

Counterfactuals on all three models: Adjusted Causal, Unadjusted causal and Associational

Brian K. Masinde

```
# clear the working space
rm(list = ls())

library(here)
library(stats) # need this to calculate Mahalanobis Distance
library(parallel) # parallelize
library(dplyr)
library(FNN)
library(cluster)
library(ggplot2)
library(rpart)
library(caret)
```

Counterfactual Data Input

```
# we need the renaming function for cleaning
melor_2015 <- read.csv(here("data", "clustered_M15_CF_data.csv"))
```

Adjusted Causal Counterfactual predictions

Importing trained models

```
# Import trained BASE models
# From folder: adjusted SCM/new base models

adj_base_models_list <- list()

# base models file path
adj_base_file_path <- here("adjusted SCM/new base models")
adj_base_wind_model <- readRDS(file.path(adj_base_file_path,
                                         "dec_base_wind_model_tuned.rds"))

adj_base_class_full_model <- readRDS(file.path(adj_base_file_path,
                                              "damage_fit_class_full.rds"))
adj_base_reg_model <- readRDS(file.path(adj_base_file_path,
```

```

        "base_reg_model.rds"))

adj_base_models_list <- list("base_wind_model" = adj_base_wind_model,
                             "base_class_full_model" = adj_base_class_full_model,
                             "base_reg_model" = adj_base_reg_model)

# Import trained Truncated models
# From folder: adjusted SCM/new trunc models

# empty list
adj_trunc_models_list <- list()

adj_trunc_file_path <- here("adjusted SCM/new trunc models")

adj_trunc_wind_model <- readRDS(file.path(adj_trunc_file_path,
                                           "trunc_wind_model_tuned.rds"))

adj_trunc_reg_model <- readRDS(file.path(adj_trunc_file_path,
                                           "trunc_reg_model.rds"))

adj_trunc_models_list <- list("trunc_wind_model" = adj_trunc_wind_model,
                              "trunc_reg_model" = adj_trunc_reg_model)

```

Counterfactual Predictions

```

# calling hurdle function
source(here("R", "adj_hurdle_function.R"))

adj_counterfactual_hurdle_preds <- adj_hurdle_function(df = melor_2015,
                                                       scm_models_base = adj_base_models_list,
                                                       scm_models_high = adj_trunc_models_list,
                                                       threshold = 0.3 # threshold in train/test models is 0.35
                                                       )

```

Unadjusted Causal Counterfactual predictions

Importing trained models

```

# Import trained BASE models
# From folder: adjusted SCM/new base models

unadj_base_models_list <- list()

# base models file path
unadj_base_file_path <- here("unadjusted SCM/new base models")

```

```

unadj_base_wind_model <- readRDS(file.path(unadj_base_file_path,
                                           "dec_base_wind_model_tuned.rds"))

unadj_base_class_full_model <- readRDS(file.path(unadj_base_file_path,
                                                  "damage_fit_class_full.rds"))
unadj_base_reg_model <- readRDS(file.path(unadj_base_file_path,
                                           "base_reg_model.rds"))

unadj_base_models_list <- list("base_wind_model" = unadj_base_wind_model,
                              "base_class_full_model" = unadj_base_class_full_model,
                              "base_reg_model" = unadj_base_reg_model)

```

```

# Import trained Truncated models
# From folder: adjusted SCM/new trunc models

```

```

# empty list
unadj_trunc_models_list <- list()

unadj_trunc_file_path <- here("unadjusted SCM/new trunc models")

unadj_trunc_wind_model <- readRDS(file.path(unadj_trunc_file_path,
                                           "trunc_wind_model_tuned.rds"))

unadj_trunc_reg_model <- readRDS(file.path(unadj_trunc_file_path,
                                           "trunc_damage_fit_reg.rds"))

unadj_trunc_models_list <- list("trunc_wind_model" = unadj_trunc_wind_model,
                              "trunc_reg_model" = unadj_trunc_reg_model)

```

```

names(unadj_trunc_models_list)

```

```

## [1] "trunc_wind_model" "trunc_reg_model"

```

```

names(unadj_base_models_list)

```

```

## [1] "base_wind_model"      "base_class_full_model" "base_reg_model"

```

```

# setting threshold for classification step
threshold = 0.30

```

```

source(here("R", "unadj_hurdle_function.R"))
unadj_counterfactual_hurdle_preds <- unadj_hurdle_function(df = melor_2015,
                                                         scm_models_base = unadj_base_models_list,
                                                         scm_models_high = unadj_trunc_models_list,
                                                         threshold = threshold # threshold in train/test models i
                                                         )

```

Associational Model Counterfactuals

Importing trained models

```
# Read the .rds models
base_reg <- readRDS(here("associational XGBOOST", "damage_fit_reg_base.rds"))
trunc_reg <- readRDS(here("associational XGBOOST", "trunc_damage_fit_reg.rds"))
clas_model <- readRDS(here("associational XGBOOST", "ass_XGBOOST_class.rds"))
```

Counterfactual predictions

```
source(here("R", "ass_hurdle_function.R"))

# setting threshold for classification step
threshold = 0.30

ass_counterfactual_hurlde_preds <- ass_hurdle_function(df = melor_2015, ass_clas_model = clas_model,
  ass_base_model = base_reg, ass_trunc_model = trunc_reg ,threshold = threshold)
```

Output

```
# output dataframe with counterfactual predictions

melor_2015_CF_results <- melor_2015 %>%
  mutate(adj_cf_preds = adj_counterfactual_hurdle_preds,
         unadj_cf_preds = unadj_counterfactual_hurdle_preds,
         ass_cf_preds = ass_counterfactual_hurlde_preds)

# write to file
write.csv(melor_2015_CF_results, here("hurdle comparisons", "melor_2015_CF_results.csv"))
```