TUTORIAL 9

In this tutorial we will cover:

1.   Importing data into R (revision)

2.   Fitting classifiers in R (revision)

3.   Evaluating classifiers in R

Mode of study:

You can work through this tutorial in your own time, or in Jan Mouton 2015 and 2017, where help will be available.

Solutions will be made available in Week 12.

Before you start

If you have not worked through Practical 9 yet, you should do so before starting this tutorial. It might also be useful to revisit Practicals 5 and 7 to revise the fitting of decision trees and kNN models in R.

DATASET

In Practical 5 we worked with the dataset `churn.csv`. The aim was to predict customer churn (i.e. customers switching from one cellular phone company to another). The dataset is used in the example on pages 73 – 78 of the textbook, and consists of 20 000 observations. The target variable is called `LEAVE` (indicating whether a customer stayed or left) and there are also measurements for 11 different attributes.

In this tutorial you will use the same dataset once again. A description of the attributes can be found in the instruction document for Practical 5.

PRACTICAL QUESTIONS

1.   **Load the `tree` and `caret` libraries.**

2.   **Import the dataset into RStudio and store it in an object called `churn`. View the structure of the dataset. Make sure the data types are correct; if not, fix it.**

3.1   **Fit a decision tree to the data, to predict the value of the target variable `LEAVE`. Use the `tree()` function with its default settings.**

3.2   **Plot the tree created in Question 3.1.**

3.3   **Construct the confusion matrix for the tree created in Question 3.1.**
      Use the `confusionMatrix()` function from the `caret` package. Remember that you will have to compute the predicted values first.

4.1   **Fit a one nearest neighbour classifier to the data to predict the value of the target variable. Use the `knn3()` function.**

4.2   **Construct the confusion matrix for the classifier from Question 4.1.**

5.1   **Fit a 5 nearest neighbour classifier to the data to predict the value of the target variable. Use the `knn3()` function.**

5.2   **Construct the confusion matrix for the classifier from Question 5.1.**

6.1   **Fit a 21 nearest neighbour classifier to the data to predict the value of the target variable. Use the `knn3()` function.**

6.2   **Construct the confusion matrix for the classifier from Question 6.1.**

DISCUSSION QUESTIONS

Use the output from the practical questions to answer the following questions. Solutions will be made available in due course.

1.   **Consider the accuracy for the one nearest neighbour classifier. What do you notice? Why is this so?**

2.   **Explain how the "no information rate" value in the confusion matrices is calculated and what it represents.**

3. **Ignoring the 1NN classifier, which of the models you fit do you think is the best one? Give a reason for your answer.**

> After completing this week's practical and tutorial you should be able to perform a basic evaluation of classifier models.