Classificação de Veículos

Luan Roger Santos Santana

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

Preparing data

Loading Data

```
data_raw <- read.csv("../data_sets/Material 02 - 5 - C - Veiculos - Dados.csv")
data_raw_new_cases <- read.csv("../data_sets/Material 03 - 5 - C - Veiculos - Dados - Novos Casos.csv")
```

Cleaning data

```
data <- data_raw[,!(names(data_raw) %in% c('a'))]
data_new_cases <- data_raw_new_cases[,!(names(data_raw_new_cases) %in% c('a'))]
print(head(data))
```

```
##   Comp Circ DCirc RadRa PrAxisRa MaxLRa ScatRa Elong PrAxisRect MaxLRect
## 1   95   48    83   178       72     10    162    42         20       159
## 2   91   41    84   141       57      9    149    45         19       143
## 3  104   50   106   209       66     10    207    32         23       158
## 4   93   41    82   159       63      9    144    46         19       143
## 5   85   44    70   205      103     52    149    45         19       144
## 6  107   57   106   172       50      6    255    26         28       169
##   ScVarMaxis ScVarmaxis RaGyr SkewMaxis Skewmaxis Kurtmaxis KurtMaxis HollRa
## 1        176        379   184        70         6        16       187    197
## 2        170        330   158        72         9        14       189    199
## 3        223        635   220        73        14         9       188    196
## 4        160        309   127        63         6        10       199    207
## 5        241        325   188       127         9        11       180    183
## 6        280        957   264        85         5         9       181    183
##   tipo
## 1  van
## 2  van
## 3 saab
## 4  van
## 5  bus
## 6  bus
```

```
print(head(data_new_cases))
```

```
##   Comp Circ DCirc RadRa PrAxisRa MaxLRa ScatRa Elong PrAxisRect MaxLRect
## 1  100   48    83   178       72     10    162    42         20       159
## 2   91   40    84   141       57      9    149    45         18       143
## 3   92   50   106   209       66     10    207    32         23       160
```

```
##   ScVarMaxis ScVarmaxis RaGyr SkewMaxis Skewmaxis Kurtmaxis KurtMaxis HollRa
## 1        176        400   184        70         6        16       187    197
## 2        170        330    70        72         9        14       189    199
## 3        223        635   220        73        14         9       188    230
##   tipo
## 1    ?
## 2    ?
## 3    ?
```

Creating data partitioning

```
set.seed(1988)
ran <- sample(1:nrow(data), 0.8 * nrow(data))
training_data <- data[ran,]
test_data <- data[-ran,]
```

Training

Using KNN

Creating the model

```
tuneGrid <- expand.grid(k = c(1,3,5,7,9))
set.seed(1988)
knn <- train(tipo ~ ., data = training_data, method = "knn", tuneGrid=tuneGrid)
print(knn)
```

```
## k-Nearest Neighbors
##
## 676 samples
##  18 predictor
##   4 classes: 'bus', 'opel', 'saab', 'van'
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 676, 676, 676, 676, 676, 676, ...
## Resampling results across tuning parameters:
##
##   k  Accuracy   Kappa
##   1  0.6225458  0.4965691
##   3  0.5963528  0.4616769
##   5  0.6039509  0.4721682
##   7  0.5955723  0.4615824
##   9  0.5882459  0.4519974
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was k = 1.
```

```
prediction.knn <- predict(knn, test_data)
cf_matrix <- confusionMatrix(prediction.knn, as.factor(test_data$tipo))
print(cf_matrix)
```

Checking the model with training data

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction bus opel saab van
##       bus   39    2    4   2
##       opel   0   15   21   1
##       saab   6   13   17   1
##       van    3    1    2  43
##
## Overall Statistics
##
##                Accuracy : 0.6706
##                  95% CI : (0.5945, 0.7406)
##     No Information Rate : 0.2824
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.5586
##
##  Mcnemar's Test P-Value : 0.5677
##
## Statistics by Class:
##
##                      Class: bus Class: opel Class: saab Class: van
## Sensitivity              0.8125     0.48387      0.3864     0.9149
## Specificity              0.9344     0.84173      0.8413     0.9512
## Pos Pred Value           0.8298     0.40541      0.4595     0.8776
## Neg Pred Value           0.9268     0.87970      0.7970     0.9669
## Prevalence               0.2824     0.18235      0.2588     0.2765
## Detection Rate           0.2294     0.08824      0.1000     0.2529
## Detection Prevalence     0.2765     0.21765      0.2176     0.2882
## Balanced Accuracy        0.8735     0.66280      0.6138     0.9331
```

Checking for new cases

```
prediction.knn_new_data <- predict(knn, data_new_cases)
data_new_cases$tipo <- NULL
result <- cbind(data_new_cases, tipo=prediction.knn_new_data)
print(result)
```

```
##   Comp Circ DCirc RadRa PrAxisRa MaxLRa ScatRa Elong PrAxisRect MaxLRect
## 1  100   48    83   178       72     10    162    42         20      159
## 2   91   40    84   141       57      9    149    45         18      143
## 3   92   50   106   209       66     10    207    32         23      160
##   ScVarMaxis ScVarmaxis RaGyr SkewMaxis Skewmaxis Kurtmaxis KurtMaxis HollRa
## 1        176        400   184        70         6        16       187    197
## 2        170        330    70        72         9        14       189    199
## 3        223        635   220        73        14         9       188    230
##   tipo
## 1 opel
## 2 saab
## 3 saab
```