XAI: Interpretable models 2

Exercise 4.- Interpretable models: Logistic regression, trees and RuleFit.

0.- Backgrounds and expected results.

This exercise must be implemented from scratch. Read it carefully and write a report answering the questions using the graphs resulting from the exercise. Include the graphs in the report. **In this exercise you must upload the code file and the report (in word or pdf format) to Poliformat**.

Be creative and **don't limit your answers to a single sentence**. Write it thinking that a company has hired your services to analyse or explain the correct functioning of the software in question.

1.- Logistic regression.

The linear regression model works well in regression setups, but fails in the classification case. A solution for classification is logistic regression.

EXERCISE:

From the COMPAS database extract the samples of black defendants (race == "African-American") and white defendants (race == "Caucasian") as you did in "Fairness2". Remember to filter out rows where days_b_screening_arrest is over 30 or under -30, leaving us with less than 6,200 samples.

Use the logistic regression model (**glm()**) to **predict recedivism** of defendants considering that we set recividism prediction as those whose **decile_score is equal o greater than 4**. Use sex, age, race and is_recid as input features. Be careful with the family parameter.

Show the resulting weights, odds ratio and std error in a table.

Finally, to facilitate the comparison of the **odds ratio**, build a **bar chart** with the values for each input feature (discard Intercept in this plot and order bars by odds_ratio).

QUESTIONS:

Interpret the results.

2.- Tree.

EXERCISE:

Use the bike rental database of practice 3 (modified as we performed there) to **build a regresion tree** using the function **ctree()** from **library partykit** to **predict** the volume of bikes rented (**cnt**) depending on the other features. To allow an easy explanation of the results fix the **maximum depth to 2.**

Plot the resulting tree using the **plot()** function.

QUESTIONS:
Interpret the results.

3.- RuleFit.

EXERCISE:
Since RuleFit estimates a linear model in the end, the interpretation is equivalent to linear models. The only difference is that the model has new features that are coming from decision rules. Decision rules are binary features: A value of 1 means that all conditions of the rule are met, otherwise the value is 0.

Apply RuleFit (function **pre()** from **library pre**) to the bike rental database of practice 3 (modified as we performed there). **Read** the paper **pre_rule_fit_function.pdf** if you do not know how **to use the pre()** function. To allow an easy explanation of the results, **fix** the **maximum depth to 2** and fix the **family** parameter to **"gaussian".**

Obtain the coefficients of the obtained rules and the importance of each rule. **Remove** all the **rules** that have a **coefficient of 0.**

Show a table with the **4 top rules** ordered by importance.

Use **Plotly to draw** the **bar plot** that shows the **variable importance**. The variable importance can be obtained from the output of the **importance()** function (show **bars in descending order of imp**).

QUESTIONS:

- How many rules have you obtained initially?
- How many rules have you obtained after removing the ones with 0 coefficient?
- Interpret the results of the top 4 rules and the variable importance plot.