

Explainable AI – Exercise Sheet 4

November 25th, 2024

Exercise 1 - Theoretical Exercise

2p 1. Shapley Value Calculation:

Imagine a group of three university students from the University of Marburg, who are collaborating on an assignment for the XAI (Explainable Artificial Intelligence) course, and they want to ensure that the credit for their collaborative efforts is fairly distributed. Each student's contribution to the assignment is associated with the assignment grade, which is on a scale of 10 (10 is the maximum grade). Their assignment values (grades) are as follows:

- If no student contributes, the assignment's grade is 0 ($f(\emptyset) = 0$).
- If Student A works alone, the assignment's grade is 2 ($f(\{A\}) = 2$).
- If Student B or Student C works alone, the assignment's grade remains at 0 ($f(\{B\}) = f(\{C\}) = 0$).
- If Student A and Student B collaborate, the assignment's grade is 7.5 ($f(\{A, B\}) = 7.5$).
- If Student A and Student C collaborate, the assignment's grade is 8 ($f(\{A, C\}) = 8$).
- If Student B and Student C collaborate, the assignment's grade is 9 ($f(\{B, C\}) = 9$).
- If all three students collaborate, the assignment's grade is 10 ($f(\{A, B, C\}) = 10$).

Calculate the Shapley value for each student in this collaboration group to fairly distribute the assignment, considering their respective contributions.

1p 2. Given a model trained on a dataset with a large number of features, what is the problem when calculating the Shapley value for each feature? What solution has been proposed in the SHAP method to deal with this problem? Please describe it.

1p 3. Name and describe 3 desirable properties in SHAP in your own words

Exercise 2 - Practical Exercise

6p 4. In this exercise, we will practice implementing Shapley value calculation and using SHAP library. We will continue using the "xai" environment created in the first week. If you missed the first week, please revisit the Environment Setup Instructions from the first-week exercise.

To run the notebook for this week, follow the steps (similar to the previous weeks):

- A. Activate the "xai" environment by executing the command "conda activate xai" (without the double quotes).
- B. Navigate to the "practical-exercise" folder.
- C. Launch Jupyter Lab by executing the command "jupyter-lab" (without the double quotes).
- D. Open the Week_4_exercise.ipynb file and complete the exercises.

Note:

- A. Please carefully read the instructions and requirements for each practical exercise.
- B. Your results, interpretation, and comments on the results of the practical exercises are more important for evaluating the exercises than your code.
- C. If you discover any mistakes or encounter errors in the exercise, or if you have any questions, please feel free to contact us on Slack.

- D. External resources (papers, blogs, websites, implementations, etc.) need to be cited if you used them.**
- E. Please follow the Generative AI policy.**