Machine learning 2

Mara Acuja

24 11 2021

```r
data <- read.csv("Data/default_features_1059_tracks.txt", header=FALSE)
data <- as.data.frame(data)
colnames(data)[69:70] <- c("Longitude", "Latitude")

# Maybe some more preprocessing could be done here.
anyNA(data) # testing if there is at least a single NA -> but in this dataset there isn't
```

```
## [1] FALSE
```

```r
anyDuplicated(data)
```

```
## [1] 0
```

```r
set.seed(1)
n <-dim(data)[1]
train <- sample(1:n, 0.8*n)
test <- (1:n)[-train]
```

Comparison of machine learning algorithms / Introduction / Theory

This report attempts to compare two regression/classification algorithms on its behaviour on a specific data set. What should be found out?

Data set

Here should be a bit of a short summary of the data set with some key characteristics of the data set.

What is it about? Which variables are included? What type of variables? What is missing? What about missing values?

Method

Short summary about the algorithms. Which are used? What do we do? Classification or Regression?

Algorithm 1

Short introduction of the first algorithm. What does it do? What are the strengths? What are weaknesses? How is it implemented, including major code snippets.

Algorithm 2

Short introduction of the second algorithm. What does it do? What are the strengths? What are weaknesses? How is it implemented, including major code snippets.

Results

Here some tables, summaries or especially graphs shoul be shown here. Maybe this section should be separated into two to show the algorithms for themselves

Discussion

Here follows the discussion of the results. What are the major findings? How did the algorithms perform? Which one was better overall? Is it always better or were the findings which were better by the other one? Which one should be implemented? How could the algorithm be tweeked to perform even better? Where were the problems during implementation? Where are the limits for the algorithms?

Conclusion

At final some conclusions about the key findings and which algorithm should be used. What was the goal? Were and how were they achieved?

RMarkdown default stuff - needs to be removed but serves up to now as draft

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.