TRABALHO DE IAA003 – Linguagem R

EQUIPE: 23

RODRIGO ROLIM VERAS

ALBINO BOGUCHESKI JUNIOR

MARCOS ANTONIO NESPOLO JUNIOR

```r
# instalar o pacotes necessários
install.packages("mlbench", repos = "http://cran.us.r-project.org")
```

1.1 Carregue a base de dados Satellite

```
##
## The downloaded binary packages are in
##  /var/folders/7c/43lddvnj7bsfph_4fmdrnsl40000gp/T//RtmpeUIvvw/downloaded_packages
```

```r
install.packages("e1017", repos = "http://cran.us.r-project.org")
```

```
## Warning: package 'e1017' is not available for this version of R
##
## A version of this package for your version of R might be available elsewhere,
## see the ideas at
## https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages
```

```r
install.packages("randomForest", repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
##  /var/folders/7c/43lddvnj7bsfph_4fmdrnsl40000gp/T//RtmpeUIvvw/downloaded_packages
```

```r
install.packages("kernlab", repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
##  /var/folders/7c/43lddvnj7bsfph_4fmdrnsl40000gp/T//RtmpeUIvvw/downloaded_packages
```

```r
install.packages("caret", repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
##  /var/folders/7c/43lddvnj7bsfph_4fmdrnsl40000gp/T//RtmpeUIvvw/downloaded_packages
```

```r
# usar os pacotes necessários
library(mlbench)
library(caret)
```

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

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

```r
# carregar os dados Satellite
data(Satellite)
```

```r
# exibir estrutura dos dados Satellite
str(Satellite)
```

```
## 'data.frame':    6435 obs. of  37 variables:
##  $ x.1    : num  92 84 84 80 84 80 76 76 76 76 ...
##  $ x.2    : num  115 102 102 102 94 94 102 102 89 94 ...
##  $ x.3    : num  120 106 102 102 102 98 106 106 98 98 ...
##  $ x.4    : num  94 79 83 79 79 76 83 87 76 76 ...
##  $ x.5    : num  84 84 80 84 80 80 76 80 76 76 ...
##  $ x.6    : num  102 102 102 94 94 102 102 98 94 98 ...
##  $ x.7    : num  106 102 102 102 98 102 106 106 98 102 ...
##  $ x.8    : num  79 83 79 79 76 79 87 79 76 72 ...
##  $ x.9    : num  84 80 84 80 80 76 80 76 76 76 ...
##  $ x.10   : num  102 102 94 94 102 102 98 94 98 94 ...
##  $ x.11   : num  102 102 102 98 102 102 106 102 102 90 ...
##  $ x.12   : num  83 79 79 76 79 79 79 76 72 76 ...
##  $ x.13   : num  101 92 84 84 84 76 80 80 80 76 ...
##  $ x.14   : num  126 112 103 99 99 99 107 112 95 91 ...
##  $ x.15   : num  133 118 104 104 104 104 118 118 104 104 ...
##  $ x.16   : num  103 85 81 78 81 81 88 88 74 74 ...
##  $ x.17   : num  92 84 84 84 76 76 80 80 76 76 ...
##  $ x.18   : num  112 103 99 99 99 99 112 107 91 95 ...
##  $ x.19   : num  118 104 104 104 104 108 118 113 104 100 ...
##  $ x.20   : num  85 81 78 81 81 85 88 85 74 78 ...
##  $ x.21   : num  84 84 84 76 76 76 80 80 76 76 ...
##  $ x.22   : num  103 99 99 99 99 103 107 95 95 91 ...
##  $ x.23   : num  104 104 104 104 108 118 113 100 100 100 ...
##  $ x.24   : num  81 78 81 81 85 88 85 78 78 74 ...
##  $ x.25   : num  102 88 84 84 84 84 79 79 75 75 ...
##  $ x.26   : num  126 121 107 99 99 103 107 103 91 91 ...
##  $ x.27   : num  134 128 113 104 104 104 113 104 96 96 ...
##  $ x.28   : num  104 100 87 79 79 79 87 83 75 71 ...
##  $ x.29   : num  88 84 84 84 84 79 79 79 75 79 ...
##  $ x.30   : num  121 107 99 99 103 107 103 103 91 87 ...
##  $ x.31   : num  128 113 104 104 104 109 104 104 96 93 ...
##  $ x.32   : num  100 87 79 79 79 87 83 79 71 71 ...
##  $ x.33   : num  84 84 84 84 79 79 79 79 79 79 ...
##  $ x.34   : num  107 99 99 103 107 107 103 95 87 87 ...
##  $ x.35   : num  113 104 104 104 109 109 104 100 93 93 ...
##  $ x.36   : num  87 79 79 79 87 87 79 79 71 67 ...
##  $ classes: Factor w/ 6 levels "red soil","cotton crop",..: 3 3 3 3 3 3 3 3 4 4 ...
```

```r
# apresentar alguma medidas estatísticas do dados Satellite
summary(Satellite)
```

```
##       x.1             x.2              x.3              x.4        
##  Min.   : 39.0   Min.   : 27.00   Min.   : 53.00   Min.   : 33.00  
##  1st Qu.: 60.0   1st Qu.: 71.00   1st Qu.: 85.00   1st Qu.: 69.00  
##  Median : 68.0   Median : 87.00   Median :101.00   Median : 81.00  
##  Mean   : 69.4   Mean   : 83.59   Mean   : 99.29   Mean   : 82.59  
##  3rd Qu.: 80.0   3rd Qu.:103.00   3rd Qu.:113.00   3rd Qu.: 92.00  
##  Max.   :104.0   Max.   :137.00   Max.   :140.00   Max.   :154.00  
##       x.5              x.6              x.7              x.8       
##  Min.   : 39.00   Min.   : 27.00   Min.   : 50.00   Min.   : 29.0  
##  1st Qu.: 60.00   1st Qu.: 71.00   1st Qu.: 85.00   1st Qu.: 69.0  
##  Median : 68.00   Median : 85.00   Median :101.00   Median : 81.0  
##  Mean   : 69.15   Mean   : 83.24   Mean   : 99.11   Mean   : 82.5  
##  3rd Qu.: 80.00   3rd Qu.:103.00   3rd Qu.:113.00   3rd Qu.: 92.0  
```

```
##  Max.   :104.00   Max.   :137.00   Max.   :145.00   Max.   :157.0
##       x.9             x.10             x.11             x.12
##  Min.   : 40.00   Min.   : 27.00   Min.   : 50.00   Min.   : 29.00
##  1st Qu.: 60.00   1st Qu.: 71.00   1st Qu.: 85.00   1st Qu.: 68.00
##  Median : 67.00   Median : 85.00   Median :100.00   Median : 81.00
##  Mean   : 68.91   Mean   : 82.89   Mean   : 98.85   Mean   : 82.39
##  3rd Qu.: 79.00   3rd Qu.:102.00   3rd Qu.:113.00   3rd Qu.: 92.00
##  Max.   :104.00   Max.   :130.00   Max.   :145.00   Max.   :157.00
##       x.13            x.14             x.15             x.16
##  Min.   : 39.00   Min.   : 27.00   Min.   : 50.00   Min.   : 29.00
##  1st Qu.: 60.00   1st Qu.: 71.00   1st Qu.: 85.00   1st Qu.: 69.00
##  Median : 68.00   Median : 85.00   Median :101.00   Median : 81.00
##  Mean   : 69.29   Mean   : 83.48   Mean   : 99.31   Mean   : 82.64
##  3rd Qu.: 80.00   3rd Qu.:103.00   3rd Qu.:113.00   3rd Qu.: 92.00
##  Max.   :104.00   Max.   :137.00   Max.   :145.00   Max.   :154.00
##       x.17            x.18             x.19             x.20
##  Min.   : 40.00   Min.   : 27.00   Min.   : 50.00   Min.   : 29.0
##  1st Qu.: 60.00   1st Qu.: 71.00   1st Qu.: 85.00   1st Qu.: 69.0
##  Median : 68.00   Median : 85.00   Median :100.00   Median : 81.0
##  Mean   : 69.05   Mean   : 83.17   Mean   : 99.15   Mean   : 82.6
##  3rd Qu.: 79.00   3rd Qu.:103.00   3rd Qu.:113.00   3rd Qu.: 92.0
##  Max.   :104.00   Max.   :130.00   Max.   :145.00   Max.   :157.0
##       x.21            x.22             x.23             x.24
##  Min.   : 39.00   Min.   : 27.00   Min.   : 50.00   Min.   : 29.00
##  1st Qu.: 60.00   1st Qu.: 71.00   1st Qu.: 85.00   1st Qu.: 68.00
##  Median : 67.00   Median : 84.00   Median :100.00   Median : 81.00
##  Mean   : 68.84   Mean   : 82.86   Mean   : 98.95   Mean   : 82.47
##  3rd Qu.: 79.00   3rd Qu.:103.00   3rd Qu.:113.00   3rd Qu.: 92.00
##  Max.   :104.00   Max.   :130.00   Max.   :145.00   Max.   :157.00
##       x.25            x.26             x.27             x.28
##  Min.   : 39.00   Min.   : 27.00   Min.   : 50.00   Min.   : 29.00
##  1st Qu.: 60.00   1st Qu.: 71.00   1st Qu.: 85.00   1st Qu.: 69.00
##  Median : 68.00   Median : 85.00   Median :100.00   Median : 81.00
##  Mean   : 69.16   Mean   : 83.37   Mean   : 99.21   Mean   : 82.66
##  3rd Qu.: 79.00   3rd Qu.:103.00   3rd Qu.:113.00   3rd Qu.: 92.00
##  Max.   :104.00   Max.   :131.00   Max.   :140.00   Max.   :154.00
##       x.29            x.30             x.31             x.32
##  Min.   : 39.00   Min.   : 27.00   Min.   : 50.00   Min.   : 29.00
##  1st Qu.: 60.00   1st Qu.: 71.00   1st Qu.: 85.00   1st Qu.: 69.00
##  Median : 68.00   Median : 85.00   Median :100.00   Median : 81.00
##  Mean   : 68.94   Mean   : 83.15   Mean   : 99.11   Mean   : 82.62
##  3rd Qu.: 79.00   3rd Qu.:103.00   3rd Qu.:113.00   3rd Qu.: 92.00
##  Max.   :104.00   Max.   :130.00   Max.   :145.00   Max.   :157.00
##       x.33            x.34             x.35             x.36
##  Min.   : 39.00   Min.   : 27.00   Min.   : 50.00   Min.   : 29.00
##  1st Qu.: 60.00   1st Qu.: 71.00   1st Qu.: 85.00   1st Qu.: 68.00
##  Median : 67.00   Median : 84.00   Median :100.00   Median : 81.00
##  Mean   : 68.73   Mean   : 82.86   Mean   : 98.93   Mean   : 82.51
##  3rd Qu.: 79.00   3rd Qu.:103.00   3rd Qu.:113.00   3rd Qu.: 92.00
##  Max.   :104.00   Max.   :130.00   Max.   :145.00   Max.   :157.00
##                 classes
##  red soil           :1533
##  cotton crop        : 703
##  grey soil          :1358
```

```
##  damp grey soil     : 626
##  vegetation stubble : 707
##  very damp grey soil:1508
```

```r
# exibir alguns dados do Satellite
head(Satellite, n = 6)
```

```
##   x.1 x.2 x.3 x.4 x.5 x.6 x.7 x.8 x.9 x.10 x.11 x.12 x.13 x.14 x.15 x.16 x.17
## 1  92 115 120  94  84 102 106  79  84  102  102   83  101  126  133  103   92
## 2  84 102 106  79  84 102 102  83  80  102  102   79   92  112  118   85   84
## 3  84 102 102  83  80 102 102  79  84   94  102   79   84  103  104   81   84
## 4  80 102 102  79  84  94 102  79  80   94   98   76   84   99  104   78   84
## 5  84  94 102  79  80  94  98  76  80  102  102   79   84   99  104   81   76
## 6  80  94  98  76  80 102 102  79  76  102  102   79   76   99  104   81   76
##   x.18 x.19 x.20 x.21 x.22 x.23 x.24 x.25 x.26 x.27 x.28 x.29 x.30 x.31 x.32
## 1  112  118   85   84  103  104   81  102  126  134  104   88  121  128  100
## 2  103  104   81   84   99  104   78   88  121  128  100   84  107  113   87
## 3   99  104   78   84   99  104   81   84  107  113   87   84   99  104   79
## 4   99  104   81   76   99  104   81   84   99  104   79   84   99  104   79
## 5   99  104   81   76   99  108   85   84   99  104   79   84  103  104   79
## 6   99  108   85   76  103  118   88   84  103  104   79   79  107  109   87
##   x.33 x.34 x.35 x.36   classes
## 1   84  107  113   87 grey soil
## 2   84   99  104   79 grey soil
## 3   84   99  104   79 grey soil
## 4   84  103  104   79 grey soil
## 5   79  107  109   87 grey soil
## 6   79  107  109   87 grey soil
```

```r
# Para reproductibilidade
set.seed(7)
```

```r
# particionar em 80% para treino e 20% para teste
indices <- createDataPartition(Satellite$classes, p=0.8, list=F)
treino <- Satellite[indices, ]
teste <- Satellite[-indices, ]
```

1.2 Crie partições contendo 80% para treino e 20% para teste

```r
# treinar modelos RandomForest, SVM e RNA
rf <- train(classes ~ ., data=treino, method="rf")
svm <- train(classes ~ ., data=treino, method="svmRadial")
rna <- train(classes ~ ., data=treino, method="nnet", trace=F)
```

1.3 Treine modelos RandomForest, SVM e RNA para predição destes dados.

```r
# predições
predict.rf <- predict(rf, teste)
predict.svm <- predict(svm, teste)
predict.rna <- predict(rna, teste)
```

```r
# matrizes de confusões de cada uma das predições

# matriz de confusão para o modelo RF
conf_matrix.rf <- confusionMatrix(predict.rf, teste$classes)

print(conf_matrix.rf)
```

1.4 Escolha o melhor modelo com base em suas matrizes de confusão.

```
## Confusion Matrix and Statistics
##
##                       Reference
## Prediction            red soil cotton crop grey soil damp grey soil
##   red soil                 300           1         3              1
##   cotton crop                0         137         1              0
##   grey soil                  3           0       263             25
##   damp grey soil             0           0         2             78
##   vegetation stubble         3           0         1              0
##   very damp grey soil        0           2         1             21
##                       Reference
## Prediction            vegetation stubble very damp grey soil
##   red soil                             4                   0
##   cotton crop                          1                   1
##   grey soil                            0                   3
##   damp grey soil                       1                  20
##   vegetation stubble                 128                   4
##   very damp grey soil                  7                 273
##
## Overall Statistics
##
##                Accuracy : 0.9182
##                  95% CI : (0.9019, 0.9326)
##     No Information Rate : 0.2383
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.8987
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                       Class: red soil Class: cotton crop Class: grey soil
## Sensitivity                    0.9804             0.9786           0.9705
## Specificity                    0.9908             0.9974           0.9694
## Pos Pred Value                 0.9709             0.9786           0.8946
## Neg Pred Value                 0.9938             0.9974           0.9919
## Prevalence                     0.2383             0.1090           0.2111
## Detection Rate                 0.2336             0.1067           0.2048
## Detection Prevalence           0.2407             0.1090           0.2290
## Balanced Accuracy              0.9856             0.9880           0.9699
##                       Class: damp grey soil Class: vegetation stubble
## Sensitivity                         0.62400                   0.90780
## Specificity                         0.98016                   0.99300
```

```
## Pos Pred Value                      0.77228                  0.94118
## Neg Pred Value                      0.96027                  0.98868
## Prevalence                          0.09735                  0.10981
## Detection Rate                      0.06075                  0.09969
## Detection Prevalence                0.07866                  0.10592
## Balanced Accuracy                   0.80208                  0.95040
##                      Class: very damp grey soil
## Sensitivity                              0.9070
## Specificity                              0.9685
## Pos Pred Value                           0.8980
## Neg Pred Value                           0.9714
## Prevalence                               0.2344
## Detection Rate                           0.2126
## Detection Prevalence                     0.2368
## Balanced Accuracy                        0.9377
```

```r
cat('\n')
```

```r
# matriz de confusão para o modelo SVM
conf_matrix.svm <- confusionMatrix(predict.svm, teste$classes)

print(conf_matrix.svm)
```

```
## Confusion Matrix and Statistics
##
##                      Reference
## Prediction             red soil cotton crop grey soil damp grey soil
##   red soil                  303           0         2              0
##   cotton crop                 0         138         2              2
##   grey soil                   2           0       261             27
##   damp grey soil              0           1         5             74
##   vegetation stubble          1           0         0              1
##   very damp grey soil         0           1         1             21
##                      Reference
## Prediction             vegetation stubble very damp grey soil
##   red soil                              5                   0
##   cotton crop                           2                   2
##   grey soil                             0                   7
##   damp grey soil                        1                  21
##   vegetation stubble                  126                   3
##   very damp grey soil                   7                 268
##
## Overall Statistics
##
##                Accuracy : 0.9112
##                  95% CI : (0.8943, 0.9262)
##     No Information Rate : 0.2383
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.8901
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
```

```
##                      Class: red soil Class: cotton crop Class: grey soil
## Sensitivity                   0.9902             0.9857           0.9631
## Specificity                   0.9928             0.9930           0.9645
## Pos Pred Value                0.9774             0.9452           0.8788
## Neg Pred Value                0.9969             0.9982           0.9899
## Prevalence                    0.2383             0.1090           0.2111
## Detection Rate                0.2360             0.1075           0.2033
## Detection Prevalence          0.2414             0.1137           0.2313
## Balanced Accuracy             0.9915             0.9894           0.9638
##                      Class: damp grey soil Class: vegetation stubble
## Sensitivity                        0.59200                   0.89362
## Specificity                        0.97584                   0.99563
## Pos Pred Value                     0.72549                   0.96183
## Neg Pred Value                     0.95685                   0.98699
## Prevalence                         0.09735                   0.10981
## Detection Rate                     0.05763                   0.09813
## Detection Prevalence               0.07944                   0.10202
## Balanced Accuracy                  0.78392                   0.94462
##                      Class: very damp grey soil
## Sensitivity                              0.8904
## Specificity                              0.9695
## Pos Pred Value                           0.8993
## Neg Pred Value                           0.9665
## Prevalence                               0.2344
## Detection Rate                           0.2087
## Detection Prevalence                     0.2321
## Balanced Accuracy                        0.9299
```

```r
cat('\n')
```

```r
# matriz de confusão para o modelo RNA
conf_matrix.rna <- confusionMatrix(predict.rna, teste$classes)

print(conf_matrix.rna)
```

```
## Confusion Matrix and Statistics
##
##                      Reference
## Prediction             red soil cotton crop grey soil damp grey soil
##   red soil                  289         132         3              1
##   cotton crop                 6           0         0              0
##   grey soil                   3           5       244            104
##   damp grey soil              0           0         0              0
##   vegetation stubble          7           3        12             13
##   very damp grey soil         1           0        12              7
##                      Reference
## Prediction             vegetation stubble very damp grey soil
##   red soil                             31                   1
##   cotton crop                           1                   0
##   grey soil                            13                 267
##   damp grey soil                        0                   0
##   vegetation stubble                   91                  15
##   very damp grey soil                   5                  18
##
## Overall Statistics
```

```
##
##                Accuracy : 0.5
##                  95% CI : (0.4723, 0.5277)
##     No Information Rate : 0.2383
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.3672
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: red soil Class: cotton crop Class: grey soil
## Sensitivity                   0.9444           0.000000           0.9004
## Specificity                   0.8282           0.993881           0.6130
## Pos Pred Value                0.6324           0.000000           0.3836
## Neg Pred Value                0.9794           0.890368           0.9583
## Prevalence                    0.2383           0.109034           0.2111
## Detection Rate                0.2251           0.000000           0.1900
## Detection Prevalence          0.3559           0.005452           0.4953
## Balanced Accuracy             0.8863           0.496941           0.7567
##                      Class: damp grey soil Class: vegetation stubble
## Sensitivity                        0.00000                   0.64539
## Specificity                        1.00000                   0.95626
## Pos Pred Value                         NaN                   0.64539
## Neg Pred Value                     0.90265                   0.95626
## Prevalence                         0.09735                   0.10981
## Detection Rate                     0.00000                   0.07087
## Detection Prevalence               0.00000                   0.10981
## Balanced Accuracy                  0.50000                   0.80082
##                      Class: very damp grey soil
## Sensitivity                             0.05980
## Specificity                             0.97457
## Pos Pred Value                          0.41860
## Neg Pred Value                          0.77196
## Prevalence                              0.23442
## Detection Rate                          0.01402
## Detection Prevalence                    0.03349
## Balanced Accuracy                       0.51718
```

```
cat('\n')
```

1.5 Indique qual modelo dá o melhor o resultado e a métrica utilizada O melhor modelo foi random forest com acurácia de 0.9182 e kappa de 0.8987. A métrica utilizada foram a acurácia e kappa.

```
dados <- read.csv("http://www.razer.net.br/datasets/Volumes.csv", sep=";", dec=",")
head(dados)
```

2.1 Carregar o arquivo Volumes.csv (http://www.razer.net.br/datasets/Volumes.csv)

```
##   NR  DAP    HT   HP       VOL
## 1  1 34.0 27.00 1.80 0.8971441
## 2  2 41.5 27.95 2.75 1.6204441
## 3  3 29.6 26.35 1.15 0.8008181
```

```
## 4  4 34.3 27.15 1.95 1.0791682
## 5  5 34.5 26.20 1.00 0.9801112
## 6  6 29.9 27.10 1.90 0.9067022
```

```r
dados$NR <- NULL
```

2.2 Eliminar a coluna NR, que só apresenta um número sequencial

```r
regression.indices <- caret::createDataPartition(dados$VOL, p=0.8, list=F)
regression.treino <- dados[regression.indices, ]
regression.teste <- dados[-regression.indices, ]
```

2.3 Criar partição de dados: treinamento 80%, teste 20%

```r
# Para reproductibilidade
set.seed(7)
```

```r
regression.rf <- caret::train(VOL ~ ., data=regression.treino, method="rf")
```

2.4 Usando o pacote "caret", treinar os modelos: Random Forest (rf), SVM (svmRadial), Redes Neurais (neuralnet) e o modelo alométrico de SPURR

```
## note: only 2 unique complexity parameters in default grid. Truncating the grid to 2 .
```

```r
regression.svm <- caret::train(VOL ~ ., data=regression.treino, method="svmRadial")
regression.rna <- caret::train(
  VOL ~ ., data=regression.treino,
  method="nnet",
  trControl=trainControl(method = "LOOCV"),
  trace=F
)
```

```r
regression.spurr <- nls(
  VOL ~ b0 + b1*DAP*DAP*HT,
  data=regression.treino,
  start=list(b0=0.5, b1=0.5)
)
```

treino do modelo Spurr

```r
summary(regression.spurr)
```

visualizar os resultados de Spurr

```
##
## Formula: VOL ~ b0 + b1 * DAP * DAP * HT
##
## Parameters:
##     Estimate Std. Error t value Pr(>|t|)
## b0 2.490e-02  5.328e-02   0.467    0.642
## b1 3.860e-05  1.466e-06  26.325   <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1377 on 78 degrees of freedom
##
## Number of iterations to convergence: 2
## Achieved convergence tolerance: 1.134e-09
```

2.5 Efetue as predições nos dados de teste predições

```r
# Para reproductibilidade
set.seed(7)
```

```r
predict.regression.rf <- predict(regression.rf, regression.teste)
predict.regression.svm <- predict(regression.svm, regression.teste)
predict.regression.rna <- predict(regression.rna, regression.teste)
predict.regression.suprr <- predict(regression.spurr, regression.teste)
```

2.6 Crie suas próprias funções (UDF) e calcule as seguintes métricas entre a predição e os dados observados

- Erro padrão de estimativa: Syx

```r
Syx <- function(reals, predicteds, n) {
  return (sqrt(sum((reals - predicteds)^2)/(n - 2)))
}
```

- Erro padrão de estimativa em porcentagem: Syx%

```r
SyxPercent <- function(reals, predicteds, n) {
  return ((Syx(reals, predicteds, n)/mean(reals))*100)
}
```

- Coeficientededeterminação:R2

```r
R2 <- function (reals, predicteds) {
  return (1 - sum((reals - predicteds)^2)/sum((reals - mean(reals))^2))
}
```

2.7 Escolha o melhor modelo.

métrica de estimativas para o modelo RandomForest - Regressão

- coeficiente de determinação

```r
R2(regression.teste$VOL, predict.regression.rf)
```

```
## [1] 0.8223603
```

- Erro padrão estimativa

```r
n <- nrow(regression.teste)
Syx(regression.teste$VOL, predict.regression.rf, n)
```

```
## [1] 0.1376052
```

- Erro padrão estimativa em porcentagem

```r
n <- nrow(regression.teste)
SyxPercent(regression.teste$VOL, predict.regression.rf, n)
```

```
## [1] 10.42195
```

métrica de estimativas para o modelo SVM - Regressão

- coeficiente de determinação

```r
R2(regression.teste$VOL, predict.regression.svm)
```

```
## [1] 0.6254546
```

- Erro padrão estimativa

```r
n <- nrow(regression.teste)
Syx(regression.teste$VOL, predict.regression.svm, n)
```

```
## [1] 0.19981
```

- Erro padrão estimativa em porcentagem

```r
n <- nrow(regression.teste)
SyxPercent(regression.teste$VOL, predict.regression.svm, n)
```

```
## [1] 15.13322
```

métricas de estimativas para o modelo nnet - Regressão

- coeficiente de determinação

```r
R2(regression.teste$VOL, predict.regression.rna)
```

```
## [1] -1.069672
```

- Erro padrão estimativa

```r
n <- nrow(regression.teste)
Syx(regression.teste$VOL, predict.regression.rna, n)
```

```
## [1] 0.4696948
```

- Erro padrão estimativa em porcentagem

```r
n <- nrow(regression.teste)
SyxPercent(regression.teste$VOL, predict.regression.rna, n)
```

```
## [1] 35.57377
```

métricas de estimativas para o modelo Spurr

- coeficiente de determinação

```r
R2(regression.teste$VOL, predict.regression.suprr)
```

```
## [1] 0.7734018
```

- Erro padrão estimativa

```r
n <- nrow(regression.teste)
Syx(regression.teste$VOL, predict.regression.suprr, n)
```

```
## [1] 0.1554151
```

- Erro padrão estimativa em porcentagem

```r
n <- nrow(regression.teste)
SyxPercent(regression.teste$VOL, predict.regression.suprr, n)
```

```
## [1] 11.77084
```

2.7 escolha o melhor modelo

Resumo dos resultados RF:

1. coeficiente de determinação: 0.8223603.
2. Erro padrão estimativa: 0.1376052.
3. Erro padrão estimativa em porcentagem: 10.42195

SVM:

1. coeficiente de determinação: 0.6254546
2. Erro padrão estimativa: 0.19981
3. Erro padrão estimativa em porcentagem: 15.13322

nnet:

1. coeficiente de determinação: -1.069672
2. Erro padrão estimativa: 0.4696948
3. Error padrão estimativa em porcentagem: 35.57377

spurr:

1. coeficiente de determinação: 0.7734018
2. Erro padrão estimativa: 0.1554151
3. Error padrão estimativa em porcentagem: 11.77084

Com base nas métricas, o modelo que se saiu melhor foi o RandomForest, com R2 igual 0.8223603, Erro padrão estimativa de 0.1376052 e Erro padrão de estima- tiva em porcentagem de 10.42195