Rensselaer Polytechnic Institute
Department of Electrical, Computer, and Systems Engineering
ECSE 4840: Introduction to Machine Learning, Fall 2024

Project Proposal and Progress Report: due Saturday, October 26th, at 11:59PM.
Submit your report on Gradescope.
Use the team function to submit a single document for a group of 2.

The projects for this course should be done **in groups of one or two** and should be related to some real-world machine learning problem (preferably one that you find interesting). The goal is for you to get your hands dirty with data collection and machine learning tools, putting the concepts that we've learned in class to work. If you prefer to investigate a hardware implementation (e.g., using an Arduino/Raspberry Pi, embedded processor, etc.) that's fine, but remember the emphasis should be on machine learning algorithms, not hardware considerations. Another critical point is that the project shouldn't be fundamentally a computer vision/deep learning project, or use a lot of algorithms we didn't talk about in class (e.g., transformers and Tensorflow). I mainly want you to demonstrate that you understand how to apply the algorithms we directly covered in class. You're welcome to use Python, C++, Matlab or whatever you want for the final project.

The final project should leverage **three or more concepts from class** using a dataset that **you collect yourself**. For example, you could compare and contrast logistic regression, a support vector machine, and a multilayer perceptron to address a multiclass classification problem, or you could apply cluster analysis and principal component analysis discover patterns in your data before solving a nonlinear regression problem.

If you want to do a project that has some overlap with a project for another course, you must clearly disclose it. This "double-dipping"is permissible only if (1) the overlap with the other project is fully described and disclosed, and (2) there is at least 50% "new content" that you're only doing for this class and you wouldn't have done otherwise. That is, you must demonstrate that the course project you submit for my class is something fundamentally new that you would not have done anyway for some other course. I'll let you know whether what you propose seems sufficient/acceptable.

Remember, the project is worth 30% of your grade and takes the place of a final exam; it can't be something that you bang out a couple days before the deadline. It's meant to represent a careful, sustained amount of work. You should have begun to collect data and make significant progress throughout October and November. You will present your project in class at the beginning of December and submit the final writeup/code/demo on the last day of class. Due to tight grading deadlines, no extensions will be granted!

The proposal/progress report for your project should be a document that contains the following sections:

1. Introduction. Now that some time has passed and the semester is half over, you should have a much clearer idea about your project and what will be viable to complete in the time remaining. Describe your project clearly and specify how it flows from concepts we discussed in the course. Remember that the focus should be on basic machine learning concepts we covered in class.

2. Data Collection. Your dataset should be something you can feasibly collect and should be fun, unique, and personal dataset you collect and curate yourself. What I *don't* want is an off-the-shelf ImageNet/Kaggle dataset that someone else collected using a web scraper. The goal is for you to get your hands dirty and be intentional about what you put into the dataset. Collecting your own data will also make you think more about the relationship between your inputs and outputs and to shape the results to your liking.

   I asked you to speculate about this for HW1 but now is the time to be more realistic and focused. What datasets or source images/video have you already assembled and what data do you have left to collect? At this point I'm expecting that you have most of what you think you'll need and clear plan for collecting the rest.

I don't want a "kitchen sink" approach to data collection; you should be clear-headed about exactly what you want the input and output of your algorithms to be, and have good reasons for why they should be related. A common issue in HW1 was the proposal of output variables that might be subjective, unrepeatable, or time-consuming to obtain (e.g., user ratings of moods, emotion, alertness, etc.) so be careful to address this if you have this type of project.

3. Technical Approach. Now that we have covered much of the syllabus, what three (or more) techniques do you anticipate demonstrating in your project? There should be some form of comparison of using different techniques for the same problem and evaluating which worked better or was more interpretable. This part should be as detailed/mathematical as you can make it. A good project should have some component of hand-written code to demonstrate that you didn't just hook together various libraries.

4. Preliminary Results. Include any initial results from your work in progress. Include any "making-of" graphs that give insight into intermediate steps. Highlight both strengths of your current algorithms/results and weaknesses that you plan to fix before the final project is due.

5. Broader Implications. Discuss the question of whether there would be any ethical implications for collecting your proposed dataset or applying a machine learning model trained on it, based on the considerations we discussed in class.

6. Plan for Completion and Further Work. What tasks are required to complete the project, and what is your timeline/plan for getting them done?

This may end up being a long-ish document, but you will add to it to create the final report. The goal is to get you thinking about the details of the project now vs. at the last minute.