Evaluation Time Series Models

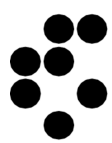
Solutions to Hands on Exercises

L. Torgo

ltorgo@fc.up.pt

Faculdade de Ciências / LIAAD-INESC TEC, LA
Universidade do Porto

Jan, 2017



Jožef Stefan Institute

Hands on Performance Estimation

the Algae data set

Load in the data set `algae` from package **DMwR** and answer the following questions:

1. Estimate the MSE of a regression tree for forecasting alga *a1* using 10-fold Cross validation. solution
2. Repeat the previous exercise this time trying some variants of random forests. Check what are the characteristics of the best performing variant. solution
3. Compare the results in terms of mean absolute error of the default variants of a regression tree, a linear regression model and a random forest, in the task of predicting alga a3. Use 2 repetitions of a 5-fold Cross Validation experiment. solution

Jožef Stefan Institute

Solutions to Exercise 1

- Estimate the MSE of a regression tree for forecasting alga *a1* using 10-fold Cross validation.

```
library(DMwR)
library(performanceEstimation)
data(algae)
algae <- algae[-c(62,199),]
res.a1 <- performanceEstimation(
    PredTask(a1 ~ .,algae[,1:12],"algaA1"),
    Workflow("standardWF",learner="rpartXse",pre="knnImp"),
    EstimationTask("mse",method=CV())
    )
```

Solutions to Exercise 1 (cont.)

- Estimate the MSE of a regression tree for forecasting alga *a1* using 10-fold Cross validation.

```
summary(res.a1)


##
## == Summary of a  Cross Validation Performance Estimation Experiment ==
##
## Task for estimating  mse  using
##  1 x 10 - Fold Cross Validation
##   Run with seed =  1234
##
## * Predictive Tasks ::  algaA1
## * Workflows  ::  rpartXse
##
## -> Task:  algaA1
##   *Workflow: rpartXse
##                mse
## avg     316.64207
## std     185.39856
## med     306.41274
## iqr     280.34431
## min      79.40367
## max     658.52265
## invalid   0.00000
```

Solutions to Exercise 2

- Repeat the previous exercise this time trying some variants of random forests. Check what are the characteristics of the best performing variant.

```
library(randomForest)
resrf.a1 <- performanceEstimation(
    PredTask(a1 ~ .,algae[,1:12],"algaA1"),
    workflowVariants("standardWF",
                     learner="randomForest",
                     learner.pars=list(ntree=c(500,750,1000)),
                     pre="knnImp"),
    EstimationTask("mse",method=CV())
    )
```

Jožef Stefan Institute

Solutions to Exercise 2 (cont.)

```
summary(resrf.a1)


##
## == Summary of a  Cross Validation Performance Estimation Experiment ==
##
## Task for estimating  mse  using
##  1 x 10 - Fold Cross Validation
##   Run with seed =  1234
##
## * Predictive Tasks ::  algaA1
## * Workflows  ::  randomForest.v1, randomForest.v2, randomForest.v3
##
## -> Task:  algaA1
##   *Workflow: randomForest.v1
##                mse
## avg     259.25690
## std     166.21130
## med     211.48892
## iqr     160.13135
## min      68.04905
## max     643.98622
## invalid   0.00000
##
##   *Workflow: randomForest.v2
##                mse
## avg     261.37276
## std     166.08527
## med     215.40643
## iqr     166.84359
## min      68.65619
```

Solutions to Exercise 2 (cont.)

- Repeat the previous exercise this time trying some variants of random forests. Check what are the characteristics of the best performing variant.

```
topPerformer(resrf.a1,"mse","algaA1")

## Workflow Object:
##  Workflow ID       ::  randomForest.v1
##  Workflow Function ::  standardWF
##       Parameter values:
##   learner.pars  -> ntree=500
##   learner  -> randomForest
##   pre  -> knnImp
```

Go back

Jožef Stefan Institute

Solutions to Exercise 3

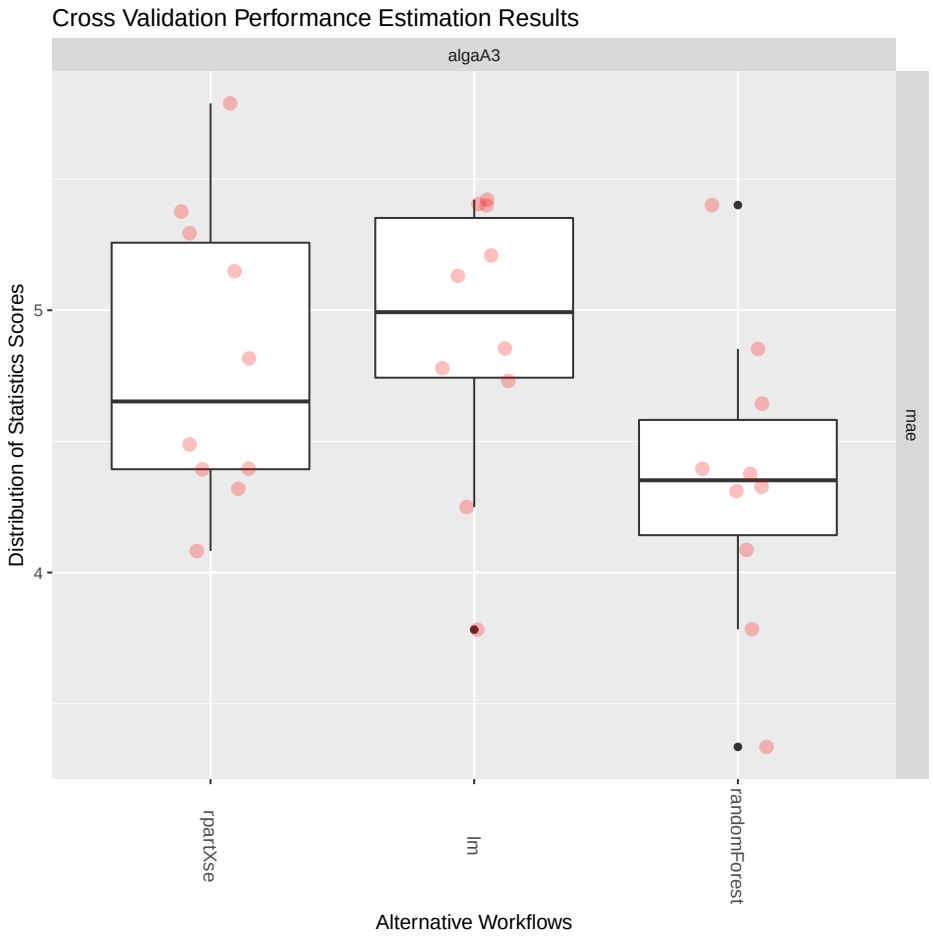
- Compare the results in terms of mean absolute error of the default variants of a regression tree, a linear regression model and a random forest, in the task of predicting alga a3. Use 2 repetitions of a 5-fold Cross Validation experiment. Plot the results

```
res.a3 <- performanceEstimation(
    PredTask(a3 ~ ., algae[,c(1:11,14)], "algaA3"),
    workflowVariants("standardWF",
                     learner=c("rpartXse","lm","randomForest"),
                     pre="knnImp"),
    EstimationTask("mae",method=CV(nReps=2,nFolds=5))
    )
```

Jožef Stefan Institute

Solutions to Exercise 3 (cont.)

```
plot(res.a3)
```



Jožef Stefan Institute

Go back