1. Record Nr. UNINA9910896193203321

Autore	Picci Giorgio
Titolo	An Introduction to Statistical Data Science : Theory and Models / / by Giorgio Picci
Pubbl/distr/stampa	Cham : , : Springer Nature Switzerland : , : Imprint : Springer, , 2024
ISBN	9783031666193 3031666194
Edizione	[1st ed. 2024.]
Descrizione fisica	1 online resource (437 pages)
Disciplina	519.5
Soggetti	Statistics Machine learning Engineering mathematics Artificial intelligence - Data processing Statistical Theory and Methods Bayesian Inference Statistical Learning Statistics in Engineering, Physics, Computer Science, Chemistry and Earth Sciences Engineering Mathematics Data Science
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	- 1. Introduction -- 2. Classical Statistical Inference -- 3. Linear Models -- 4. Conditioning and Regularization -- 5. Linear Hypotheses and LDA -- 6. Bayesian Statistics -- 7. Principal Component Analysis -- 8. Non Linear Inference -- 9. Time Series.
Sommario/riassunto	This graduate textbook on the statistical approach to data science describes the basic ideas, scientific principles and common techniques for the extraction of mathematical models from observed data. Aimed at young scientists, and motivated by their scientific prospects, it provides first principle derivations of various algorithms and procedures, thereby supplying a solid background for their future specialization to diverse fields and applications. The beginning of the

book presents the basics of statistical science, with an exposition on linear models. This is followed by an analysis of some numerical aspects and various regularization techniques, including LASSO, which are particularly important for large scale problems. Decision problems are studied both from the classical hypothesis testing perspective and, particularly, from a modern support-vector perspective, in the linear and non-linear context alike. Underlying the book is the Bayesian approach and the Bayesian interpretation of various algorithms and procedures. This is the key to principal components analysis and canonical correlation analysis, which are explained in detail. Following a chapter on nonlinear inference, including material on neural networks, the book concludes with a discussion on time series analysis and estimating their dynamic models. Featuring examples and exercises partially motivated by engineering applications, this book is intended for graduate students in applied mathematics and engineering with a general background in probability and linear algebra.