CSE 251B Winter 2021 Project guidelines
(much of this is cribbed from the excellent guidelines for CSE 254 by Charles Elkan!).

These are some guidelines to do the final project.

Projects should be done in teams of five or six people. Teamwork is very important in science. However, don't divide the project up so that people can work completely independently! Constant communication within the team is very important!

Because you only have two weeks for a final project, you should not be too ambitious!

However, the project should reflect what the course is about! All projects should involve deep networks, whether feedforward or recurrent.

I understand that many people post the code to their work on the web, or that some of the code is available in other simulators besides PyTorch. Just running someone else's code is ***NOT ACCEPTABLE AS A PROJECT!*** If you use someone else's code, then you should document that in your report. Not doing so is plagiarism, and you will receive a zero on your final project if we discover that you have not documented code you have used. If you use someone else's code, you should not just replicate what they have done, but explore some variations (hopefully, improvements!), apply it to new data, or try it on a completely different problem than it was designed for.

The ***most important part of a project is easy access to training data.*** It is often surprisingly difficult to find training data that is sufficiently large to train a deep network - especially if you have some new idea you want to try. Anything you decide to try, the ***first*** thing you need to do is determine whether there is training data that you can just download off the web. Trying to create your own training data is a Very Bad Idea.

You are welcome (and encouraged!) to use PyTorch. You may use other simulators if you want, but please note the next paragraph.

Projects could be:

1. An attempt at replicating a recent result. ***WARNING: This is often very hard!*** Crucial details are often left out of papers. Authors may not always be precise in what they say. Hence, replicating a previous result is often a very good learning experience, and motivates you to put these details in your own papers! Again, many of these recent results are done by people at Google, who have access to thousands of machines. So in replicating something, you probably will want to scale it back. In particular, it is a good idea to test things out on a smaller dataset to make sure your code is doing what you expect before you waste a lot of our precious resources on code that doesn't work. Here is a list of curated Deep Learning papers that you can check for project ideas. Unfortunately, due to the avalanche of deep learning papers, the maintainers of this list have given up

maintaining it! So it ends around 2017.

2. Of course, the coolest stuff being done right now mostly involve recurrent networks with LSTM or GRU units, or reinforcement learning. PyTorch makes this much easier. It would be a good idea to check Andrej Karpathy's blog on the unreasonable effectiveness of recurrent networks.

3. Something of your own design. Do **NOT** try to predict: the stock market, baseball scores, or to recognize the key of a musical piece. These sort of projects always fail. Pick something that seems easy, but not too easy. You will usually find it is harder than you think.

I will not list possible topics, because then everyone will try to do those instead of something cleverer (if that's a word). There are potential projects all over the web. And there are datasets all over the web.

The schedule for the projects is as follows.

(0) Early and often, you should come and discuss potential project ideas with me or the TAs or any other expert. Email is not great: I have quite a few unread emails. If you send me an email, also send me a text to tell me you sent me an email (and identify yourself in the text): 619-823-3033.

These dates are (very similar to those) given in the syllabus - there should be no surprises here:

Handed out: Wednesday, February 24th;
Project proposal due Monday, March 1st;
Progress report due Friday, March 5th.
Draft final report due Friday, March 12th.
Final report and presentation of project March 17th and 19th: Time to be determined by Poll.

(1) On Monday***, March 1st,*** submit your project proposal via gradescope. You should specify:

- what the main idea of the project is
- where you are going to get the data (in fact, ***you should already have it by then***), and ***what you are going to do instead if you can't get it.***

- what the architecture of the network will be
- one or two recent (or old, depending) papers that you will build on, with full bibliographic data.

The proposal should be written in well-organized continuous English, as opposed to just an outline. Most of its text should be reusable in your final report. The proposal should be at least two pages long when formatted using NIPS format. In any case, at this point,

you should make sure you get feedback from me or the TAs - bug me if I don't respond! It is crucial that I don't give you enough rope to hang yourself.

(2) Start work immediately on your project. The first phase of the project is especially important. In this phase, some important tasks include:

* formulating clear, sensible goals, or if appropriate, hypotheses to test,
* finding useful data **(very important!)** and designing experiments
* selecting your platform (e.g., PyTorch)
* finding online code, which, if it has everything you need, you need to then decide what you are going to do to do it somewhat differently! I.e., are you:
  -going to apply it to a new dataset that wasn't used in the original paper
  -apply it to the same dataset but vary the architecture to see if you can improve it
* just applying existing software to the same dataset that was used originally is ***definitely not ok.***

These are tasks that can and should be performed mostly in parallel, not sequentially.

(3) On ***Friday, March 5th*** you should submit a progress report via gradescope. This must take into account comments received from the instructor and/or the TAs. By this time, you should have already started running the network, debugging, etc. If you make architecture changes to an existing system, use a smaller subset of the dataset to make sure the network is training properly. This will allow you to get faster turnaround for debugging purposes.

(4) On ***Friday, March 12th,*** you should submit a draft of your project report to Gradescope. The report should be polished and should resemble a good submission for a conference. You should have preliminary data to report. A good example of evaluation criteria for papers are the NIPS paper evaluation criteria. Read, think carefully about, and follow all the principles of good writing in the "Nuts and Bolts" guide to rhetoric by Michael Harvey, available as an ebook for $8.40 on Amazon. I found a pdf on the web of this book by googling the title. Here are a few other web-based guides to writing: Seven tips to writing a good paper (Psychology Today), Charles King's "How to write a research paper", Simon Peyton Jones's "How to write a great research paper" (there is a youtube version of him giving this as a talk).

(5) On Wednesday***, March 17th, or Friday, March 19th,*** you should submit your final project paper, and be prepared with a 5 minute talk on it, time and zoom link TBA. **PLEASE DO NOT COME AND THEN LEAVE AFTER YOUR TALK.** That's rude to the other teams that prepared projects, and you might miss something really, really cool.

The final version of your report should be a PDF file in NIPS format that is no more than a couple of megabytes. If it is much bigger than this, you need to format your figures differently!

You should spend about eight hours per week per person on the final project. If you are spending much less time, you are not putting in the effort of a full-time student. If you are spending much more time, you should think more about efficiency and prioritization.While doing the project, remember that winning at research is similar to winning in many other fields of endeavor.

* Build on an idea that has been successful in previous work.
* Make the description of your work understandable, attractive, and memorable; pick catchy names.
* Keep the work simple. Let the basic ideas shine through.
* More papers = more ways to have impact and be noticed.