

xgboost Link Lifetime Prediction

```
library(mlr3)
library(mlr3learners)
library(mlr3pipelines)

library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.3.2      v purrr 0.3.4
## v tibble 3.0.4       v dplyr 1.0.2
## v tidyr 1.1.2        v stringr 1.4.0
## v readr 1.4.0        v forcats 0.5.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()      masks stats::lag()
```

Reading the Data

```
data_dir = "../datasets/"

dataset_context = read_csv(
  str_c(data_dir, "dataset_context.csv"),
  col_types = cols(scenario=col_factor(), provider=col_factor())
) %>% rowid_to_column(var="row_id_original")

dataset_context_prediction = dataset_context %>% select(
  row_id_original,
  drive_id,
  provider,
  scenario,
  velocity_mps,
  acceleration_mpss,
  rsrp_dbm,
  rsrq_db,
  rssnr_db,
  cqi,
  ta,
  rsrp_neighbor,
  rsrq_neighbor,
  link_lifetime
)

# remove missing values
dataset_context_prediction = drop_na(dataset_context_prediction)
glimpse(dataset_context_prediction)
```

```
## Rows: 68,334
## Columns: 14
## $ row_id_original    <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15...
## $ drive_id           <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
## $ provider           <fct> o2, o2, tmobile, o2, tmobile, o2, tmobile, o2, tm...
## $ scenario           <fct> campus, campus, campus, campus, campus, campus, c...
## $ velocity_mps       <dbl> 6.76, 7.65, 3.35, 8.57, 3.81, 10.08, 9.01, 10.73,...
## $ acceleration_mpss <dbl> 0.00, 0.89, 0.00, 0.92, 0.45, 1.51, 5.15, 0.65, 1...
## $ rsrp_dbm           <dbl> -98, -101, -91, -101, -91, -94, -88, -94, -88, -9...
## $ rsrq_db            <dbl> -10, -12, -6, -12, -6, -9, -6, -9, -6, -8, -6, -8...
## $ rssnr_db           <dbl> -1, -1, 12, -1, 12, 5, 18, 5, 18, 1, 18, 1, 20, 3...
## $ cqi                <dbl> 9, 6, 11, 6, 11, 12, 15, 12, 15, 10, 15, 10, 12, ...
## $ ta                 <dbl> 9, 9, 7, 9, 7, 9, 7, 9, 7, 9, 7, 9, 7, 24, 9, 7, ...
## $ rsrp_neighbor      <dbl> -99, -104, -Inf, -104, -Inf, -100, -Inf, -100, -I...
## $ rsrq_neighbor      <dbl> -12, -14, -Inf, -14, -Inf, -17, -Inf, -17, -Inf, ...
## $ link_lifetime      <dbl> 18.01, 17.00, 19.01, 16.00, 18.00, 15.00, 16.99, ...
```

Train and Validation Split

```
dataset_context_train = dataset_context_prediction %>%
  filter(drive_id %in% 1:8) %>%
  select(!drive_id)

dataset_context_validation = dataset_context_prediction %>%
  filter(drive_id %in% 9:10) %>%
  select(!drive_id)

task_train = TaskRegr$new(
  id = "link_lifetime_train",
  backend = dataset_context_train,
  target = "link_lifetime"
)
task_train$col_roles$name = c("row_id_original")
task_train$col_roles$feature = setdiff(
  task_train$col_roles$feature,
  "row_id_original"
)
task_train

## <TaskRegr:link_lifetime_train> (54733 x 12)
## * Target: link_lifetime
## * Properties: -
## * Features (11):
##   - dbl (9): acceleration_mpss, cqi, rsrp_dbm, rsrp_neighbor, rsrq_db,
##     rsrq_neighbor, rssnr_db, ta, velocity_mps
##   - fct (2): provider, scenario

task_validation = TaskRegr$new(
  id = "link_lifetime_test",
  backend = dataset_context_validation,
  target = "link_lifetime"
)
task_validation$col_roles$name = c("row_id_original")
task_validation$col_roles$feature = setdiff(
```

```

    task_validation$col_roles$feature,
    "row_id_original"
  )
task_validation

## <TaskRegr:link_lifetime_test> (13601 x 12)
## * Target: link_lifetime
## * Properties: -
## * Features (11):
##   - dbl (9): acceleration_mpss, cqi, rsrp_dbm, rsrp_neighbor, rsrq_db,
##     rsrq_neighbor, rssnr_db, ta, velocity_mps
##   - fct (2): provider, scenario

```

Pipeline Creation

```

make_pipeline = function() {
  factor_encoding = po(
    "encode",
    method = "one-hot",
    affect_columns = selector_type("factor")
  )
  xgboost = lrn("regr.xgboost")
  pipe = factor_encoding %>% xgboost
  return(pipe)
}

make_default_learner = function() {
  learner_default = GraphLearner$new(
    make_pipeline()
  )
  learner_default$param_set$values = mlr3misc::insert_named(
    learner_default$param_set$values,
    list(regr.xgboost.nrounds=100)
  )
  return(learner_default)
}
default_learner = make_default_learner()

```

Training the Learner

```

default_learner$train(task_train)

## [21:07:59] WARNING: amalgamation/./src/objective/regression_obj.cu:174: reg:linear is now deprecated

```

Results on Validation Data

```

prediction = default_learner$predict(task_validation)

prediction$score(msr("regr.rsq"))

## regr.rsq
## 0.5564764

```

```
prediction$score(msr("regr.mae"))
```

```
## regr.mae
```

```
## 25.80789
```

```
prediction$score(msr("regr.rmse"))
```

```
## regr.rmse
```

```
## 38.64939
```

```
prediction_results = as.data.table(prediction)
```

```
ggplot(prediction_results, aes(x=truth, y=response)) +  
  geom_point()
```

