Explainable AI – Exercise Sheet 5

December 2nd, 2024

Exercise 1 - Theoretical Exercise

1. Gradient-based methods We have learned about four gradient-based methods in the lecture. Why are there so many methods like this, and what are the limitations of each method, along with their motivations?

   Complete the table below with a short description:

Method	Year	Motivation	Methodology	Limitation
Saliency Map				
Smooth Grad				
Intergrated Gradient				
Grad Cam				

Table 1: Gradient-based methods

2. Explain the experiment in Section 6.4 in the Integrated Gradient paper https://dl.acm.org/doi/pdf/10.5555/3305890.3306024 in your own words. What do you conclude?

3. Shortly explain the problem with Integrated Gradients that is outlined in this blogpost https://distill.pub/2020/attribution-baselines/. You can also use images in your explanation.

Exercise 2 - Practical Exercise

4. In this exercise, we will practice gradient-based methods in computer vision classification tasks. We will continue using the ”xai” environment that has been installed with SHAP and LIME from last weeks.

   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_5_exercise.ipynb file and complete the exercises.