$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix} = \begin{pmatrix} a_{11} \cdot b_{11} + a_{12} \cdot b_{21} + a_{13} \cdot b_{31} & a_{11} \cdot b_{12} + a_{12} \cdot b_{22} + a_{13} \cdot b_{32} \\ a_{21} \cdot b_{11} + a_{22} \cdot b_{21} + a_{23} \cdot b_{31} & a_{21} \cdot b_{12} + a_{22} \cdot b_{22} + a_{23} \cdot b_{32} \end{pmatrix}$$

Computing Matrix Algebra

practical notes foR data science

MDT-DS

Mario De Toma | April 2020

https://www.mdt-datascience.com/

Computing Matrix Algebra

practical notes foR data science

Mario De Toma

April 2020

Contents

Preface

The booklet you are reading is actually a cheatsheet about computing matrix algebra operations such as matrix multiplication, inversion and factorization.

It is written foR (aspiring) data scientists where with "foR" (capital letter R) I mean the side of data science addicted to R and its gorgeous ecosystem especially including Rcpp, RcppArmadillo and RcppEigen which are actually interfaces to C++ and to the powerful C++ libraries of linear algebra.

Acknowledgments

I wrote this booklet as a personal *#StayAtHome* initiative but I couldn't be less thankful to this world health emergency.

I would like to thank, instead, my wife for everything, not specifically for the psicological support in writing this booklet.

Finally I would like to thank the readers of this booklet because they surely are special people having interest in such a tidy math object: **the matrix**.

Thanks for reading!

Computing Matrix Algebra practical notes foR data science

https://www.mdt-datascience.com/

Chapter 1

introduction

Nobody can be a poet without feeling strong affection for words, at the same time nobody can be serious about data science without becoming close friend to matrices.

The matrix is a tidy mathematical object with its own set of rules, its own algebra that enables knowing matrix structure and interacting with other matrices, vectors and scalars.

This booklet help in strengthen this relationship interacting with matrices by computing some important matrix algebra operations.

1.1 booklet concept

Data are easily represented in computers as math object such matrices and vectors and data scientists interact with data through computational tools.

This booklet is an agile review of algebra operations on matrices through coding in R, Rcpp, Armadillo and Eigen in order to get a better understanding of the matrix object.

It is neither a text about matrix algebra theory (no theorems and proofs) nor a text on matrix numerical methods (numerical stability issue and program optimization topic are not addressed). Some resources for deepen theoretical knowledge are referenced in appendix.

While looping through matrix columns and rows and calling classes and methods of linear algebra libraries the reader could, with some effort (intentionally not all is explained), get more aquainted with the algebra of matrices and so strengthen the practical knowledge of this important math basis for data science.

1.2 structure of chapters

Every chapter covers one operation in matrix algebra. Every chapter is articulated in 5 sections focusing respectively on:

- quick math concept explanation;
- operation computation in R;
- operation implementation in Rcpp;
- computing with Armadillo C++ library;
- computing with Eigen C++ library.

In this way the reader can review the math, experiment function calls in domain specific computational tool, realize the complexity of implementing matrix operation from scratch and get a feel of using optimized libraries.

1.3 R and Rcpp

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS.

Rcpp offers a seamless integration of R and C++. Many R data types and objects can be mapped back and forth to C++ equivalents which facilitates both writing of new code as well as easier integration of third-party libraries.

1.4 C++ linear algebra libraries

One of the main rule in programming is never to reinvent the wheel. Therefore a data scientist needs to be acquainted with available numerical matrix algebra libraries. Armadillo and Eigen are superb choices for C++.

Armadillo is a high quality linear algebra library (matrix maths) for the C++ language, aiming towards a good balance between speed and ease of use.

Eigen is a C++ template library for linear algebra. It supports all matrix sizes, from small fixed-size matrices to arbitrarily large dense matrices, and even sparse matrices. It is fast, reliable and elegant.

1.5 assumption about the reader

In order to fully enjoy the active reading of the booklet, the reader should have:

- some background knowledge of matrix algebra;
- some background knowledge of computer science;
- some familiarity with R and C++;
- a little bit of nerdy personality traits.

1.6 about the author

Mario De Toma holds a master level degree in engineering from *Politecnico di Milano* University and a professional certificate in data science from HarvardX.

Mario gained a huge experience in managing projects, services and people working for IT consulting and service corporations.

The reader can visit the author's blog (https://www.mdt-datascience.com/) for more adventures in data and quantitative analysis.