Machine Learning Algorithms on the Wisconsin Diagnostic Dataset

Shaojun Xie ( )

2021-11-01

Contents

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

Introduction

The machine learning algorithms were trained to detect breast cancer using the Wisconsin Diagnostic Breast Cancer (WDBC) dataset[20]. The dataset consists of features which were computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. The said features describe the characteristics of the cell nuclei found in the image.

```
dat <- read.csv("data/WBCD_data.csv")
knitr::kable(dat)
```

id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothnes
842302	M	17.990	10.38	122.80	1001.0	
842517	M	20.570	17.77	132.90	1326.0	
84300903	M	19.690	21.25	130.00	1203.0	
84348301	M	11.420	20.38	77.58	386.1	
84358402	M	20.290	14.34	135.10	1297.0	
843786	M	12.450	15.70	82.57	477.1	
844359	M	18.250	19.98	119.60	1040.0	
84458202	M	13.710	20.83	90.20	577.9	
844981	M	13.000	21.82	87.50	519.8	
84501001	M	12.460	24.04	83.97	475.9	
845636	M	16.020	23.24	102.70	797.8	
84610002	M	15.780	17.89	103.60	781.0	
846226	M	19.170	24.80	132.40	1123.0	
846381	M	15.850	23.95	103.70	782.7	
84667401	M	13.730	22.61	93.60	578.3	
84799002	M	14.540	27.54	96.73	658.8	
848406	M	14.680	20.13	94.74	684.5	
84862001	M	16.130	20.68	108.10	798.8	
849014	M	19.810	22.15	130.00	1260.0	
8510426	B	13.540	14.36	87.46	566.3	
8510653	B	13.080	15.71	85.63	520.0	
8510824	B	9.504	12.44	60.34	273.9	
8511133	M	15.340	14.26	102.50	704.4	
851509	M	21.160	23.04	137.20	1404.0	
852552	M	16.650	21.38	110.00	904.6	
852631	M	17.140	16.40	116.00	912.7	
852763	M	14.580	21.53	97.41	644.8	
852781	M	18.610	20.25	122.10	1094.0	
852973	M	15.300	25.27	102.40	732.4	
853201	M	17.570	15.05	115.00	955.1	
853401	M	18.630	25.11	124.80	1088.0	
853612	M	11.840	18.70	77.93	440.6	
85382601	M	17.020	23.98	112.80	899.3	
854002	M	19.270	26.47	127.90	1162.0	
854039	M	16.130	17.88	107.00	807.2	
854253	M	16.740	21.59	110.10	869.5	
854268	M	14.250	21.72	93.63	633.0	
854941	B	13.030	18.42	82.61	523.8	
855133	M	14.990	25.20	95.54	698.8	
855138	M	13.480	20.82	88.40	559.2	
855167	M	13.440	21.58	86.18	563.0	
855563	M	10.950	21.35	71.90	371.1	
855625	M	19.070	24.81	128.30	1104.0	
856106	M	13.280	20.28	87.32	545.2	
85638502	M	13.170	21.81	85.42	531.5	
857010	M	18.650	17.60	123.70	1076.0	
85713702	B	8.196	16.84	51.71	201.9	
85715	M	13.170	18.66	85.98	534.6	
857155	B	12.050	14.63	78.04	449.3	
857156	B	13.490	22.30	86.91	561.0	
857343	B	11.760	21.60	74.72	427.9	
857373	B	13.640	16.34	87.21	571.8	
857374	B	11.940	18.24	75.71	437.6	

2.2 Check how may variables are there.

```
dim(dat)
```

```
## [1] 569  33
```

There are `R nrow(dat)` rows and “

Chapter 3

Literature

Here is a review of existing methods.

Chapter 4

Methods

We describe our methods in this chapter.

Math can be added in body using usual syntax like this

4.1 math example

p is unknown but expected to be around 1/3. Standard error will be approximated

$$SE = \sqrt{(\frac{p(1-p)}{n})} \approx \sqrt{\frac{1/3(1-1/3)}{300}} = 0.027$$

You can also use math in footnotes like this[1].

We will approximate standard error to 0.027[2]

[1] where we mention $p = \frac{a}{b}$

[2] p is unknown but expected to be around 1/3. Standard error will be approximated

$$SE = \sqrt{(\frac{p(1-p)}{n})} \approx \sqrt{\frac{1/3(1-1/3)}{300}} = 0.027$$

Chapter 5

Applications

Some *significant* applications are demonstrated in this chapter.

5.1 Example one

5.2 Example two

Chapter 6

Final Words

We have finished a nice book.