Spring 2021 CSCI 740

Machine Learning Final Project Proposal

Assigned Date: February 25, 2021 Due Date: 11:50pm, March 18, 2021

Individual Submission

(each team member should individually submit the team document)

Total Points (towards project): 10/100 points

Abstract: Your team's final project should study the effect of one or more machine learning algorithms on an application while focusing on evaluation and visualization aspects that are the tenets of the course. The project will involve

a) identifying a machine learning problem,
b) reading appropriate related literature and identifying one algorithm or a family of algorithms to address the problem
c) implementing an algorithm or set of comparable algorithms and making modifications/extensions to add a novel component and performing extensive evaluation and visualization of the results.

A: To get started:

1. identify **project members**: you can have a three or four person project (identify members and roles for each person in the project).
    - You can use this link to indicate your availability and project topic interest (use Searching for Team section first and then Completed Teams). You can also use the project discussion board to determine project topic interest.
2. identify **project topic**:
    - **Application project**. This is by far the most common: Pick an application that interests you, and explore how best to apply learning algorithms to solve it. e.g. Disease Risk Stratification, Epidemic Prognosis, Automated stock trading, Movie Recommendation, SPAM prediction, Grade Performance Prediction
    - **Algorithmic project**. Pick a problem or family of problems, and implement an existing sophisticated learning algorithm with a novel variant of an existing algorithm or a completely novel approach, to solve it.
    - Topics may include any of the course topics supervised learning: generative/discriminative learning, parametric/non- parametric learning, neural networks, and support vector machines; unsupervised learning: clustering, dimensionality reduction, kernel methods; learning theory: bias/variance tradeoffs; VC theory; large margins; reinforcement learning and adaptive control and fairness and bias.
    - Take a look at this document for a list of topics and tools gathered from a previous semester here.
3. determine **scope of the project (project goals): this** will depend on number of members in project. The project represents 30% of the final grade for each student (while each assignment is about 15%) – the scope and amount of work for each student should reflect this grade level. It will be good for the team to nail down the dataset the will be used for this project by the time you submit this project proposal.
4. **project deliverables should include**:
    - code,
    - final demo/visualization,
    - evaluations of your approach,
    - documentation (submit a double spaced report (5 pages per group member),
    - 2 minute video or a poster: the video should be end-to-end pitch of your project with motivation, solution approach (novelty/significance), evaluation and conclusion/future work sections. It should not be just a walk-through of your demo. A poster should be similar to a conference style poster with similar sections.
    - Final presentation

- You should plan to maintain a github page for your project, provide regular updates on your achievements and challenges, submit a double spaced report (5 pages per group member) and give a final presentation with a demo on the date the final is scheduled.

Other Notes:

- You can build on existing prior research that you have done in other classes, as long as new work is being done for this class project and sufficiently self-contained, but in this case, you must clearly state in your proposal and final report what part of the project was done before this course and for this course. You should identify a novel aspect of this project. You can however, use this project for publications of your main research.
- Contributions related to CSCI 740 topics must be identified – each member's contribution should be clearly specified for teams. **Ph.D. students will have an additional required deliverable in that should plan a 15 minute presentation on a peer-reviewed paper related to their project topic in class.** Final grade will have both a team contribution component and individual contribution component.
- I am happy to discuss topic relevance and scope or any other question related to the project as you develop your proposal via email or a zoom meeting.
- The code for the project must be mostly your own. You can use any of the existing libraries including sci-kit learn and other packages. Any use of outside code must be discussed with me and acknowledged.

B: What to submit by project proposal deadline? Initial **2 page draft of your proposal** that includes the following sections:

1. Project Title, Names of Project Members
2. Describe your project.
3. Each team member's role and contribution should be clearly.
4. Enumerate the CSCI 740 related topics (questions and solution approaches) each individual will contribute to the project.
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.

C: Looking ahead:

- You will receive my feedback on the proposal by March 25, 2021**.**
- We will schedule multiple in-class updates of the project starting in early May with the final project presentation scheduled for the day of the final exam. These updates will have final project points associated with them.

D: Grading Rubric of project(similar to Stanford project and most conference paper submission review process) :

- Correctness and technical quality of the work: Does the technical material make sense? Are the things tried reasonable? Are the proposed algorithms or applications clever and interesting? Do the authors convey novel insight about the problem and/or algorithms?
- Significance of Problem: Did the authors choose an interesting or a "real" problem to work on, or only a small "toy" problem? Is this work likely to be useful and/or have impact?
- Novelty of the approach and results: Is this project applying a common technique to a well-studied problem, or is the problem or method relatively unexplored?
  In order to highlight these components, it is important you present a solid discussion regarding the learnings from the development of your method, and summarizing how your work compares to existing approaches.

E: References:

- Model for topics and posters (this is ML specific, your projects can be on broader AI topics): Stanford ML course http://cs229.stanford.edu/projects.html