A Minimal Book Example

Yihui Xie

2021-04-09

Contents

1 Prerequisites **7**

2 Introduction to tidymodels **9**

3 Statististical learning **11**

4 Linear Regression **13**

4.1 Libraries 13

4.2 Simple linear regression 13

4.3 Multiple linear regression 13

4.4 Interaction terms 13

4.5 Non-linear transformations of the predictors 13

4.6 Qualitative predictors 13

4.7 Writing functions 13

5 Classification **15**

5.1 The Stock Market Data 15

5.2 Logistic Regression 15

5.3 Linear Discriminant Analysis 15

5.4 Quadratic Discriminant Analysis 15

5.5 K-Nearest Neighbors 15

5.6 An Application to Caravan Insurance Data 15

6 Resampling Methods **17**

6.1 The Validation Set Approach 17

6.2 Leave-One-Out Cross-Validation 17

6.3 k-Fold Cross-Validation 17

6.4 The Bootstrap 17

7 Linear Model Selection and Regularization **19**

7.1 Best Subset Selection 19

7.2 Forward and Backward Stepwise Selection 19

7.3 Choosing Amoung Models Using the Validation Set Approach and Cross-Validation 19

7.4 Ridge Regression 19

7.5 The Lasso 19

7.6 Principal Components Regression 19

7.7 Partial Least Squares 19

8 Moving Beyond Linearity **21**

8.1 Polynomial Regression and Step Functions 21

8.2 Splines 21

8.3 GAMs 21

9 Tree-Based Methods **23**

9.1 Fitting Classification Trees 23

9.2 Fitting Regression Trees 23

9.3 Bagging and Random Forests 23

9.4 Boosting 23

10 Support Vector Machines **25**

10.1 Support Vector Classifier 25

10.2 Support Vector Machine 25

10.3 ROC Curves 25

10.4 SVM with Multiple Classes 25

10.5 Application to Gene Expression Data 25

11 Unsupervised Learning **27**

11.1 Principal Components Analysis 27

11.2 K-Means Clustering 27

11.3 Hierarchical Clustering 27

11.4 PCA on the NCI60 Data 27

11.5 Clustering the Observations of the NCI60 Data 27

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.org/tinytex/.

Chapter 2

Introduction to tidymodels

Chapter 3

Statististical learning

Here is a review of existing methods.

Chapter 4

Linear Regression

4.1 Libraries

4.2 Simple linear regression

4.3 Multiple linear regression

4.4 Interaction terms

4.5 Non-linear transformations of the predictors

4.6 Qualitative predictors

4.7 Writing functions

Chapter 5

Classification

5.1 The Stock Market Data

5.2 Logistic Regression

5.3 Linear Discriminant Analysis

5.4 Quadratic Discriminant Analysis

5.5 K-Nearest Neighbors

5.6 An Application to Caravan Insurance Data

Chapter 6

Resampling Methods

6.1 The Validation Set Approach

6.2 Leave-One-Out Cross-Validation

6.3 k-Fold Cross-Validation

6.4 The Bootstrap

Chapter 7

Linear Model Selection and Regularization

7.1 Best Subset Selection

7.2 Forward and Backward Stepwise Selection

7.3 Choosing Amoung Models Using the Validation Set Approach and Cross-Validation

7.4 Ridge Regression

7.5 The Lasso

7.6 Principal Components Regression

7.7 Partial Least Squares

Chapter 8

Moving Beyond Linearity

8.1 Polynomial Regression and Step Functions

8.2 Splines

8.3 GAMs

Chapter 9

Tree-Based Methods

9.1 Fitting Classification Trees

9.2 Fitting Regression Trees

9.3 Bagging and Random Forests

9.4 Boosting

Chapter 10

Support Vector Machines

10.1 Support Vector Classifier

10.2 Support Vector Machine

10.3 ROC Curves

10.4 SVM with Multiple Classes

10.5 Application to Gene Expression Data

Chapter 11

Unsupervised Learning

11.1 Principal Components Analysis

11.2 K-Means Clustering

11.3 Hierarchical Clustering

11.4 PCA on the NCI60 Data

11.5 Clustering the Observations of the NCI60 Data