

CIS 419/519 Introduction to Machine Learning

Course Project Guidelines

1 Project Overview

One the main goals of this course is to prepare you to apply machine learning algorithms to real-world problems. The final course project will provide you the opportunity explore such an application of machine learning to a problem of your own choice.

Projects must be completed in **teams of three students**. Ultimately, all teams (regardless of size) are expected to produce a project of equivalent scope. If you have a particularly ambitious project idea that cannot be completed by a team of three people, you may propose a team of four students, but you must have a strong justification for such a larger team. **You may not complete the project solo or as a pair, unless one of your project partners drops the class.**

Milestones and Deadlines

- **Project Proposal:** due Wednesday, March 4, 2020 11:59pm (no late submissions)
- **Project Status Check-In:** due Monday, April 6, 2020 11:59pm (no late submissions)
- **Final Report & Summary Slides:** due Mon, April 27, 2020 11:59pm (no late submissions)
- **Project Extra Credit:** due Friday, May 1, 2020 11:59pm (no late submissions)

Grading Breakdown

- Project proposal: 10%
- Project status report: 10%
- Final summary slides: 10%
- Final report: 70%

Evaluation Criteria

- Technical quality (i.e., 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 (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 work (Is the proposed application and approach novel or especially innovative?)
- Clarity of presentation (Is the presentation clear? Could we reconstruct the method entirely from the report?)

Students enrolled in the graduate version of the course (CIS 519) will be expected to complete a project of significantly higher scope, quality, and polish than students in CIS 419. Specifically, CIS 519 projects are expected to be of sufficient quality for a machine learning workshop publication. Teams may include students from both CIS 419 and CIS 519, but projects from combined undergraduate/graduate teams will be graded under the CIS 519 criteria.

Although I encourage you to implement your project in python using `scikit-learn` or `PyTorch`, you may use other software or programming languages if you have a particularly compelling reason.

2 Choosing a Topic

Your first task is to identify a topic for your project. One of the best ways to identify a topic is to choose an application domain that interests you and identify problems in that domain. Then, explore how to apply learning algorithms to best solve it. Let the problem drive your choice of technique, rather than the other way around. Most projects will be based on particular applications.

Alternatively, you can also choose a problem or set of problems and then develop a new learning algorithm (or novel variant of an existing learning algorithm) to solve it. Although CIS 520 is intended more to prepare you to develop novel learning methods than CIS 419/519, you may choose to develop a novel learning method (or novel variant) if you want a challenge.

Regardless, most projects will combine aspects of both applications and algorithms. Your project **must** include an evaluation on real-world data (i.e., not a “toy” domain or synthetic data).

2.1 Ideas

Many fantastic course projects will come from students choosing either an application that they’re interested in, or picking some sub-field of machine learning that they want to explore more, and working on that topic. If you’ve been thinking about starting a research project, this project may also provide you an opportunity to do so.

Alternatively, if you’re already working on a research project that machine learning might be applicable to, then working out how to apply learning to it will often make a very good project topic. Similarly, if you currently work in industry and have an application on which machine learning might help, that could also make a great project.

Here are a few other sources of project ideas:

Course projects/suggestions from similar courses at other universities

- Stanford, 2016: <http://cs229.stanford.edu/projects2016.html>
- Stanford, 2015: <http://cs229.stanford.edu/projects2015.html>
- C. Guestrin, CMU: <http://www.cs.cmu.edu/~guestrin/Class/10701/projects.html#datasets>
- Ray Mooney, UT: <http://www.cs.utexas.edu/~mooney/cs391L/project-topics.html>
- Amy McGovern, OU: http://www.mcGovern-fagg.org/amy/courses/cs5033_fall2019/index.html

Eric's list of project suggestions

- Extend an active learning technique (which queries the user for labels) to use other sources of feedback that are richer than binary labels, such as equivalence sets, distribution examples, measures of "typicality" of the instance, or some other idea of your own.
- There are multiple ways to combine kernels together to create new kernels (addition, multiplication, etc.). Develop an SVM-based learning algorithm that tries a number of kernels and their combinations in a principled manner to find the optimal separator for a data set.
- Multi-view learning is typically applied to supervised or semi-supervised classification scenarios. Instead, apply it to unsupervised clustering or constrained clustering.
- Write a reinforcement learning agent to play Mario or Tetris. Or, write a deep RL agent to solve one of the problems on the OpenAI Gym (<https://gym.openai.com/docs/>).
- Design an algorithm for transfer learning that improves image classification in some categories of the Caltech 256 data set based on transfer from other categories, or object recognition in the MIT objects and scenes data set, or indoor scene recognition. Transfer could also be used to improve image segmentation in the Berkeley image segmentation data set, or to do style transfer in images to make a photograph look like it was painted in the style of a famous artist.
- Often times, users have an idea of the classifier they are looking for, even if the data does not directly support it. Design an interactive method for building a model in collaboration with a user. For example, perhaps the user knows that particular attributes should be in the first few splits of the decision tree, even if there isn't enough data to support it, so the tree could be interactively built in collaboration with the user. Or, perhaps the user knows that particular factors are especially important.
- Use a deep learning technique called generative adversarial networks to simulate the generation of novel images of diseases. These images could then be used for training physicians to diagnose the disease. Or, generate novel images based on keywords that could be used in presentations (e.g., a bus that looks like a cat face).

Look through papers from recent machine learning conferences

- Neural Information Processing Systems: <http://papers.nips.cc/>
- Int. Conf. on Machine Learning 2019: <http://proceedings.mlr.press/v97/>
- Int. Conf. on Machine Learning 2018: <http://proceedings.mlr.press/v80/>
- Int. Conf. on Machine Learning 2017: <http://proceedings.mlr.press/v70/>
- Int. Conf. on Machine Learning 2016: <http://jmlr.org/proceedings/papers/v48/>
- Int. Conf. on Machine Learning 2015: <http://jmlr.org/proceedings/papers/v37/>
- Int. Conf. on Machine Learning 2014: <http://jmlr.org/proceedings/papers/v32/>
- Int. Conf. on Machine Learning 2013: <http://jmlr.org/proceedings/papers/v28/>

Final Advice Pick a topic that you can get excited and passionate about! Be brave and feel free to propose ambitious things that you're excited about. Finally, if you are not certain what would make a good project, we encourage you to e-mail us or come to instructor/TA office hours to talk about project ideas.

3 Project Proposal

Your first deliverable is a one-page project proposal that includes the following information: project title, names of all teammates, and a description of what you plan to do. Your proposal must be one page in length, single-spaced in 12 point font, with 1 inch margins.

You should write a compelling proposal that describes your project in detail and demonstrates that you have the understanding and ability to complete it. Your proposal should also discuss sources of real-world data for your chosen application or how you plan to obtain real-world data. Since you may wish to use machine learning methods that we have not yet covered, you may need to read ahead. Do not worry if there are particular aspects of the project that you can't answer currently (such as which ML method is best); this is a proposal for future work, after all. However, your proposal should demonstrate that you've started to think through the various issues involved with your project/data and present a compelling argument in support of it.

Consult Heilmeier's Catechism¹ (excluding the cost and time estimate) to determine what should be in your proposal. Imagine that you are bidding for funding, so your proposal should be a compelling argument that convinces me your project is a good idea, important, and that you have the capability to complete it successfully. And, you must do all of that in only one page.

You will be submitting your proposal using www.gradescope.com. Log onto gradescope, and submit the PDF files to the CIS 519 assignment entitled "Project Proposal". **Only ONE person from each team should submit, but during that submission process, that one person must choose your other teammates by name, turning this into a group submission.**

In the past, we've had significant problems with the submitter not selecting their group members. **Therefore, if the submitter omits this step of selecting group members, they will lose 5% off their score on that component of the project.**

4 Project Status Check-In Report

The project status report is due a few weeks before the final submission, as is intended to make certain that your project is on-track. It should describe what you've accomplished so far and very briefly state what you have left to do.

You should write your status report as if it is an early draft of your final project report. Specifically, you can write it as if you're writing the first few pages of the project report, so that you can re-use most of the text in your final report. Your status report should be at most 2 pages long.

Please write the status report (and final report) keeping in mind that the intended audience is the professors and TAs. (Thus, you should not spend two pages explaining logistic regression.)

Your status report should be in the same L^AT_EX template as your final report (posted on the course website; see the next section for details).

You will be submitting your status report using www.gradescope.com. Log onto gradescope, and submit the PDF files to the CIS 519 assignment entitled "Project Status Report". Only ONE person from each team should submit.

Important: **During this submission process, you must choose your other teammates by name, turning this into a group submission. If the submitter omits this step of selecting group members, they will lose 5% off their score on that component of the project.**

¹http://en.wikipedia.org/wiki/George_H._Heilmeier#Heilmeier.27s_Catechism

5 Final Submission

Your final submission will consist of two deliverables: (1) a final report, and (2) a set of summary slides. Remember that late days cannot be used for the final project submission.

5.1 Final Report

Your final project report can be at most 4 pages long (include all text, appendices, figures, and anything else), with 1 additional page that can contain nothing but references, and must be written in the provided L^AT_EX template. If you did this work in collaboration with someone else, or if someone else (such as another professor) had advised you on this work, your report must fully acknowledge their contributions.

At a minimum your final report must describe the problem/application and motivation, survey related work, discuss your approach, and describe your results/conclusions/impact of your project. It should include enough detail such that someone else can reproduce your approach and results. For inspiration on what should be included, see the project reports available on the links provided in Section 2.1. You will likely end up with a better report if you start by writing a 6-7 page report and then edit it down to 4 pages of well-written and concise prose.

In addition, your report must also include a figure that graphically depicts a major component of your project (e.g., your approach and how it relates to the application, etc.). Such a summary figure makes your paper much more accessible by providing a visual counterpart to the text. Developing such a concise and clear figure can actually be quite time-consuming; I often go through around ten versions before I end up with a good final version.

We know that most students work very hard on the final projects, and so we are careful to give each report sufficient attention. We (specifically, the professors) will personally read every word of every report. After the class, we are also considering posting the final reports online so that you can read about each others work. If are okay with having your final report posted online, be sure to give us explicit permission when you submit, as described below.

5.2 Summary Slides

In addition to the final report, you are also required to prepare a two-slide overview of your project. Think of these slides as a concise presentation of your project, highlighting the problem you worked on, your approach, and your results / contributions.

You may use any format you wish for the slides, but you are limited to only two slides. The goal is not to cram as much as possible into two slides, but to provide a clear and concise presentation of the main points of your project. You should avoid any font smaller than 16 pt, and most of your text should be around 18pt or larger.

The best slides will use lots of graphics along with some text. You are welcome to re-use these graphics in your project report, and you may reuse the summary figure from your report in your slides.

Although this is only two slides, you should be aware that it is actually quite difficult to present an entire project in such a concise manner while still being clear. Do not leave these slides to the last minute; you will likely need to make several versions of these slides until you narrow them down to the essentials, and so they might actually take a while.

5.3 Submission Instructions

Save your report as a PDF file of 5 pages. Save your summary slides as an additional 2 page PDF, and append them to your report, creating a single PDF of 7 pages.

You will be submitting your status report using www.gradescope.com. Log onto gradescope, and submit the PDF files to the CIS 519 assignment entitled “Project Final Report”. Only ONE person from each team should submit the final report and slides.

Important: **During this submission process, you must choose your other teammates by name, turning this into a group submission. If the submitter omits this step of selecting group members, they will lose 5% off their score on that component of the project.**

6 Project Extra Credit (Worth *up to* 6% extra credit on the entire final project)

Create a 4 minute video on your CIS 419/519 project and upload it to Vimeo the deadline.

Your video can be as fancy or as simple as you want (bonus points for creativity!), but it should provide an overview of your project. It should also emphasize the application, and clearly articulate its requirements and the ML issues related to it (since this is a course in applied ML). Everyone in your group should present some aspect of the project in the video. For example, you could have a live-action demonstration, or simply record some audio to present over your project overview slides. You can easily find free software on the web to edit your video, or use something like iMovie.

Make sure that your video is 4 minutes or less! Be creative and showcase your project and in all its machine learning glory!

To turn in your video, please upload it to Vimeo (<https://vimeo.com/>). Then give us the link to your video in this form: https://docs.google.com/a/seas.upenn.edu/forms/d/1ByMk0xbJ_2AEY0GN7qzQK_xw4rPT7GI4Q0EAHynw-g8/viewform. That is all.

Here are more detailed instructions:

- Go to Vimeo.
- Create an account by clicking the green JOIN button.
- You will receive an email with a link to verify your account. You have to verify before you can upload videos. Once you are signed in, click “Upload” at the top of the page.
- Click the “Choose a Video to Upload” button and choose your video
- Once it is uploaded, you can change the privacy settings, if you wish.
- That’s it! Fill in the title and tags, etc. and use the form above to tell us where to find it.

One other note: this extra credit is purely optional and is designed to be a fun value-added activity. We will curve grades without considering extra credit, and so people who elect not to complete any extra credit will not be penalized in any way.