——————————————————————————————————
————

Chapter 1, page 18

The sentence is:

"Here, the rows indicate examples of automobiles, while the columns record **various each** automobile's features, such as price, mileage, color, and transmission type."

The sentence should be:

"Here, the rows indicate examples of automobiles, while the columns record **each** automobile's features, such as price, mileage, color, and transmission type."

——————————————————————————————————
————

Chapter 4, page 116

The sentence is:

"**This type of figures** grew in popularity recently, since it provides a way to observe trending topics on social media websites."

The sentence should be:

"**This type of figure** grew in popularity recently, since it provides a way to observe trending topics on social media websites."

——————————————————————————————————
————

Chapter 4, page 121

The sentence is:

"Still, as shown in the following table, these steps **are is** fairly straightforward: … "

The sentence should be:

"Still, as shown in the following table, these steps **are** fairly straightforward: ..."

———————————————————————————————————

Chapter 5, page 130

The sentence is:

For example, a diagonal line could be created by a decision that asks, "is the number of celebrities **is greater** than the estimated budget?" If so, then "it will be a critical success."

The sentence should be:

For example, a diagonal line could be created by a decision that asks, "is the number of celebrities **greater** than the estimated budget?" If so, then "it will be a critical success."

———————————————————————————————————

Chapter 5, page 139

The sentence is:

"Omitting this seed will cause your training and testing **split** to differ from those shown in the remainder of this chapter:"

The sentence should be:

"Omitting this seed will cause your training and testing **splits** to differ from those shown in the remainder of this chapter:"

———————————————————————————————————

Chapter 5, page 156

The sentence is:

"The RIPPER algorithm can create much more complex rules than can the 1R algorithm, **as in** can consider more than one feature."

The sentence should be:

"The RIPPER algorithm can create much more complex rules than can the 1R algorithm, **as it** can consider more than one feature."

———————————————————————————————————————

Chapter 7, page 219

The two concatenated sentences are:

"The late science fiction author Arthur C. Clarke wrote, "any sufficiently advanced technology is indistinguishable from **magic."This** chapter covers a pair of machine learning methods that may appear at first glance to be magic."

The sentences should be (space included):

"The late science fiction author Arthur C. Clarke wrote, "any sufficiently advanced technology is indistinguishable from **magic."** **This** chapter covers a pair of machine learning methods that may appear at first glance to be magic."

———————————————————————————————————————

Chapter 7, page 220

The sentence is:

"With any luck, you'll realize that you don't need a black belt in statistics to tackle black box **machine's** learning methods — there's no

need to be intimidated!"

The sentence should be:

"With any luck, you'll realize that you don't need a black belt in statistics to tackle black box **machine** learning methods—there's no need to be intimidated!"

— —

Chapter 7, page 220

The sentence is:

"This question is the basis of the controversial **Turing test**, proposed in 1950 by the pioneering computer scientist Alan Turing, **proposed in 1950 by the pioneering computer scientist Alan Turing,** which grades a machine as intelligent if a human being cannot distinguish its behavior from a living creature's."

The sentence should be:

"This question is the basis of the controversial **Turing test**, proposed in 1950 by the pioneering computer scientist Alan Turing, which grades a machine as intelligent if a human being cannot distinguish its behavior from a living creature's."

— —

Chapter 7, page 220

The sentence is:

"This question is the basis of the controversial **Turing test**, proposed in 1950 by the pioneering computer scientist Alan Turing, **proposed in 1950 by the pioneering computer scientist Alan Turing,** which grades a machine as intelligent if a human being cannot distinguish its

behavior from a living creature's."

The sentence should be:

"This question is the basis of the controversial **Turing test**, proposed in 1950 by the pioneering computer scientist Alan Turing, which grades a machine as intelligent if a human being cannot distinguish its behavior from a living creature's."

– –

Chapter 7, page 239

The sentence is:

"A **Support Vector Machine** (**SVM**) can be imagined as a surface that creates
a boundary between points of data **plotted in multidimensional** that represent examples and their feature values."

The sentence should be:

"A **Support Vector Machine** (**SVM**) can be imagined as a surface that creates
a boundary between points of data **plotted in multidimensional space** that represent examples and their feature values."

– –

Chapter 7, page 245

The sentence is:

"Modifying this value will adjust the penalty, for example, **the fall** on the wrong side of the hyperplane."

The sentence should be:

"Modifying this value will adjust the penalty, for example, **for falling**

on the wrong side of the hyperplane."

_____
_____

Chapter 7, page 251

The sentence is:

"However, we can skip this step for now, because the R package that we will use for fitting the SVM model will perform the **rescaling automatically**."

The sentence should be (remove extra spacing):

"However, we can skip this step for now, because the R package that we will use for fitting the SVM model will perform the **rescaling automatically**."

_____
_____

Chapter 8, page 260

The sentence is:

"The result of a market basket analysis is a collection of **association rules** that specify patterns found in the relationships among **items he itemsets**."

The sentence should be:

"The result of a market basket analysis is a collection of **association rules** that specify patterns found in the relationships among **the itemsets**."

_____
_____

Chapter 8, page 268

The sentence is:

"**Although this may seem reasonable this, if we use the data in this form, we will encounter problems later may seem reasonable,** R chose to create four variables because the first line had exactly four comma-separated values."

The sentence should be:

"**Although this may seem reasonable, if we use the data in this form, we will encounter problems later.** R chose to create four variables because the first line had exactly four comma-separated values."

— —

Chapter 8, page 270

The sentence is:

"With an additional step, we can determine that the average transaction contained *43,367 / 8,835 = 4.409* distinct grocery items. Of course, if we look a little further down the output, we'll see that the mean number of items per transaction has already been provided."

The sentence should be:

"With an additional step, we can determine that the average transaction contained *43,367 / 9,835 = 4.409* distinct grocery items. Of course, if we look a little further down the output, we'll see that the mean number of items per transaction has already been provided."

— —

Chapter 8, page 282

The sentence is:

- "The operator `%in%` means that at least one of the items must be

found in the list you defined. If you want any rules matching either berries or yogurt, you could write items **%in%c(**"berries", "yogurt")."

The sentence should be (note the space):

- "The operator %in% means that at least one of the items must be found in the list you defined. If you want any rules matching either berries or yogurt, you could write items **%in% c(**"berries", "yogurt")."

————————————————————————————————————

Chapter 10, page 314

The sentence is:

"It's ideal to **fid** a learner that is extremely confident when making a correct prediction, but timid in the face of doubt."

The sentence should be:

"It's ideal to **fit** a learner that is extremely confident when making a correct prediction, but timid in the face of doubt."

————————————————————————————————————

Chapter 10, page 321

The sentence is:

"For more information on caret, please refer to: Kuhn M. Building predictive models in R using the caret package. *Journal of Statistical* **Software**. 2008; 28."

The sentence should be (note the italics):

"For more information on caret, please refer to: Kuhn M. Building predictive

models in R using the caret package. *Journal of Statistical Software*. 2008; 28."

_____

Chapter 10, page 331

The sentence is:

"It is possible to calculate F-scores using different weights for precision and recall, but choosing the weights could be tricky **at the best** and arbitrary at worst."

The sentence should be:

"It is possible to calculate F-scores using different weights for precision and recall, but choosing the weights could be tricky **at best** and arbitrary at worst."

_____

Chapter 10, page 331

The sentence is:

"The ROCR package provides an easy-to-use suite of functions **for visualizing for visualizing** the performance of classification models."

The sentence should be:

"The ROCR package provides an easy-to-use suite of functions **for visualizing** the performance of classification models."

_____

Chapter 10, page 331

The sentence is:

"It includes functions for computing **large set** of the most common

performance measures and visualizations."

The sentence should be:

"It includes functions for computing **large sets [or "a large set"]** of the most common performance measures and visualizations."

_ _

Chapter 10, page 332

The ROC curve graph is missing an x-axis label.

_ _

Chapter 10, page 334

The sentence is:

"The end result is **an** ROC plot with a dashed reference line:"

The sentence should be:

"The end result is **a** ROC plot with a dashed reference line:"

_ _

Chapter 10, page 337

The sentence is:

"To avoid this problem, it is better to divide the original data so that in addition to the training **datasets** and the test **datasets**, a **validation dataset** is available."

The sentence should be:

"To avoid this problem, it is better to divide the original data so that in addition to the training **dataset** and the test **dataset**, a **validation**

dataset is available."

——

Chapter 10, page 344

The sentence is:

"A special case of bootstrapping known as the **0.632 bootstrap** accounts for this by calculating the final performance measure as a function of performance on both the training data (which is overly optimistic) and the test data (which is overly pessimistic)."

The sentence should be (notice commas added for clarity):

"A special case of bootstrapping, known as the **0.632 bootstrap**, accounts for this by calculating the final performance measure as a function of performance on both the training data (which is overly optimistic) and the test data (which is overly pessimistic)."

——

Chapter 11, page 351

The sentence is:

"As this practice sometimes results in the selection of models that achieve marginal performance improvements via large increases in model complexity, alternative model selection functions are provided."

The sentence should include at least a brief mention of these "alternative model selection functions", perhaps as a follow-up sentence to not leave the reader hanging.

——

Chapter 12, page 380

The code and sentence is:

"""

The sqlQuery() function uses typical SQL queries, as shown in the following command:

> my_query <- "select * from my_table where my_value = 1"
> results_df <- sqlQuery(channel = my_db, query = <mark>sql_query</mark>,
    stringsAsFactors = FALSE)
The resulting results_df object is a data frame containing all of the rows selected using the SQL query stored in <mark>sql_query</mark>.

"""

The code and sentence should be:

"""

The sqlQuery() function uses typical SQL queries, as shown in the following command:

> my_query <- "select * from my_table where my_value = 1"
> results_df <- sqlQuery(channel = my_db, query = <mark>my_query</mark>,
    stringsAsFactors = FALSE)
The resulting results_df object is a data frame containing all of the rows selected using the SQL query stored in <mark>my_query</mark>.

"""

_ _

Chapter 12, page 391

The sentence is:

"Notice that the quote characters **have escaped** using the \" notation:"

The sentence should be:

"Notice that the quote characters **have been escaped** using the \"
notation:"

_____
_____

Chapter 12, page 398

The sentence is:

"**R** has a reputation for being slow and memory-inefficient, a
reputation that is at least somewhat earned."

The sentence should be:

"**Base R** has a reputation for being slow and memory-inefficient, a
reputation that is at least somewhat earned."

_____
_____

Chapter 12, page 411

In the section entitled "Parallel cloud computing with MapReduce
and Hadoop", there is no mention of sparklyr, Keras, or TensorFlow.
These need to be included in 2017.

_____
_____