CPE 695 WS Final Project Instructions Due: May 8, 2024 Spring 2024

Overview

As previously stated in the proposal assignment, your team project should aim to use **one or more machine learning methods to analyze data and solve a problem** using **real data**. It should include all steps we need in order to properly apply an ML method:

1. Task formulation
2. Data analysis and preprocessing / feature extraction
3. Method formulation
4. Training and hyperparameter finetuning
5. Evaluation

Use the methods you have learned in this course to create a solution to your chosen task, which each team proposed in the "Team Project Proposal" assignment.

Submission Format

Each team should submit three files for the final project (submit through Canvas):

1. Written report (PDF)
2. Presentation video (Video file, e.g. MP4 format)
3. Source code (Jupyter Notebook)

---

1. The written report should have the following sections:

Introduction: Define the task you seek to solve, including goals for the project outcome

Data: Describe the dataset you used, including:

- The features of the dataset
- Preprocessing steps you conducted to effectively use the data
- The size of the training, validation, and testing splits

Method: This section should include:

- The machine learning methods you employed. Justify your choice via either qualitative reasoning or quantitative preliminary experimentation
- Details of the problem formulation relative to the method chosen
- Optimization (training) procedure, including hyperparameter finetuning steps such as Grid Search or Genetic Algorithm

Results: Discuss the results of your experiments, both the solution found and evaluation metrics proving your method's efficacy.

Conclusion: Summarize the project and results, and provide some ideas about future work which you could conduct to further improve your method.

---

2. The presentation video should include slides (Powerpoint or comparable) and be 12-16 minutes in duration. **Each team member must present** for an approximately equal share of the time (3-5 minutes). The video should include both your slides and video + audio of each speaker as they present.

You can use any method to record this video, but I would recommend one of the following:

1. All team members log into a Zoom meeting, and one of them screen-shares the slides. Present your project and record the meeting. Export the video file.
2. Record the video with slides, speaker video, and audio in Powerpoint. This can be done using the tools in the “Record” tab in Powerpoint.

---

3. The source code should be a Jupyter Notebook containing all code you used to generate your results, including data preprocessing, machine learning method implementation, and evaluation code.

Grading

Projects grades will be determined based on the following factors:

- **Proper data preparation:** was the data preprocessed appropriately?
- **System formulation:** is the system formulated in a reasonable way to solve the problem?
- **Suitability of chosen ML method(s):** can we expect the method to be a good fit for the task?
- **Implementation:** was the method executed correctly?
- **Results analysis:** did the authors conduct thorough analysis to evaluate the performance of their solution?
- **Adherence to submission format:** are all parts of the submission present (report, video, code), and do they follow the guidelines laid out above?

To maximize your grade, make sure you can answer “Yes” to every one of these questions.

Note that your source code must be **original and written by your team**; if your code contains >= 60% code copied from other sources, a large grading penalty will be applied.