DATA 300 Final Project Proposal

Due Date: 11:59 pm, October 24, 2024

Individual Submission Instructions for Final Project

Total Points: 10/100

Abstract

For your team's final project, you will explore how one or more machine learning algorithms impact a chosen application, focusing on **evaluation** and **visualization**. The project will involve:

1. Identifying a machine learning problem.
2. Researching related literature to select one or more suitable algorithms to address the problem.
3. Implementing an algorithm or set of comparable algorithms (and making modifications/extensions)
4. Performing extensive evaluation and visualization of the results.

Getting Started

1. **Form a Team:** Your team can consist of **3-6 members**. Each member should have a defined role in the project.
2. **Choose a Project Topic:**

- **Application Project:** Choose an application of interest and explore how machine learning algorithms can be applied to solve it. Examples include disease risk stratification, automated stock trading, movie recommendation, SPAM detection, or grade prediction.
- **Algorithmic Project:** Focus on a particular algorithm or family of algorithms, make improvements, and apply it to a specific machine learning problem.
- You are welcome to explore and implement machine learning or deep learning topics beyond the course material. For instance, areas like deep learning models, Natural Language Processing (NLP), Graph Neural Networks (GNNs), and beyond.

3. **Define Project Scope:** Your project should reflect **35% of your final grade**. Finalize your dataset and define the project's scope based on team size.

Deliverables

- **Code** and **Documentation**: Submit your code and a double-spaced report (5 pages per group member).
- **Visualization and Demo**: Prepare a poster summarizing the project. The poster should resemble a conference-style presentation.
- **Final Presentation**: Present your project on the designated exam date.
- **GitHub Page**: Maintain a GitHub repository for your project with regular updates.

Additional Notes

- You can build on prior work from other courses as long as the contributions for this project are clearly new.

- Most of the code should be your own. You may use libraries such as `scikit-learn` and other packages, but make sure to credit any external code used.

Proposal Submission

Submit a 2-page proposal including:

1. Project title and team members
2. Project description
3. Roles and contributions of each team member
4. Course-related topics to be explored
5. Describe the dataset you will use
6. Provide a timeline for the project
7. Describe what you plan to demo on the final exam date and other deliverables
8. Describe your plan to evaluate the project

Looking Ahead

You'll receive feedback on your proposal within two weeks. Multiple in-class updates will be scheduled, with the final presentation on the exam date.

Grading Rubric

- **Correctness and Technical Quality**: Are the methods and solutions reasonable and sound?
- **Significance**: Is the problem interesting and impactful? or is it a trivial "toy" problem?
- **Novelty**: Does the project offer original insights or a novel approach?

Project Examples

Stanford CS229 2019, Stanford CS229 2020, Stanford CS229 2021