



## Practice Problems 5

Using the **RWeka** package can cause trouble. We have collected some [work-arounds](#), [hints](#), and [installation instructions](#) that may help. The key is to use compatible versions of Java and R. Alternatively (and more simply) use either a Windows computer or run in on the [cloud version of RStudio](#) or the [cloud version of R](#) -- or skip that question.

### Problem 1 (20 Pts)

Build an R Notebook of the bank loan decision tree example in the textbook on pages 135 to 148; the CSV file is available for download below. Show each step and add appropriate documentation. Note that the provided dataset uses values 1 and 2 in default column whereas the book has *no* and *yes* in the default column. To fix any problems replace "no" with "1" and "yes" with "2" in the code that for *matrix\_dimensions*. Alternatively, change the line

```
error_cost <- matrix(c(0, 1, 4, 0), nrow = 2, dimnames = matrix_dimensions) to error_cost <- matrix(c(0, 1, 4, 0), nrow = 2).
```

If your tree produces poor results or runs slowly, add `control=Weka_control(R=TRUE)`.

### Problem 2 (10 Pts)

Build an R Notebook of the poisonous mushrooms example using rule learners in the textbook on pages 160 to 168. Show each step and add appropriate documentation. The CSV file is available below. If you have issues with the **RWeka** package on MacOS, consider using a Windows computer, [RStudio.cloud](#) or skip this question.

**Tip:** In case anyone gets this error on the 1R implementation:

```
>mushroom_1R <- OneR(type ~., data = mushrooms)
```

Error in .jcall(0, "Ljava/lang/Class;", "getClass") : weka.core.UnsupportedAttributeTypeException:  
weka.classifiers.rules.OneR: ...

Change your characters to factors. Here's an [explanation why factors are needed](#).

## Problem 3 (35 Pts)

So far we have explored four different approaches to classification: *k*NN, Naive Bayes, C5.0 Decision Trees, and RIPPER Rules. Comment on the differences of the algorithms and when each is generally used. Provide examples of when they work well and when they do not work well. Add your comments to your R Notebook. Be specific and explicit; however, no code examples are needed.

## Problem 4 (35 Pts)

Much of our focus so far has been on building a single model that is most accurate. In practice, data scientists often construct multiple models and then combine them into a single prediction model. This is referred to as a **model ensemble**. Two common techniques for assembling such models are *boosting* and *bagging*. Do some research and define what model ensembles are, why they are important, and how boosting and bagging function in the construction of assemble models. Be detailed and provide references to your research. You can use this [excerpt](#) from [Kelleher, MacNamee, and D'Arcy, Fundamentals of Machine Learning for Predictive Data Analytics](#) as a starting point. This book is an excellent resource for those who want to dig deeper into data mining and machine learning.

## Submission Details

- Practice Problems are for learning and practice and therefore are not graded and no submission is required. You are encourage to discuss and review them with your peers. Additionally, they are reviewed during weekly recitations. If you desire, you may ask for individual feedback from the instructional staff during office hours. Completing practice problems will prepare you for the graded practicums and their completion is critical to doing well on the practicums and the final project.
- If you have issues with the *RWeka* package due to Java incompatibilities, you may wish to use the online R environment [rdr.io](http://rdr.io) which offers an R console as well as a Jupyter Notebook installation or the [cloud version of RStudio](#). Also, look at the [collection of tips](#) that we have compiled for you.

## Useful Resources

- [Credit Data Set \(CSV\)](#)
- [Mushroom Data Set \(CSV\)](#)
- [RWeka Usage Hints](#)
- [Cloud version of RStudio](#)



## Learning

[Blackboard](#)

[Lynda.com](#)

[Data Camp](#)

## Support

[Contact Instructor](#)

[Virtual Office](#)

[Book Appointment](#)



© COPYRIGHT 2017-2020 by Northeastern University

Created by [Martin Schedlbauer, PhD](#)

FREE FOR ACADEMIC USE WITH ACKNOWLEDGEMENT AND NOTICE.