Introduction to Artificial Intelligence
Project 15: Heart attack prediction

Daniel Marczak

Summer 2023

1 Project description

The goal of the project is to create a model for heart attack prediction. Given a set of attributes of the patients (e.g. age, blood pressure, cholesterol level, etc.) as an input, the model should output a binary information if the patient has increased risk of hearth attack.

The data is available on Kaggle. The student should perform extensive data analysis and preprocessing: visualize, clean, analyze and split the data. The student can apply data augmentation or find additional data, if possible.

The choice of a models is a part of a task. The architecture should be chosen carefully taking into account various factors: data, computational resources, final performance, etc.

The evaluation of the model should contain the metrics commonly used for the given problem (e.g. test accuracy) as well as the visualization of the outputs.

All the important decisions made in the project should be properly motivated and tested experimentally. For example questions such as *Does data augmentations help?* or *Is architecture/method A better than B in this problem?* should be answered theoretically if possible and supported by experiments. An ablation study of the final solution is also very insightful (for an example take a look at Section 4.4 and Table 5).

2 Schedule

The project is done in pairs and has three stages. Maximum number of points from the project is 25. The project timeline, with points for each stage, is as follows:

- Preliminary assumptions submission deadline is May 16th. The meeting will be arranged on per pair basis on 19th May. (0-5 points)
- The midterm solution should be delivered by May 30th. The meeting will be arranged on per pair basis on 2nd June. (0-5 points)

- The final solution delivery deadline is June 13th. The meeting will be arranged on per pair basis between 16 and 20 June. (0-15 points)

Delay in handing in each stage results in a penalty of 20% total points available from the stage. **After each deadline, we will have project meetings on MS Teams**. The meetings are **mandatory** to complete the project, and we will do them in a week following the handing in of each stage. Please contact me to arrange the meeting. In case of any questions about projects, either ask me directly on Teams.

3 Technical details

Solutions should be implemented in Python. Jupyter Notebooks are not allowed. Please ensure that your code adheres to basic standards of lean coding in accordance to PEP 8 and add instructions on how to run the code. The report should follow the structure of a scientific paper. Please use templates from NeurIPS / IEEE / Springer / Nature etc for the report. The code, alongside the reports, should be uploaded on github. Please send the links to the repository as submissions for each stage of the project.

4 Assessment criteria

Preliminary report

The preliminary report should be a PDF file containing:

- Description of the task that you are working on, eg: what is the goal of the project and how do you frame the task you want to solve (regression, binary classification, multilabel classification. (1 point)
- Description of the dataset (1 point):
- Does it require any preprocessing or cleaning? Are the labels in the dataset balanced? How do you plan to address the data imbalance in the project? How do you plan to split the data for the project? How would you evaluate the performance of your solution? Consider both the metrics and data splits and look for an upper bound of your solution, eg. what is the best score on Kaggle. (1 point)
- Description of your solution (2 points): Describe briefly the usual methods used for the problem. Select two or three of them that you want to implement and compare. Add a more detailed description of those methods and the data preprocessing required for them.

Midterm solution

At this stage, you should present your progress, eg. what have you already implemented and how it works. The grading of this stage will take into account:

- Your progress in coding and evaluation of the method - at least one of the selected algorithms should be implemented at this stage. (1 point)
- Intermediate results and their presentation. Please add the results to the project report. The results should be presented accordingly, as plots or tables. Make sure your results are reproducible by adding a seed to your code. (1 point)
- Presentation and discussion on the challenges, results, and findings during the project so far, and potential corrections to the previous assumptions. This should also be included in the report. (2 points)
- Discussion on the plans for finishing the project, eg. what experimental results do you want to show in the final report? (1 point).

Final solution

The final solution should be the code and report in the form of a scientific paper. The grading of this part will be:

- Submission of the code for your solution and correctness of the code. (5 points)
- The final report should contain the information from the previous stages. The grading of report will be as follows:
  - The proper form of the report follows the format of a scientific paper (introduction, description of the data and algorithms that you will use, intermediate results and insights from the midterm solution, final experimental results, and conclusions). Don't include the code in your report. (2 points)
  - Experiments, presentation, and the discussion of the results. (4 points)
  - Comparison of the algorithms that you implemented, based on the results from the experimental section. (2 points)
  - Final conclusions - which algorithm performed the best, is the solution satisfactory and how could it potentially be improved? (2 points)

In case of any questions contact me via MS Teams