

1. Record Nr.	UNINA9910798272303321
Autore	Massaron Luca
Titolo	Regression analysis with Python : learn the art of regression analysis with Python // Luca Massaron, Alberto Boschetti
Pubbl/distr/stampa	Birmingham : , : Packt Publishing, , 2016
ISBN	1-78398-074-5
Edizione	[1st edition]
Descrizione fisica	1 online resource (312 p.)
Collana	Community experience distilled
Soggetti	Python (Computer program language) Regression analysis
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Includes index.
Nota di contenuto	Cover; Copyright; Credits; About the Authors; About the Reviewers; www.PacktPub.com; Table of Contents; Preface; Chapter 1: Regression - The Workhorse of Data Science; Regression analysis and data science; Exploring the promise of data science; The challenge; The linear models; What you are going to find in the book ; Python for data science; Installing Python; Choosing between Python 2 and Python 3; Step-by-step installation; Installing packages; Package upgrades; Scientific distributions; Introducing Jupyter or IPython; Python packages and functions for linear models ; NumPy; SciPy StatsmodelsScikit-learn; Summary; Chapter 2: Approaching Simple Linear Regression; Defining a regression problem; Linear models and supervised learning; Reflecting on predictive variables; Reflecting on response variables; The family of linear models; Preparing to discover simple linear regression; Starting from the basics; A measure of linear relationship; Extending to linear regression; Regressing with StatsModels; The coefficient of determination; Meaning and significance of coefficients; Evaluating the fitted values; Correlation is not causation; Predicting with a regression model Regressing with Scikit-learnMinimizing the cost function; Explaining the reason for using squared errors; Pseudoinverse and other optimization methods; Gradient Descent at work; Summary; Chapter 3: Multiple Regression in Action; Using multiple features; Model building with Statsmodels; Using formulas as an alternative; The correlation

matrix; Revisiting gradient descent; Feature scaling; Unstandardizing coefficients; Estimating feature importance; Inspecting standardized coefficients; Comparing models by R-squared; Interaction models; Discovering interactions; Polynomial regression
Testing linear versus cubic transformation
Going for higher-degree solutions; Introducing underfitting and overfitting; Summary; Chapter 4: Logistic Regression; Defining a classification problem; Formalization of the problem: binary classification; Assessing the classifier's performance; Defining a probability-based approach; More on the logistic and logit functions; Let's see some code; Pros and cons of logistic regression; Revisiting Gradient Descent; Multiclass Logistic Regression; An example; Summary; Chapter 5: Data Preparation; Numeric feature scaling; Mean centering; Standardization
Normalization
The logistic regression case; Qualitative feature encoding; Dummy coding with Pandas; DictVectorizer and one-hot encoding; Feature hasher; Numeric feature transformation; Observing residuals; Summarizations by binning; Missing data; Missing data imputation; Keeping track of missing values; Outliers; Outliers on the response; Outliers among the predictors; Removing or replacing outliers; Summary; Chapter 6: Achieving Generalization; Checking on out-of-sample data; Testing by sample split; Cross-validation; Bootstrapping; Greedy selection of features ; The Madelon dataset
Univariate selection of features

Sommario/riassunto

Learn the art of regression analysis with Python
About This Book
Become competent at implementing regression analysis in Python
Solve some of the complex data science problems related to predicting outcomes
Get to grips with various types of regression for effective data analysis
Who This Book Is For
The book targets Python developers, with a basic understanding of data science, statistics, and math, who want to learn how to do regression analysis on a dataset. It is beneficial if you have some knowledge of statistics and data science.
What You Will Learn
Format a dataset for regression and evaluate its performance
Apply multiple linear regression to real-world problems
Learn to classify training points
Create an observation matrix, using different techniques of data analysis and cleaning
Apply several techniques to decrease (and eventually fix) any overfitting problem
Learn to scale linear models to a big dataset and deal with incremental data
In Detail
Regression is the process of learning relationships between inputs and continuous outputs from example data, which enables predictions for novel inputs. There are many kinds of regression algorithms, and the aim of this book is to explain which is the right one to use for each set of problems and how to prepare real-world data for it. With this book you will learn to define a simple regression problem and evaluate its performance. The book will help you understand how to properly parse a dataset, clean it, and create an output matrix optimally built for regression. You will begin with a simple regression algorithm to solve some data science problems and then progress to more complex algorithms. The book will enable you to use regression models to predict outcomes and take critical business decisions. Through the book, you will gain knowledge to use Python for building fast better linear models and to apply the results in Python or in any computer language you prefer.
Style and approach
This is a practical tutorial-based book. You will be given an example problem and then supplied with the relevant code and how to walk through it. The details are provided in a step by step manner, followed by a thorough explanation of the math underlying the solution. This approach will help you leverage your own data using the same techniques.
