STATISTICAL INFERENCE AS SEVERE TESTING. HOW TO GET BEYOND THE PROBABILITY AND STATISTICS FOR COMPUTER SCIENCE

Forsyth, David , 2018
Springer International Publishing
eBook ISBN 978-3-319-64410-3.   Hardcover ISBN 978-3-319-64409-7
Softcover ISBN 978-3-319-87788-4.   XXIV, 367

This is a textbook of use in probability and statistics courses for computer science students. It provides the needed background for dealing with stochastic methods for analyzing both qualitative and quantitative data. It contents 15 chapters.

The book introduces the needed background in probability and statistics which are of use in different areas of computing science and data science. Among them are thematics, uncommon, in other text books, as the treatment of simulation (discussing its use for obtaining insight in the probability behavior of the random variables involved) emphasizing in Markov Chains (as they are needed for granting convergence of different simulation algorithms.

Classification, Support Vector Machines, Agglomerative Methods and K-Means Clustering are discussed and vector building for complex signaling and other advanced data analysis tools. Their implementations in practice are discussed.

Basic inference tools, covering point and confidence intervals estimation and tests of hypothesis, are presented.

Among the more classic statistical tools presented are: linear regression, and other approaches in practical problems, Principal Components Analysis, Multivariate Scaling , Principal Coordinate Analysis etc.

The author presented in each chapter worked examples as well as problems to be solved and involved Programming issues.

A.   Santiago
Universidad Autónoma de Guerrero

PROBABILITY AND STATISTICS FOR DATA SCIENCE: MATH + R + DATA

Norman Matloff , 2019
Chapman & Hall/CRC Data Science Series) 1st Edition
ISBN-13: 978-0367260934   ISBN-10: 036726093X XV+426

The book supports the needs of Data Science students for dealing with the basic probability issues (random variables, probability distributions, expectation, mean squared error, etc.). Important Data Science statistical tools are included. Among them multivariate statistical models as linear and logistic regression, Principal Component Analysis , Random Graph Models, Hidden Markov Models, and Neural Networks.

The book will be very useful for non-mathematic data scientists as it minimizes the usual presentation of results using theorems, lemmas, propositions. But do not eliminate mathematics form the curriculum!. Students need to be at home in calculus, matrix algebra, and programming. They should deal at each step with the discussion of each concept and model developed with mathematical precision, without theoretical consenting.

C.N. Bouza-Herrera
Universidad de La Habana