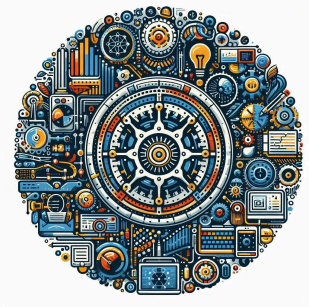
Advanced Topics on Machine Learning (M.IA009) 2024/2025



Practical Assignment 2 (18/11/2024)
Rita P. Ribeiro

Objectives

The goal of this project is to explore the insights provided by different explainable artificial intelligence (XAI) techniques in a supervised learning context on tabular data.

Tasks

You have a set of main tasks to accomplish, as described next.

Task 1: Dataset and Learning Task
Define your case study and train your first machine learning models.

- Choose a tabular data set from the options listed below and identify the associated classification or regression task your project will focus on.
  - DS1. Real Estate Listings in Portugal
  - DS2. Women's NBA Shots
  - DS3. Airline Passenger Satisfaction

  If you have other suggestions for the dataset, you should contact me.
- Use pre-modelling XAI techniques to assess feature importance analysis before training the models.

Task 2: In-Modeling Explanations
Train a glass-box model on your case study and critically analyze the interpretability provided by the obtained model.

Task 3: Post-Modeling Explanations
Train a black-box model on your case study and then apply post-hoc XAI techniques, as described below.

Task 3.1. Apply at least one simplification-based technique and assess how well it approximates the black-box model.

Task 3.2. Apply at least two feature-based techniques and compare the insights obtained from the explanations generated by the two methods, whether they are consistent or not.

Task 3.2 Apply at least one example-based technique and discuss the explanation obtained for two different examples.

Task 4: Quality of the explanations
Select at least one functionally-grounded metric to evaluate explanations generated by one of the post-hoc XAI techniques. Discuss the results and suggest ways this metric could inform interpretability improvements.

Deliverables

For this project you should deliver the following elements:

- Report.
- Notebook: All code documented. Description of the datasets used.
- Video Presentation.
- Slides of the presentation and checkpoints. This should be submitted as a pdf file.

Report

This should be submitted as a pdf file. Document with no more than 6 pages and letter of size 11pt. Avoid output dumps. Recall that the report is going to be evaluated by your very busy professors and that they will focus on the main points of the assignment. Always highlight the results that you consider relevant. Please note:

– Each task should be clearly identified in the document structure and all the figures and tables be self-explained.

– The conclusions should be a short high-level account of what was observed. Mainly, you should compare the insights provided by each method, highlighting strengths and weaknesses for interpreting the model and discussing the relative usefulness of each XAI approach for a simple model and dataset.

– It is not necessary to describe the technical details of the methods (unless necessary, but you should know their concepts and how they work). It is more important to point out the differences in the XAI methods and the reasons for the results in terms of XAI methods characteristics.

Notebook

A fully operational Jupyter notebook with the selected experiments as clear and concise as possible. Described how the data was processed. Try to structure the document in different parts highlighting all the steps of the process. It is recommended a creation of a github webpage with all the documentation that supports the project, including additional scripts that were developed.

Presentation

A **4 minutes video** (or link to a video), **per element of the group** with a recorded presentation of the respective part of the work. The presentations of the group, when combined, describe the whole of the group's work.

Slides

The project slides presentation and the slides at checkpoints to be submitted at different dates.

Evaluation

Components:

Report 30%

- Narrative 10%
- Writing style 10%
- Presentation 10% (includes the checkpoints and the final presentation).

Technical 70%

- Diversity of the employed techniques 20%
- Structure of the experimental setup 10%
- Correctness 30%
- Critical analysis of results 10%

You are free to propose new ideas and functionalities to implement. However, this should be discussed with the responsible of the course during the classes and checkpoints. New contributions will be evaluated in the item "Diversity of the results for the experiments".

Groups

Assignments are submitted by **groups of 3 students**. Different elements may have different grades. Other group sizes will not be considered.

It is advisable that the students from the same group perform overlapping work and only after that, exchange ideas with each other. Group work is important for learning from other people.

The case studies should be distributed uniformly according to the number of groups, i.e. 5 groups will choose DS1, 5 groups DS2 and 5 groups DS3.

Submissions

Formal final deadline is the **15th of December 2024**, to be submitted in Moodle, and only in Moodle. Submissions after that date will be multiplied by a monotonously decreasing factor that starts in 1.

Checkpoints:
Until the end of the semester, there will be two checkpoints, according to the schedule below:

- Checkpoint 1 (25th of November 2024): definition of dataset and supervised learning algorithms; pre-modelling analysis;

- Checkpoint 2 (9th of December 2024): choice of XAI techniques.

Each group should present an update with the status of the project. You will have around 5 minutes for this presentation, where you can show your current results and list your main difficulties.

After each checkpoint and for the project deadline you have to submit the slides on Moodle. Only one submission per group is required, however you should indicate if any element of the group did not participate or contributed differently from the other elements.

Ethical principles

When submitting, students commit themselves to follow strong ethical principles. All the work must be done by the elements of the group alone. All members of the group will be involved with the whole of the work. All the materials used and consulted must be credited in the work.