

1. Record Nr.	UNINA9910768197303321
Autore	Friedland Gerald
Titolo	Information-Driven Machine Learning : Data Science as an Engineering Discipline // by Gerald Friedland
Pubbl/distr/stampa	Cham : , : Springer International Publishing : , : Imprint : Springer, , 2024
ISBN	3-031-39477-1
Edizione	[1st ed. 2024.]
Descrizione fisica	1 online resource (281 pages)
Disciplina	006.31
Soggetti	Artificial intelligence - Data processing Machine learning Data structures (Computer science) Information theory Expert systems (Computer science) Artificial intelligence Data Science Machine Learning Data Structures and Information Theory Knowledge Based Systems Artificial Intelligence
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Preface -- 1 Introduction -- 2 The Automated Scientific Process -- 3 The (Black Box) Machine Learning Process -- 4 Information Theory -- 5 Capacity -- 6 The Mechanics of Generalization -- 7 Meta-Math: Exploring the Limits of Modeling -- 8 Capacity of Neural Networks -- 8 Capacity of Neural Networks -- 10 Capacities of some other Machine Learning Methods -- 11 Data Collection and Preparation -- 12 Measuring Data Sufficiency -- 13 Machine Learning Operations -- 14 Explainability -- 15 Repeatability and Reproducibility -- 16 The Curse of Training and the Blessing of High Dimensionality -- 16 The Curse of Training and the Blessing of High Dimensionality -- Appendix A Recap: The Logarithm -- Appendix B More on Complexity -- Appendix C Concepts Cheat Sheet -- Appendix D A Review Form that Promotes

This groundbreaking book transcends traditional machine learning approaches by introducing information measurement methodologies that revolutionize the field. Stemming from a UC Berkeley seminar on experimental design for machine learning tasks, these techniques aim to overcome the 'black box' approach of machine learning by reducing conjectures such as magic numbers (hyper-parameters) or model-type bias. Information-based machine learning enables data quality measurements, a priori task complexity estimations, and reproducible design of data science experiments. The benefits include significant size reduction, increased explainability, and enhanced resilience of models, all contributing to advancing the discipline's robustness and credibility. While bridging the gap between machine learning and disciplines such as physics, information theory, and computer engineering, this textbook maintains an accessible and comprehensive style, making complex topics digestible for a broad readership. Information-Driven Machine Learning explores the synergistic harmony among these disciplines to enhance our understanding of data science modeling. Instead of solely focusing on the "how," this text provides answers to the "why" questions that permeate the field, shedding light on the underlying principles of machine learning processes and their practical implications. By advocating for systematic methodologies grounded in fundamental principles, this book challenges industry practices that have often evolved from ideologic or profit-driven motivations. It addresses a range of topics, including deep learning, data drift, and MLOps, using fundamental principles such as entropy, capacity, and high dimensionality. Ideal for both academia and industry professionals, this textbook serves as a valuable tool for those seeking to deepen their understanding of data science as an engineering discipline. Its thought-provoking content stimulates intellectual curiosity and caters to readers who desire more than just code or ready-made formulas. The text invites readers to explore beyond conventional viewpoints, offering an alternative perspective that promotes a big-picture view for integrating theory with practice. Suitable for upper undergraduate or graduate-level courses, this book can also benefit practicing engineers and scientists in various disciplines by enhancing their understanding of modeling and improving data measurement effectively.
