Project description

As you are aware, this class includes a term project. The projects will be done in groups of 3-5 people. If you have a very good reason to work on the project by yourself, I may be able to make an exception, but you have to get my permission first.

Your projects will revolve around a specific dataset. This can be a very straightforward task of achieving the best possible performance given data, or you may focus on a more exploratory project applying advanced ML techniques to complex problems. In some cases you will tackle a known problem with an existing dataset attached to it. In other cases, you may decide to create your own dataset, annotated it, and wrangle it into a form suitable for ML.

Requirements

Your project must tackle a problem of a reasonable size and it must involve building a machine learning model. Your model performance has to be evaluated properly. This means that you should select a correct evaluation metric and test your model against a held-out test set. The test set should be allocated *before* any ML experiments are conducted with your data.

Evaluation

Your project will be evaluated majorly based on the quality of your project report and your oral in-class presentation. You will also be required to give periodic in-class project updates. When you are done with your project (during the week of the final exams), please submit a 3 page (longer if necessary) project report along with the code you have written. The project report must include the following sections:

- Introduction
- Dataset description
- Baseline approach description
- Method description
- Evaluation
- Discussion
- Conclusion
- Appendix (contribution of each team member)

During the last week of the semester you will be given a project self-evaluation form. This form will allow you to evaluate the contributions of each team member. This form will be optional. If no self-evaluation form is submitted, I will assume that group members contributed equally to your project.

Timeline

Your work on the project will consist of three phases:

Selection phase

- Bring project proposals to class
- Discuss individual proposals in class
- Form groups
- Each group picks a project
- Each group notifies the instructor about their choice
- Project abstract and group members are posted on Slack

Execution phase

- Dataset creation (optional)
- Data cleanup and normalization
- Train/dev/test split
- Model building
- Evaluation

Presentation phase

- Periodic oral updates
- In class oral presentation