MA8701 Advanced methods in statistical inference and learning

L19 with Kjersti Aas

Mette Langaas IMF/NTNU

20 March, 2023

Contents

LIME

R-code

Fitting random forest with the ranger package to the bike data.

```
# download manually
#"https://github.com/christophM/interpretable-ml-book/blob/master/data/bike.Rdata"
load("bike.Rdata")
colnames(bike)
```

```
##  [1] "season"          "yr"              "mnth"            "holiday"
##  [5] "weekday"         "workingday"      "weathersit"      "temp"
##  [9] "hum"             "windspeed"       "cnt"             "days_since_2011"
```

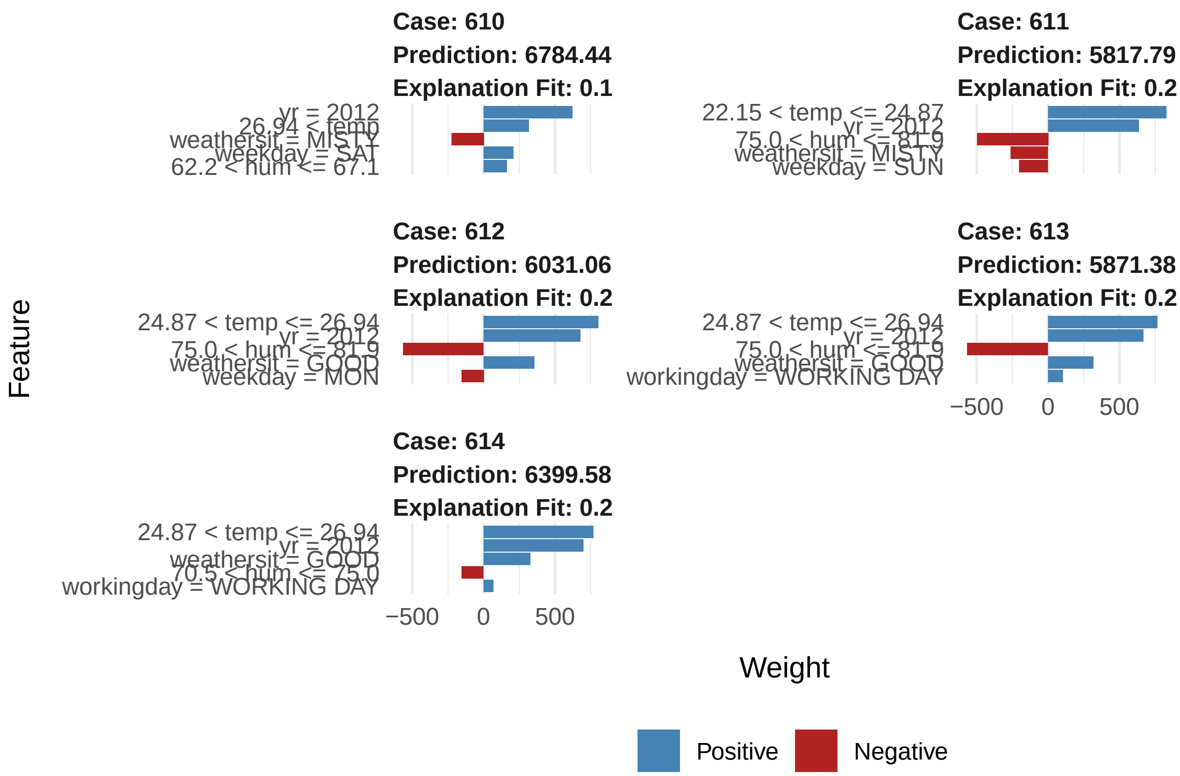
```
n=dim(bike)[1]
bikeTrain=bike[1:600,]
bikeTest<-bike[601:n,]

library(lime)
library(ranger)
predict_model.ranger <- function(x,newdata,type)
{
pred.rf <- predict(x, data = newdata)
switch(
type,
raw = data.frame(Response = res$class, stringsAsFactors = FALSE),
prob = as.data.frame(pred.rf$predictions[,2])
)
}

model_type.ranger <- function(x, ...)
```

```
{
'regression'
}

model<- ranger(cnt ~ ., data = bikeTrain, num.trees = 50, num.threads = 6,
verbose = TRUE,
probability = FALSE,
importance = "impurity",
mtry = sqrt(27))

print(model)
```

```
## Ranger result
##
## Call:
##  ranger(cnt ~ ., data = bikeTrain, num.trees = 50, num.threads = 6,      verbose = TRUE, probability
##
## Type:                             Regression
## Number of trees:                  50
## Sample size:                      600
## Number of independent variables:  11
## Mtry:                             5
## Target node size:                 5
## Variable importance mode:         impurity
## Splitrule:                        variance
## OOB prediction error (MSE):       343093.2
## R squared (OOB):                  0.8967285
```

Using `lime` to explain the random forest for test observationos 10:14 using `n_features=5` and `kernel_width=3`.

```
explainer <- lime::lime(
bikeTrain,
model = model,
#bin_continuous = FALSE
bin_continuous = TRUE,
n_bins = 10,
quantile_bins=TRUE
)
explanationLime <- explain(
bikeTest[10:14,-11],
explainer = explainer,
#n_labels = 1,
n_features = 5,
n_permutations = 5000,
feature_select = "auto",
kernel_width = 3)
lime::plot_features(explanationLime,
ncol = 2)
```



References for further reading

- The original LIME article Ribeiro, Singh, and Guestrin (2016)
- this blog post by the authors https://www.oreilly.com/content/introduction-to-local-interpretable-model-agnostic-explanations-lime/

Counterfactuals

Supplemental reading is Dandletal2020 and Wachter, Mittelstadt, and Russell (2018).

Software in R at https://github.com/susanne-207/moc.

References

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. 2016. ""Why Should i Trust You?": Explaining the Predictions of Any Classifier." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–44. KDD '16. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2939672.2939778.

Wachter, Sandra, Brent Mittelstadt, and Chris Russell. 2018. "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR." *Harvard Journal of Law & Technology* 31 (2). http://dx.doi.org/10.2139/ssrn.3063289.