Distributed Machine Learning in R with Apache Spark: An Introduction Using sparklyr and rsparkling

Abc

Yihui Xie

2018-07-30

Contents

Preface

This is a *sample* book written in **Markdown**. You can use anything that Pandoc's Markdown supports, e.g., a math equation $a^2 + b^2 = c^2$.

The **bookdown** package can be installed from CRAN or Github:

```
install.packages("bookdown")
# or the development version
# devtools::install_github("rstudio/bookdown")
```

Remember each Rmd file contains one and only one chapter, and a chapter is defined by the first-level heading `#`.

To compile this example to PDF, you need XeLaTeX. You are recommended to install TinyTeX (which includes XeLaTeX): https://yihui.name/tinytex/.

Part I

Part I

TBD.

Chapter 1

Introduction to Apache Spark

TBD.

1.1 What is Spark

1.2 Installing Spark

Chapter 2

Interfacing R with Spark

TBD.

2.1 The sparklyr package

2.2 Data wrangling in Spark with dplyr

Chapter 3

Machine Learning Essentials

TBD.

Part II

Part II

TBD.

Chapter 4

Machine learning in Spark via MLlib

TBD.

Chapter 5

Machine learning in Spark via rsparkling

TBD.