\[Link\]](#). 162
- [57] D. R. Heath-Brown. Primes represented by $x^3 + 2y^3$. *Acta Mathematica*, 186:1–84, 2001. [\[Link\]](#). 191
- [58] T. W. Hilberdink and M. L. Lapidus. Beurling Zeta functions, generalised primes, and fractal membranes. *Preprint*, pages 1–31, 2004. arXiv [\[Link\]](#). 114, 115, 195
- [59] Christian Hill. *Learning Scientific Programming with Python*. Cambridge University Press, 2016. [\[Link\]](#). 18
- [60] Robert V. Hogg, Joseph W. McKean, and Allen T. Craig. *Introduction to Mathematical Statistics*. Pearson, eighth edition, 2016. [\[Link\]](#). 52
- [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. [\[Link\]](#). 210
- [62] Peter Humphries. The distribution of weighted sums of the Liouville function and Pólya’s conjecture. *Preprint*, pages 1–33, 2011. arXiv [\[Link\]](#). 196
- [63] Timothy D. Johnson. Introduction to spatial point processes. *Preprint*, page 2008. NeuroImaging Statistics Oxford (NISOx) group [\[Link\]](#)[\[Mirror\]](#). 166
- [64] Chigozie Kelechi. Towards efficiency in the residual and parametric bootstrap techniques. *American Journal of Theoretical and Applied Statistics*, 5(5), 2016. [\[Link\]](#). 96
- [65] Denis Kojevnikov, Vadim Marmer, and Kyungchul Song. Limit theorems for network dependent random variables. *Journal of Econometrics*, 222(2):419–427, 2021. [\[Link\]](#). 166
- [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. 180
- [67] Faraj Lagum. *Stochastic Geometry-Based Tools for Spatial Modeling and Planning of Future Cellular Networks*. PhD thesis, Carleton University, 2018. [\[Link\]](#). 165
- [68] Günther Last and Mathew Penrose. *Lectures on the Poisson Process*. Cambridge University Press, 2017. 165
- [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. [\[Link\]](#). 111, 113
- [70] Jing Lei et al. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113:1094–1111, 2018. [\[Link\]](#). 96
- [71] G. Last M.A. Klatt and D. Yogeshwaran. Hyperuniform and rigid stable matchings. *Random Structures and Algorithms*, 2:439–473, 2020. [\[Link\]](#)[\[PowerPoint\]](#). 165
- [72] Jorge Mateu, Frederic P Schoenberg, and David M Diez. On distances between point patterns and their applications. *Preprint*, pages 1–29, 2010. [\[Link\]](#). 166
- [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 [\[Link\]](#). 169
- [74] Masahiro Mine. Probability density functions attached to random Euler products for automorphic L-functions. *Preprint*, pages 1–38, 2020. arXiv [\[Link\]](#). 195, 196
- [75] Christoph Molnar. *Interpretable Machine Learning*. ChristophMolnar.com, 2022. [\[Link\]](#). 96, 220
- [76] Marc-Andreas Muendler. Linear difference equations and autoregressive processes. 2000. University of Berkeley [\[Link\]](#). 48
- [77] V. Kumar Murty. Seminar on Fermat’s last theorem. In *Canadian Mathematical Society – Conference Proceedings*, volume 17, Toronto, Canada, 1995. [\[Link\]](#). 192
- [78] Peter Mörters and Yuval Peres. *Brownian Motion*. Cambridge University Press, 2010. Cambridge Series in Statistical and Probabilistic Mathematics, Volume 30 [\[Link\]](#). 125, 129

- [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](#)]. 166, 179, 211
- [80] Jesper Møller and Rasmus P. Waagepetersen. *An Introduction to Simulation-Based Inference for Spatial Point Processes*. Springer, 2003. 166
- [81] Jesper Møller and Rasmus P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. CRC Press, 2007. 166
- [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. 159
- [83] 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](#)]. 176
- [84] Guillermo Navas-Palencia. Optimal binning: mathematical programming formulation. *Preprint*, pages 1–21, 2020. arXiv:2001.08025 [[Link](#)]. 36
- [85] Nathan Ng. Large gaps between the zeros of the Riemann zeta function. *Journal of Number Theory*, 128(3):509–556, 2007. [[Link](#)]. 199
- [86] 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. 165
- [87] Fred Park. Shape descriptor / feature extraction techniques. 2011. UCI iCAMP 2011 [[Link](#)]. 83
- [88] Yuval Peres and Allan Sly. Rigidity and tolerance for perturbed lattices. *Preprint*, pages 1–20, 2020. arXiv:1409.4490 [[Link](#)]. 159, 165
- [89] Carl Rasmussen and Christopher Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006. [[Link](#)]. 51
- [90] Kamron Saniee. A simple expression for multivariate Lagrange interpolation. *SIAM Undergraduate Research Online*, 2007. SIURO [[Link](#)]. 102
- [91] Karl Sigman. Notes on the Poisson process. New York NY, 2009. IEOR 6711: Columbia University course [[Link](#)]. 165
- [92] Luuk Spreeuwers. *Image Filtering with Neural Networks: Applications and Performance Evaluation*. PhD thesis, University of Twente, 1992. 72
- [93] Dietrich Stoyan, Wilfrid S. Kendall, Sung Nok Chiu, and Joseph Mecke. *Stochastic Geometry and Its Applications*. Wiley, 2013. 176
- [94] E.C. Titchmarsh and D.R. Heath-Brown. *The Theory of the Riemann Zeta-Function*. Oxford Science Publications, second edition, 1987. 57, 114, 184
- [95] Chris Tofallis. Fitting equations to data with the perfect correlation relationship. *Preprint*, pages 1–11, 2015. Hertfordshire Business School Working Paper [[Link](#)]. 12
- [96] 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](#)]. 13, 16
- [97] 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](#)]. 16
- [98] Remco van der Hofstad. *Random Graphs and Complex Networks*. Cambridge University Press, 2016. [[Link](#)]. 168
- [99] Yu Vizilter and Sergey Zheltov. Geometrical correlation and matching of 2D image shapes. 2012. ResearchGate [[Link](#)]. 85
- [100] 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](#)]. 135
- [101] Lan Wua, Yongcheng Qi, and Jingping Yang. Asymptotics for dependent Bernoulli random variables. *Statistics and Probability Letters*, pages 455–463, 2012. [[Link](#)]. 125
- [102] Oren Yakir. Recovering the lattice from its random perturbations. *Preprint*, pages 1–18, 2020. arXiv:2002.01508 [[Link](#)]. 165
- [103] 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. 179
- [104] 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](#)]. 46

- [105] D. Yogeshwaran. Geometry and topology of the boolean model on a stationary point processes : A brief survey. *Preprint*, pages 1–13, 2018. Researchgate [[Link](#)]. [166](#)
- [106] 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. [161](#)

Index

- α -compositing, 56
 m -interlacing, 70, 166, 174, 177
- A/B testing, 223
AdaBoost, 35, 224
additive number theory, 191, 198
adversarial learning, 221, 226
AI-art, 143
algebraic number, 116
analytic continuation, 185, 192
analytic function, 114
anisotropy, 153, 173, 179
anti-aliasing, 54, 58, 202, 207
association rule, 215
attraction (point process), 164
attraction basin, 57
attractor distribution, 130, 168, 174, 181
augmented data, 87, 136, 226
auto-correlation, 48, 115
auto-regressive process, 48, 130, 224
- Bailey–Borwein–Plouffe formulas, 116
Bayesian classification, 73
Bayesian inference, 43
 - hierarchical models, 225
 - naive Bayes, 215Bernoulli trials, 209
Berry-Esseen inequality, 112
Bessel function, 180
Beurling primes, 115, 195
binning, 214
 - optimum binning, 36, 224binomial distribution, 174
bootstrapping, 14, 95, 175
 - percentile method, 100boundary effect, 161, 162, 167, 168, 172–174, 178
Brownian motion, 50, 125, 129, 134, 135, 150, 219, 224
 - Lévy flight, 130Brun's theorem, 191
- Cauchy distribution, 130
Cauchy-Riemann equations, 114
causality, 225
Cayley-Hamilton theorem, 46
CDF regression, 16
censored data, 167
central limit theorem, 130, 159, 209
chaotic dynamical system, 134, 150
character
 - principal, 187characteristic function, 180
- characteristic polynomial, 46, 48, 49, 131
Chebyshev's bias (prime numbers), 114, 187
checksum, 222
Chi-squared test, 161
Chowla conjecture, 189
classification, 226
clique (graph theory), 169
cluster process, 163, 167, 174
clustering, 226
Collatz conjecture, 122
collision graph, 151
color model
 - RGB, 54, 208
 - RGBA, 54, 55, 75, 208color opacity, 143
color transparency, 19, 136, 208
complex random variable, 111, 195
computational complexity, 216
computer vision, 11, 82
confidence band, 175
confidence interval, 43, 172, 224
confidence level, 209, 212
confidence region, 14, 28, 172, 173, 209
 - dual region, 43, 210, 224conformal map, 13
confusion matrix, 214, 226
connected components, 151, 167–169, 173, 174, 217
contour level, 143, 212
contour plot, 212
convergence
 - abscissa, 187, 192
 - absolute, 184, 185
 - alternating series, 185
 - conditional, 113, 187
 - Dirichlet test, 185convergence acceleration, 58
convex linear combination, 107
convolution of distributions, 180
counting measure, 160
covariance matrix, 89, 209
covering (stochastic), 176
covering problem, 175
credible interval, 43
credible region (Bayesian), 210, 224
cross-validation, 14, 150, 214
cuban primes, 191
curve fitting, 25
- data video, 134, 143
decision tree, 35

Dedekind zeta function, 114, 195
 deep neural network, 72, 225
 dense set (topology), 189
 density estimation, 166
 diamond-square algorithm, 135
 Diehard tests of randomness, 113
 dimensionality reduction, 15
 Dirichlet character, 114, 115, 186, 191
 modulo 4, 187, 192, 193
 Dirichlet eta function, 199
 Dirichlet functional equation, 114, 192, 193
 Dirichlet series, 111
 Dirichlet's theorem, 114, 187, 189, 193
 Dirichlet- L function, 114, 186, 193
 dissimilarity metric, 216
 distributed architecture, 213
 distribution
 Cauchy, 130
 Fréchet, 50, 130
 Gaussian, 209
 generalized logistic, 89, 163
 Hotelling, 210
 Laplace, 180
 logistic, 15
 Lévy, 130
 modified Bessel, 180
 Poisson-binomial, 182
 Poisson-exponential, 159
 Rademacher, 111, 112
 Rayleigh, 174, 175, 181
 Weibull, 50, 130, 174
 domain of attraction, 168
 dot product, 13
 dummy variable, 35
 dyadic map, 115
 dynamical systems, 115, 168
 chaotic systems, 134, 150
 dyadic map, 115
 ergodicity, 115
 logistic map, 115
 shift map, 115
 stochastic, 135
 edge effect (statistics), 167
 eigenvalue, 12, 51, 89, 226
 power iteration, 91
 elbow rule, 167, 174, 219
 elliptic curve, 191
 empirical distribution, 15, 95, 159, 162, 168, 174, 179, 181, 193
 multivariate, 112
 empirical quantiles, 100
 ensemble methods, 35, 82
 entropy, 153, 179, 215
 equidistribution modulo 1, 118
 equilibrium distribution, 135
 Erdős-Rényi model, 169
 ergodicity, 115, 135, 172, 174, 180
 Euler product, 111, 185, 192
 random, 195
 Euler's transform, 199
 evolutionary process, 134
 experimental design, 223
 experimental math, 55, 183
 explainable AI, 12, 73, 82, 89, 138, 176, 220
 exploratory analysis, 222
 exponential decay, 39
 exponential sums, 192
 extrapolation, 107
 extreme value theory, 130, 181
 feature attribution, 220
 feature importance, 220
 feature selection, 14, 96, 213
 Fermat's last theorem, 192
 fixed-point algorithm, 58, 88, 225
 flag vector, 215, 223
 Fourier transform, 180
 fractal dimension, 50
 fractional part function, 117
 Frobenius norm, 89
 Fruchterman and Rheingold algorithm, 217
 Fréchet distribution, 50, 130
 fuzzy classification, 55
 Gamma function, 50, 130
 Gaussian circle problem, 198
 Gaussian distribution, 209
 Gaussian mixture, 69
 Gaussian primes, 114, 194
 Gaussian process, 48, 224
 general linear model, 12
 generalized linear model, 12, 47
 generalized logistic distribution, 89, 173
 generative adversarial networks, 138, 221
 generative model, 143, 226
 geostatistics, 101
 Glivenko-Cantelli theorem, 193
 Goldbach's conjecture, 191
 goodness-of-fit, 55, 214
 GPU-based clustering, 70
 gradient boosting, 224
 gradient operator, 14
 graph, 167
 collision graph, 151
 connected components, 151, 173, 217
 directed, 151
 edge, 167
 Fruchterman-Reingold, 151
 nearest neighbor graph, 169, 173
 node, 167, 169
 random graph, 168
 random nearest neighbor graph, 168
 tree, 152
 undirected, 167–169, 174
 vertex, 167
 graph database, 217
 graph theory, 167
 GraphViz, 151
 greedy algorithm, 198
 grid search, 137
 half-tone (music), 205

Hartman–Wintner theorem, 125
 hash table, 153, 178, 215, 216
 sparse, 215
 Hausdorff distance, 86
 hexagonal lattice, 167
 hidden decision trees, 35, 36, 224
 hidden layer, 72
 hidden process, 159, 177, 181
 hierarchical clustering, 72, 216
 Hilbert primes, 194
 histogram equalization, 70, 72
 Hoeffding inequality, 128
 homogeneity (point process), 166
 Hotelling distribution, 210
 Hurst exponent, 50
 hyperparameter, 28, 55, 102, 137
 identifiability, 177, 179
 ill-conditioned problem, 25, 51, 91, 225
 image segmentation, 72
 index
 index discrepancy, 179
 intensity (stochastic process), 159, 166, 173
 interarrival times, 129, 159, 168, 172, 179
 standardized, 180
 interlaced processes, 166
 Internet of Things, 159
 inverse distance weighting, 103
 inverse square law, 147
 iterated logarithm, 112, 113, 125
 Itô integral, 51
 K-means clustering, 30, 31
 key-value pair, 36, 215
 Kolmogorov–Smirnov test, 112, 161, 168
 Kronecker’s theorem, 189, 197
 Lagrange interpolation, 51
 Lagrange multiplier, 14, 225
 Laplace distribution, 180
 Lasso regression, 14, 226
 lattice, 165
 perturbed lattice, 159
 shifted, 167
 stretched, 167
 law of the iterated logarithm, 112, 113, 125, 189, 195
 Le Cam’s theorem, 160
 least absolute residuals, 100
 link function, 12, 15
 Liouville function, 186, 197
 log-polar map, 13
 logistic distribution, 15, 166
 logistic map, 115
 logistic regression, 15
 unsupervised, 32
 logit function, 225
 Lévy distribution, 130
 Lévy flight, 130
 Map-reduce, 213
 marketing attribution, 223
 Markov chain, 48
 MCMC, 111
 Mathematica, 212
 MaxCliqueDyn algorithm, 169
 maximum likelihood estimation, 211, 225, 226
 mean squared error, 14, 29
 medoid, 30
 Mersenne twister, 28, 115, 118, 127
 Mertens function, 186
 minimum contrast estimation, 137, 176, 179, 211
 mixture model, 28, 44, 135, 143, 159, 166, 167, 174, 212, 226
 blending, 135
 model fitting, 55, 225
 model identifiability, 14
 modulus (complex number), 131, 185
 Monte Carlo simulations, 111, 225
 morphing (computer vision), 134
 multiplicative function
 completely multiplicative, 111, 113, 186, 187, 198
 Rademacher, 111
 Möbius function, 186
 N-body problem, 147
 n-gram (NLP), 216
 naive Bayes, 215, 224
 natural language processing, 35, 216
 nearest neighbor interpolation, 100, 103
 nearest neighbors, 159, 169, 175, 225
 nearest neighbor distances, 173–175, 177, 181
 nearest neighbor graph, 173
 NetworkX, 151
 neural network, 72
 hidden layer, 72
 hyperparameter, 74
 neuron, 72, 225
 sparse, 68
 very deep, 72
 node (decision tree), 36, 224
 perfect node, 43
 usable node, 37
 normal number, 112, 189, 193
 strongly normal, 113
 numerical stability, 46
 Omega function, 186, 192
 order statistics, 181
 ordinary least squares, 48, 100
 outliers, 181, 219
 overfitting, 14, 179, 224
 palette, 134, 208
 parametric bootstrap, 19, 28, 96, 175, 224
 partial least squares, 12
 path (graph theory), 167
 percentile bootstrap, 100
 permutation
 entropy, 179
 random permutation, 178
 perturbed lattices, 159
 Plotly, 143
 point count distribution, 160, 173, 176

point process
 attractive, 174
 cluster process
 Matérn, 165
 Neyman-Scott, 165
 non-homogeneous, 166
 perturbed lattice process, 165
 radial, 166
 renewal process, 165
 repulsive, 164
 Poisson point process, 129, 159, 173
 Poisson-binomial distribution, 159, 182
 Poisson-exponential distribution, 159
 positive semidefinite (matrix), 47, 90
 power iteration, 91
 preconditioning, 91
 prediction interval, 14, 95, 100
 predictive power, 36, 43, 214, 215, 220
 prime test (of randomness), 113, 126
 principal component analysis, 46, 220, 224
 probability generating function, 126
 proxy space, 212
 pseudo-inverse matrix, 47
 pseudo-random numbers, 127, 219
 congruential generator, 118
 Diehard tests, 113
 Mersenne twister, 118, 127, 136
 prime test, 113, 126
 strongly random, 113, 116
 TestU01, 113
 Pólya conjecture, 188
 quadratic irrational, 115, 118
 quantile, 210, 225
 empirical, 100
 weighted, 100
 quantile function, 159, 163, 174
 quantile regression, 14
 R-squared, 14, 137
 Rademacher distribution, 112
 Rademacher function, 111, 189, 195
 random, 113
 random graph, 168, 169
 random multiplicative function, 111
 Rademacher, 113
 random permutation, 178
 random variable
 complex, 111
 random walk, 125, 137, 224
 first hitting time, 126, 129
 zero crossing, 125
 Rayleigh distribution, 174, 175, 181
 Rayleigh test, 174
 records, 181
 regression splines, 12
 regular expression, 216, 222
 reinforcement learning, 226
 renewal process, 165
 repulsion (point process), 164, 175
 repulsion basin, 185
 resampling, 95, 175
 Riemann Hypothesis, 106
 Generalized, 113, 187, 191, 193
 Riemann zeta function, 111, 114, 193, 195
 root mean squared error, 55
 scaling factor, 173, 181
 seed (random number generator), 136
 semi-supervised learning, 226
 shape signature, 83
 Shapley value, 220
 Shepard's method, 103
 shift map, 115
 sigmoid function, 225
 simplex, 196
 singular value decomposition, 12, 226
 singularity, 153
 six degrees of separation, 218
 smoothing parameter, 102
 spatial statistics, 101, 165
 spectral domain, 135
 square root (matrix), 47, 90
 square-free integer, 112, 189
 stable distribution, 130, 136, 180
 state space, 135
 stationary distribution, 51
 stationary process, 48, 130, 135, 151, 161, 166, 173
 stepwise regression, 97
 stochastic convergence, 135
 stochastic function, 50
 stochastic geometry, 176
 stochastic process, 159
 stochastic residues, 177
 stop word (NLP), 216
 stretching (point process), 167
 superimposition (point processes), 166
 supervised classification, 70
 surface plot, 143
 swarm optimization, 26, 225
 synthetic data, 12, 26, 28, 87, 89, 126, 136, 143, 150, 183, 210
 synthetic metric, 215
 Tarjan's algorithm, 217
 tensor, 73
 text normalization, 216
 Theil-Sen estimator, 100
 time series, 49
 auto-regressive, 49, 130
 disaggregation, 106
 Hurst exponent, 50
 non-periodic, 24
 total least squares, 12
 training set, 100, 150, 214
 transcendental number, 116
 transformer, 72
 tree (graph theory), 152
 twin primes, 191
 universality property, 185, 189, 191
 unsupervised clustering, 70

unsupervised learning, 32, 226

validation set, 14, 55, 100, 150, 214

Vandermonde matrix, 46, 51

vertex, 159, 167, 168, 181

video compression

FFmpeg, 54, 58

Waring's problem, 191

Watts and Strogatz model, 218

Weibull distribution, 50, 130, 174, 181

weighted least squares, 12

weighted quantiles, 100

weighted regression, 15

white noise, 26, 48, 130, 224

wide data, 225

XOR operator, 118