Predictive Analytics

Solutions to Hands On Exercises

L. Torgo

ltorgo@dal.ca

Faculty of Computer Science / Institute for Big Data Analytics
Dalhousie University

May, 2021

ACADIA
UNIVERSITY

Hands on Linear Regression and Random Forests

the Servo data set

Load in the data set `Servo` from package **mlbench** and answer the following questions:

1. How would you obtain a random forest with 750 trees to forecast the value of *Class* (it is a numeric variable) solution
2. Repeat the previous exercise but now using a linear regression model. solution
3. Obtain the predictions of the two previous models for the data used to obtain them. Draw a scatterplot comparing these predictions solution
4. Split the data in train and test sets (80%-20%). Obtain the two previous models on the training data and get their predictions for the test set. Compare the predictions of the models. solution

Solutions to Exercise 1

- How would you obtain a random forest with 750 trees to forecast the value of *Class* (it is a numeric variable)

```
library(tidymodels)
data(Servo, package="mlbench")
s <- as_tibble(Servo)

rfSpec <-
  rand_forest(trees = 750) %>%          # the type of model
  set_engine("ranger") %>%   # the implementation to use
  set_mode("regression")     # type of task

rf <-
  rfSpec %>% fit(Class ~ ., data = s)
```

Go back

Solutions to Exercise 2

- Repeat the previous exercise but now using a linear regression model.

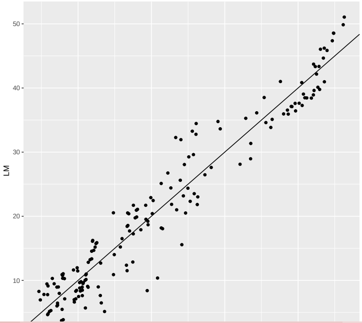
```
lmSpec <-
  linear_reg() %>%      # the type of model
  set_engine("lm")      # the implementation to use

lm <- lmSpec %>% fit(Class ~ ., data = s)   # fit the model to the data
```

Go back

Solutions to Exercise 3

- Obtain the predictions of the two previous models for the data used to obtain them. Draw a scatterplot comparing these predictions

```
psrf <- predict(rf,s)
pslm <- predict(lm,s)
preds <- psrf %>% rename(RF = .pred) %>%
  bind_cols(pslm) %>% rename(LM = .pred)
library(ggplot2)
ggplot(preds, aes(x=RF,y=LM)) + geom_point() + geom_abline(slope=1, intercept = 0)
```



Solutions to Exercise 4

- Split the data in train and test sets (80%-20%). Obtain the two previous models on the training data and get their predictions for the test set. Compare the predictions of the models.

```
spl <- initial_split(s, prop = 0.8)
sTR <- training(spl)
sTS <- testing(spl)
rf <- rfSpec %>% fit(Class ~ ., sTR)
lm <- lmSpec %>% fit(Class ~ ., sTR)
psrf <- predict(rf,sTS)
pslm <- predict(lm,sTS)
preds <- psrf %>% rename(RF = .pred) %>% bind_cols(pslm) %>% rename(LM = .pred)
sTS %>% bind_cols(preds) %>% metrics(Class,RF)
```

```
## # A tibble: 3 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>          <dbl>
## 1 rmse    standard       3.77
## 2 rsq     standard       0.950
## 3 mae     standard       2.69
```

```
sTS %>% bind_cols(preds) %>% metrics(Class,LM)
```

```
## # A tibble: 3 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>          <dbl>
## 1 rmse    standard       5.41
## 2 rsq     standard       0.849
```