# Machine Learning: Course Project

(Worth 50% of your final grade)

## Project Intention & Opportunity

The project is a fundamental element of this course, and is also designed to be an exceptional opportunity: *a well-done project is ideally intended to lead to a publication*. For example, here are projects from an introductory ML class at Stanford: http://cs229.stanford.edu/proj2019aut/ , and here are projects from an introductory ML class (that uses the same textbook as us) at UC Berkeley: https://courses.d2l.ai/berkeley-stat-157/project.html .Some of these represent the target quality you should be aiming for. On the Stanford page, they describe the projects as follows: “A very good CS229 project will be a publishable or nearly-publishable piece of work. Each year, some number of students continue working on their projects after completing CS229, submitting their work to conferences or journals.” This is exactly what we are aiming for as well.

*Opportunity: A “GPU compute prize” + additional mentoring towards publication: If you have a project that is suitable for publication in a top-tier venue, I expect to be able to provide you with access to significant GPU resources for that project (at most one or two groups) to assist in publication submission (if I anticipate that resources will make a critical difference in the quality of results). Furthermore, in that case one of the TA’s and/or I would be glad to provide you with continued mentoring after the course is completed, to prepare and submit your work for publication. (In case you are new to graduate school, or planning to apply to graduate school: publications are effectively an important “currency” of research-- they increase job opportunities, funding/scholarship/grant opportunities, further graduate school admission opportunities, and even make it easier for your reference-letter-writers to write you stronger letters.)*

## Project Groups & Group Size

The projects will be done in groups of 3 or 4. Other size groups are possible, as long as you obtain permission from the instructor; to do that you must provide a short letter of justification (e.g. Why is this topic X better suited for this particular group of n people than 3 or 4? What will this group do differently from what a group of 3 or 4 would likely do? What demonstrated experience do the individuals proposing the bigger/smaller group bring to the table that warrants special consideration?). The contributions of each member of a group will be articulated and evaluated separately, so that each member in a group will receive an individual grade. The expectations for a group increases in proportion to the size of the group.

To help you choose your teammates, I have invited all of you to write and share a short statement about yourselves in a shared introductions document. This can include relevant experience in math, CS, ML, etc, as well as research interests, any application areas that you are passionate about, and what you hope to achieve in this course. You might also wish to

include the time zone that you are currently living in. *If you cannot find a complete team by Sunday, please contact me.*

## Project Topics

The range of possible topics is very wide. Here are some general comments on how to approach topic selection:

- Look at the Stanford intro course project page or the Berkeley project page for inspiration.
- You can begin by focusing on an application and then finding effective ML approaches for that problem, or you can choose an ML problem and develop a new algorithm to solve it (the latter is harder than the former).
- Another approach (more suitable for people with more experience in ML): start by choosing papers that interest you and suggest and build an extension of one of them; alternatively, you can experimentally analyse their failure modes and suggest/implement techniques to improve. *Challenge: many DL approaches require significant compute resources, which are not easily available; I have obtained some GPU compute resources for everybody in this course.* There are many other interesting things that you can do, however, ranging from exploring internal representations to fine-tuning existing models and developing small and creative synthetic datasets.
- Whatever you do, I strongly suggest you pick a topic or area that you are truly passionate about: this will make it more interesting for you, more fun for you, and will motivate you to work harder.
- I encourage you all to discuss project and topic ideas--both generally and in detail--on Team chats as you see fit. The TA's and I will be reading the chat and may occasionally be able to comment this way.
- Feel free to take a look at this work-in-progress document that includes lists of topics and papers that I am interested in. You can also add comments to this document, as well.

F.A.Q. (to be updated)

- Can we choose a Kaggle problem?
    - A lot of work in a typical project goes into defining the problem, possibly preparing a dataset, figuring out how to evaluate. So if you choose a Kaggle problem, this is done for you, and so you will be evaluated primarily on either how well you score relative to the leaderboard (or on how novel your technique was, in which case it is your new algorithm that is probably the focus of your project, and kaggle set is just provides application/baseline). If your project plan is to score well on the leaderboard, then keep in mind how much compute you have available to you… (but if you're going for that, then you've probably already

thought through all that because you have a lot of experience training models, right? :)

## Project Proposal (P0) (due ~~Feb 5~~ Thurs Feb 11 at 10am AST)

The proposal is due next Thursday. It is not marked, but rather will be given feedback within a day of “approve/revise”. It must include the following:

- **P0: Declare Group & Tentative Topics**
    - *Optional Early Feedback:*
        - 1 proposed topic
        - 1 additional backup topic
            - Include 1-3 bullet points per topic (ideally: short, clear and detailed)
            - Include 2-5 relevant papers you were able to find so far for each topic
        - If you can send this by Sunday, we can try to reply within roughly a day with a tentative approve/revise.

    - By Thursday morning, please submit a 1-3 page proposal that includes:

        - the main parts/steps of your project, and estimates of how long they will each take.
        - description of your dataset(s) if relevant/appropriate
        - estimates of what the “initial implementation” will consist of (ie. by week of Feb 23/25) and what the “final implementation” will ideally consist of (i.e. by week of March 23/25). There will be a chance to revise the plans in early March.
        - Anticipated challenges; anticipated “easy parts”; backup plan for anticipated challenges. This might include explanations of what you are uncertain about: are there some parts you are sure will be straightforward? are there other parts you are less confident about? Are there some parts that you just don’t know yet how they will go, because they depend on a previous step that you have not done yet? Keep in mind that things don’t always work; negative results are OK in a course project especially if you have a plan for how to explain them well.
        - A tentative description of who will do what, and by when they will do it

        - Your proposal must include at least 1 “immediate next step” for team member that you can start working on even while you are waiting for our feedback.
        - A list of ~3-5 related papers and/or legitimate blog posts (check on Teams if not sure about appropriateness of a particular venue)

*Note that slightly more detailed requirements for subsequent project phases may be given as the term progresses, if appropriate.*

- ## Project Initial Implementation (P1)
  **(Due Week of Feb 23/25, referred to as “Mid-term Outline” in Syllabus)**

  **What to do:** As described in your proposal, in this step you will complete whatever initial implementation is appropriate, e.g. possibly one or more of the following:
    - forking or re-implementing existing implementations for baselines on proposed dataset (e.g. a new application of an existing technique)
    - creating a new toy dataset and running basic models on that (e.g. for a new ML technique)
    - pre-processing data set as needed
    - evidence of some training
    - ADVICE: keep in mind that training a model can take hours or even days. To have even something basic working in 3-4 weeks means you need to start working on this right away. Sometimes you train for 2 days and then it crashes, or you find there was a bug, and nothing actually happened!
    - Make sure to keep track of what you tried; I strongly suggest you include a list of things you tried as supplementary (appendix) material, even if they were not all successful.

  **What to submit:** You will begin by submitting a document that includes the following:
    - A clear, succinct description of your initial implementation: what you did, how you did it, what existing codebase(s) you used, what you implemented yourself
    - Summaries of any experimental results you obtained, along with your brief interpretations, and indications for next steps
    - A link to your code available on github in case a TA or Instructor are interested in seeing it (code does not have to be clean; research code can sometimes be messy, especially at first)

- ## Project Update (P2)
  **(Due Week of March 9/11)**
    - A short list of modifications of your plan relative to your original proposal.
    - The purpose of the update is not to “make work”, but to make sure you are on track, and have adjusted reasonably to what you have discovered and learned so far. It is also an opportunity to make sure you are aware of your own learning and working process and experience.
    - Please include a very brief discussion of:
        - what has worked so far? what has not worked?
        - what has been an effective way of working? how can you be more effective, both individually, and as a team? what do you need in order to be more effective?
        - what is the best use of remaining time?

- ## Final Project Videos & Reports (P3)
  **(Due Week of March 23/25)**

    - Your final project will be presented in the form of a video and a report. As there is no exam in this course, it is expected that this will be a substantial amount of work.

- **Project Videos (Due March 22)**
    - Submit a video describing your work. It should have a maximum duration of roughly 5 minutes. You will submit it as an unlisted youtube link. More details on the final project video can be discussed as the term progresses.

- **Project Report (Due March 27)**
    - Submit a report either in the NeurIPS conference format (with same page length constraints, etc), or--if you obtain permission in advance--in a conference format of your choice, that would be an appropriate venue for submission of your work. As with many conferences, you will also be allowed to submit an unlimited amount of appendix material, but there is no guarantee that the appendices would be read. More details on the final report can be discussed as the term progresses.

## Project Marking Notes

*While the project will be marked out of 100, the following criteria add up to 145*, ie. considerably more than 100 marks, because there are different ways of doing an excellent project. Your marks will be added up, and you will receive up to a maximum of 125/100. An example will be given below.

**Marking Rubric [subject to change slightly]:**

- [10 marks] - Creativity
- [10 marks] - Baselines
- [10 marks] - New & Appropriate Synthetic / Toy Data set
- [20 marks] - Quality of writing
- [10 marks] - Quality of figures: relevance, clarity, presentation
- [10 marks] - Insight provided: either in writing and analysis, and/or in choice of experiments and/or dataset
- [10 marks] - Novel contribution
- [20 marks] - Evaluation: Choice & Thoroughness of Experiments and Analysis of Results
- [5 marks] - Attention to Detail (in research choices made, in analyses)
- [20 marks] - Quality of presentation materials (video, interactive notebook, etc)

- While there are many ways to get points, deductions (e.g. for irrelevance to course material) are also possible. For example, while linear regression might be an appropriate and important baseline, a project that only does linear regression, or that only demonstrates a standard single feed-forward network trained on a minimally-processed dataset, would be heavily penalized. For feedback on possible deductions, pay attention to the intermediate feedback you are receiving.

So for example, suppose a project made by a group of 4 people is extremely creative (10/10), completely omits expected baselines (0), proposes a very interesting and insightful new dataset (10/10), is written OK (10/20), has great figures (5/5), good insight (8/10), OK contribution (5/10), mixed evaluation (good experiments, no baselines, decent analysis 12/20), good attention to detail (4/5), and excellent video (18/20), that project would get 82/100. That is, even though it did not include baselines, it compensated with other elements. Note that the group size would have been taken into account when giving all of these marks.

Note also that some of these criteria are closely interrelated. For example, a lack of baselines will impact the possible score on evaluation; but they are not the same. Similarly, an evaluation can be quite thorough, but without an insightful analysis, and so Evaluation and Insight are also related, etc.

While the basic rubric will be as above, undergraduate and graduate students will be marked based on different, and appropriate, standards.

The mark described so far is a baseline mark for all members of the group. Individual marks may then be raised or lowered relative to that baseline, depending on the articulated contribution of the individual.

**Changing the Marking Scheme**. Ultimately, the goal of this marking scheme is to allow you to play to your strengths and reward you for good work, and also to accomodate for the differences between topics, e.g. some topics are inherently creative but hard to evaluate, etc. In exceptional cases, it is possible that additional “marking categories” could be added to the above list, if it feels that a particular project is truly strong, yet its strengths are not apparent when considering it through any of the above already-listed categories. If that happens, then I will likely add that category in some way for future versions of this course. Also, if that is the case, then please provide a clearly-written request--well before you submit your final project, ideally around the time of your Project Update--explaining why the above categories are insufficient for grading your project, and what additional category or categories you would suggest. The goal is to have a project which could eventually become either a conference publication or a high-quality interactive blog post.