

A Minimal Book Example

Grégoire Virepinte

2018-07-24

Contents

Introduction	5
Why this book?	5
Why write this in English?	5
What to find here?	5
1 Prerequisites	7
2 Model Selection	9
2.1 The Bias-Variance Trade-Off	9
2.2 Cross-validation	9
2.3 The use of information criteria	9
2.4 Bootstrap	9
2.5 ROC Curve	9
3 Regression	11
3.1 Linear regression	11
3.2 Logistic regression	11
3.3 Polynomial regression	11
3.4 Constraints: Ridge, Lasso, Elastic Net	11
3.5 Non-linear regression	11
3.6 Random and fixed effects	11
4 Missing values	13
5 Gradient boosting	15
5.1 Fundamental idea	15
5.2 Basic algorithm	15
5.3 An implementation of XGBoost	15
6 Random Forest	17
6.1 Idea	17
6.2 Implementation	17

Introduction

Why this book?

The basic aim of this book is to collect a few revision cards about Machine Learning. Those are mainly things that I have learned during the past year and forgotten not long after, which really is a shame.

Why write this in English?

There are a few reasons for that:

- I could potentially publish this in the future, one day, which means that I might as well write it in English from the very beginning,
- I have been burned enough times with UTF-8-incompatible encoding (accents missing, that sort of thing), and I don't want that to happen again.
- Most of the Machine Learning vocabulary seems to be in English anyway, and I don't want to have to resort to Frenglish.
- Because I can.

What to find here?

This should be the basic idea of the book:

- A succinct mathematical description of the concepts mentioned in the book,
- A special emphasis on the conditions of application if I can find some, because I always seem to have problems with those,
- An example in R with the proper code and libraries.

Chapter 1

Prerequisites

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/>.

Chapter 2

Model Selection

2.1 The Bias-Variance Trade-Off

2.2 Cross-validation

2.3 The use of information criteria

2.4 Bootstrap

2.5 ROC Curve

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter `??`. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter `??`.

Figures and tables with captions will be placed in `figure` and `table` environments, respectively.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure 2.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 2.1.

```
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

You can write citations, too. For example, we are using the **bookdown** package (Xie, 2018) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).

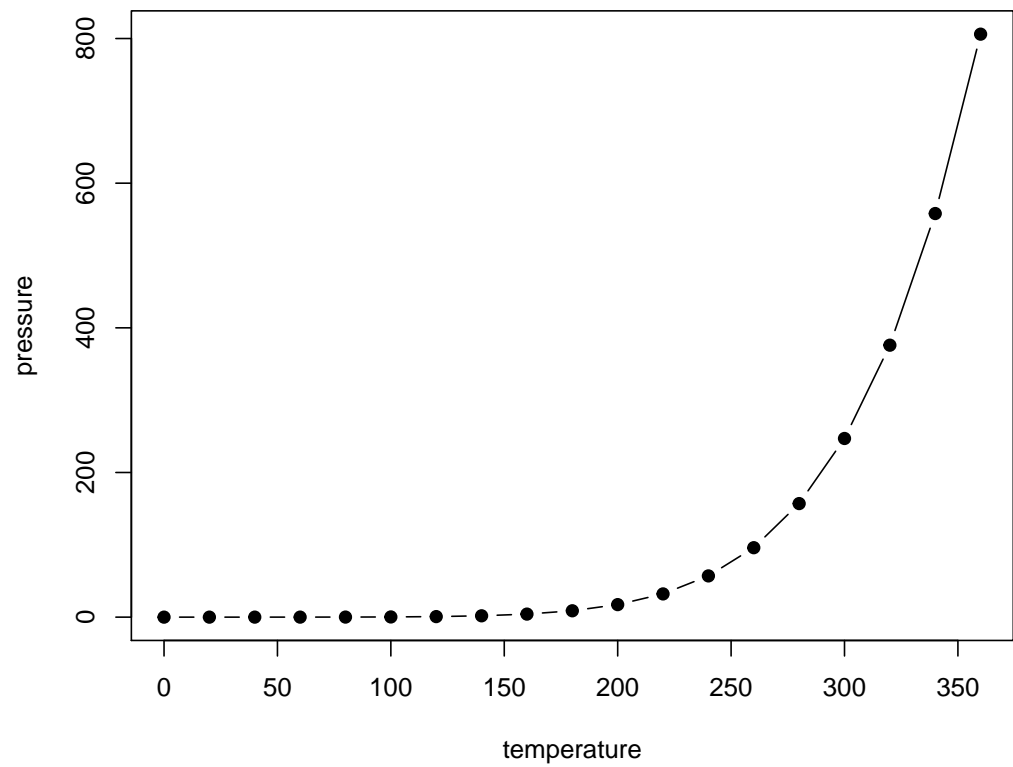


Figure 2.1: Here is a nice figure!

Table 2.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

Chapter 3

Regression

3.1 Linear regression

3.2 Logistic regression

3.3 Polynomial regression

3.4 Constraints: Ridge, Lasso, Elastic Net

3.5 Non-linear regression

3.6 Random and fixed effects

Chapter 4

Missing values

Chapter 5

Gradient boosting

5.1 Fundamental idea

5.2 Basic algorithm

5.3 An implementation of XGBoost

Chapter 6

Random Forest

6.1 Idea

6.2 Implementation

Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2018). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.7.