

January 27, 2021

ERRATA

churnTrain & churnTest

Scripts 6.2, 6.3, 6.5, 6.6, and 7.3 as well as several end-of-chapter exercises make use of the datasets *churnTrain* and *churnTest* described in Section 6.2.2. Prior to R Version 4, these datasets were included in the C50 package. The factor ordering for the output variable *churn* was modified so as to associate factor 1 with ‘yes’ and 2 with ‘no’. The purpose of the modified ordering was to place emphasis on the rare outcome—yes.

These datasets are now included in your *datasets & databases .zip* file. When you first import these datasets into R Studio, you will see the default ordering for *churn* where factor 1 is associated with ‘no’ and 2 with ‘yes’. Script 6.2 contains the code necessary to change the factor ordering so it matches the original data. Once the code in Script 6.2 executes, the output of all scripts associated with these datasets will match what is seen in your text. You can easily verify the change in factor assignment by looking at the output of *str(churnTrain)* or *str(churnTest)* prior to and after executing *script 6.2*.

Factor Variables

Prior to R Version 4, data frames automatically represented character data as factors. With Version 4, this is no longer the case. This is a problem for those machine learning tools that require the output variable to be represented as a factor variable. RWeka’s MultiLayerPerceptron function is a case in point. Script 8.4 shows how the *as.factor* function is used to make sure the output variable for a created data frame is structured as a factor variable.

Page 83: First sentence of the third complete paragraph

duev should be *due*

Page 110: Second to last line

Table 5.2 should be *Table 5.4*

Page 131: SplitInfo(Income Range)

$+ 2/15 \log_2/15]$ should be $+ 2 \cdot 2/15 \log_2 2/15]$

Page 175: Middle of page

Let’s look at the second rule given as
should be

Let’s look at the third rule given as

Page 213:

my.nnet\$act.fct should be *my.nnet\$act.fct*

Page 262:

we use Equation 10.8 to determine ... should be *we use Equation 10.3 to determine*

Page 284: 3rd paragraph

Table 11.3 should be *Table 11.5*

Page 290:

As a final point, notice that all predictiveness and predictability scores are computed with respect to the parent node. For example, the predictiveness score of 1.0 for nuclei = two found in N_3 reflects the fact that all instances incorporated into N_2 with nuclei = two have followed the path from N_2 to N_3 .

should be

As a final point, notice that all predictiveness and predictability scores are computed with respect to the parent node. For example, the predictiveness score of 1.0 for nuclei = two found in N_5 reflects the fact that all instances incorporated into N_1 with nuclei = two have followed the path from N_1 to N_5 .

Page 332: Equation B.7

The equation is missing a minus sign between e_i and mae

Page 333:

Equation B.10 is missing a minus sign between e_i and mae

Page 333:

The reference to Table 9.2 in Appendix B should be replaced with this table:

Instance Number	Actual Output	Computed Output	Absolute Error	Squared Error
1	0.0	0.024	0.024	0.0005
2	1.0	0.998	0.002	0.0000
3	0.0	0.023	0.023	0.0005
4	1.0	0.986	0.014	0.0002
5	1.0	0.999	0.001	0.0000
6	0.0	0.050	0.050	0.0025
7	1.0	0.999	0.001	0.0000

8	0.0	0.262	0.262	0.0686
9	0.0	0.060	0.060	0.0036
10	1.0	0.997	0.003	0.0000
11	1.0	0.999	0.001	0.0000
12	1.0	0.776	0.224	0.0502
13	1.0	0.999	0.001	0.0000
14	0.0	0.023	0.023	0.0005
15	1.0	0.999	0.001	0.0000

You can download the most up to date versions of the supporting materials at:

<https://krypton.mnsu.edu/~sa7379bt/>

Please direct any questions or concerns you have about the text or any of the supplementary materials to me at: richard.roiger@mnsu.edu