Regressão de Biomassa

Luan Roger Santos Santana

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

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

```
##
## Attaching package: 'caret'
```

```
## The following objects are masked from 'package:Metrics':
##
##     precision, recall
```

Preparing data

Loading Data

```
data_raw <- read.csv("../data_sets/Material 02 - 4 - R - Biomassa - Dados.csv")
data_raw_new_cases <- read.csv("../data_sets/Material 02 - 4 - R - Biomassa - Dados - Novos Casos.csv")
```

Cleaning data

```
data <- data_raw
data_new_cases <- data_raw_new_cases
print(head(data))
```

```
##    dap   h   Me biomassa
## 1  6.4 5.0 1.04     7.07
## 2  7.3 5.0 1.04    10.30
## 3  7.8 5.5 1.04    13.90
## 4  9.2 7.6 1.04    18.61
## 5  9.9 8.2 1.04    30.97
## 6 10.6 8.7 1.04    46.44
```

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

```
##    dap    h   Me biomassa
## 1  6.4  7.0 1.04        ?
## 2  7.3 10.0 1.04        ?
## 3  7.8  5.5 1.04        ?
## 4 12.2  7.5 1.04        ?
```

Creating data partitioning

```
set.seed(1988)
# ran <- sample(1:nrow(data), 0.8 * nrow(data))
ind <- createDataPartition(data$biomassa, p=0.80, list = FALSE)
```

```
training_data <- data[ind,]
test_data <- data[-ind,]
```

Training

Using KNN

Creating the model

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

```
## k-Nearest Neighbors
##
## 240 samples
##   3 predictor
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 240, 240, 240, 240, 240, 240, ...
## Resampling results across tuning parameters:
##
##   k  RMSE      Rsquared   MAE
##   1  334.1897  0.8923980  102.4732
##   3  423.9360  0.8612042  114.1809
##   5  431.4029  0.8478399  112.8244
##   7  442.5294  0.8501821  112.6002
##   9  455.5224  0.8361783  113.0790
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was k = 1.
```

```
prediction.knn <- predict(knn, test_data)
library(Metrics)
rmse(test_data$biomassa, prediction.knn)
```

Checking the model with training data

```
## [1] 1352.071
```

```
r2 <- function(predito, observado) {
    return(1 - (sum((predito-observado)^2) / sum((predito-mean(observado))^2)))
}
r2(prediction.knn,test_data$biomassa)
```

R^2 function

```
## [1] 0.3551759
```

Checking for new cases

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

```
##    dap    h   Me biomassa
## 1  6.4  7.0 1.04    12.79
## 2  7.3 10.0 1.04     7.84
## 3  7.8  5.5 1.04    13.90
## 4 12.2  7.5 1.04    46.75
```