

SUNY Albany
MAT810 Spring 2022

Advanced Topics in Machine Learning

Your class project is an opportunity for you to explore an interesting machine learning algorithm of your choice in the context of a real-world big data set. We will provide some project ideas, but the best idea would be to combine the topics of this course with problems in your own interest. Your class project must be about new things you have done this semester; you can't use results you have developed in previous semesters. The key is to implement these machine learning algorithms from scratch.

Projects can be done by you as an individual, or in teams of two students. Your project will be worth 30% of your final class grade. Your final project is a writeup in the format of a NIPS paper (8 pages maximum in NIPS format, including references; this page limit is strict), due May 6th. You can use the following NIPS template (<https://nips.cc/Conferences/2015/PaperInformation/StyleFiles>). You can submit the jupyter notebook with a detailed documentation as supplement, but your final project report should be self-contained.

In addition, you must turn in a brief project proposal (1-page maximum) by April 21. Read the list of available data sets and potential project ideas below. You are encouraged to use the data sets from the case studies (will be on the blackboard soon). If you prefer to use a different data set, we will consider your proposal, but you must have access to this data already, and present a clear proposal for what you would do with it.

Project Proposal

Project proposal format: Proposals should be one page maximum, due by April 21. Include the following information:

1. Project title and data set.
2. Project idea and machine learning algorithms
3. Papers/textbook to read. Include 1-3 relevant papers. You will probably want to read at least one of them before submitting your proposal.
4. Teammate: will you have a teammate? If so, whom? Maximum team size is two students.
5. Projects must have a "big data" aspect. This means either (i) high dimensional data, (ii) a very large number of observations. You should not use MNIST and swiss roll now.
6. Projects must have a "machine learning" aspect, and cannot just be exploratory data analysis. Ideally, the ideas should be strongly based on concepts covered in this course, i.e., un/semi-supervised learning.

Final project

Final project format: NIPS format, due by May 6th.

1. Design or raise your scientific problems (a good problem is often more important than solving it).
2. Implement the proposed algorithm from scratch.
3. Explain the mathematical theory in your own language. It is even better to demonstrate it by designing the examples.
4. Show your results with careful analysis supporting the results toward answering your problems.