

Intermediate Tree 2



This is a continuation of the intermediate decision tree [exercise](#).

Answers to the exercises are available [here](#).

If you obtained a different (correct) answer than those listed on the solutions page, please feel free to post your answer as a comment on that page.

Exercise 1

use the `predict()` command to make predictions on the Train data. Set the method to "class". Class returns classifications instead of probability scores. Store this prediction in `pred_dec`.

Exercise 2

Print out the confusion matrix

Exercise 3

What is the accuracy of the model. Use the confusion matrix.

Exercise 4

What is the misclassification error rate? Refer to Basic_decision_tree exercise to get the formula.

Exercise 5

Lets say we want to find the baseline model to compare our prediction improvement. We create a base model using this code

```
length(Test$class)
```

```
base=rep(1,3183)
```

Use the `table()` command to create a confusion matrix between the base and `Test$class`.



Learn more about decision trees in the online courses

- [Regression Machine Learning with R](#) (it includes two lectures on definitions, characteristics, mathematical formulae, graphical representations, fitting, forecasting, and accuracy of decision trees)
- [Machine Learning A-Z™: Hands-On Python & R In Data Science](#) (including 3 lectures, ~45 mins, on decision tree regression in both R and Python)
- [Data Science and Machine Learning Bootcamp with R](#) (126 lectures and 17.5 hrs of video, including several lectures on decisions trees and random forests)

Exercise 6

What is the number difference between the confusion matrix accuracy of dec and base?

Exercise 7

Remember the `predict()` command in question 1. We will use the same mode and same command except we will set the method to "regression". This gives us a probability estimates. Store this in `pred_dec_reg`

Exercise 8

load the ROCR package.

Use the `prediction()`, `performance()` and `plot()` command to print the ROC curve. Use `pred_dec_reg` variable from Q7. You can also refer to the previous exercise to see the code.

Exercise 9

`plot()` the same ROC curve but set `colorize=TRUE`

Exercise 10

Comment on your findings using ROC curve and accuracy. Is it a good model? Did you notice that ROC prediction() command only takes probability predictions as one of its arguments. Why is that so?