

# Project Guidelines

## COMP 562 (Spring 2019)

Mahmoud Mostapha and Jorge Silva\*

### 1 Overview

One of the primary 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 to explore such an application of machine learning to a problem of your own choice. In job interviews, it is often your course projects that you end up discussing, so the project has some importance even beyond this class. That said, it is better to pick a project that you will be able to go deep with (trying different methods, error analysis, etc.), than choosing a very ambitious project that requires so much setup that you will only have time to try one or two approaches.

#### 1.1 Milestones and deadlines

- Project proposal: no later than April 5, 2019
- Project report: no later than May 7, 2019 (Exam date according to registrar's office), and strongly encouraged to submit by May 1, 2019.

#### 1.2 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 into the problem or/and 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 or/and have an 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?)

Although it is encouraged to implement your project in python using scikit-learn or using Keras, you may use other software or programming languages if you have a particularly compelling reason.

### 2 Choosing a topic

Your first task as a team 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 solve it best.** 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 this class is not intended to prepare you to develop novel learning methods, you may choose to develop a novel learning method (or novel variant) if you want a challenge. Regardless, **most projects will combine**

---

\*adapted from Eric Eaton (UPenn)

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). The techniques used should be relevant to our class, so most likely you will be building a prediction system. A deep learning model would also be acceptable, though we will not be covering these topics until later in the semester.

## 2.1 Project ideas

Many fantastic course projects will come from students choosing either an application that they are interested in or picking some sub-field of machine learning that they want to explore more and working on that topic. If you have been thinking about starting a research project, this project may also provide you an opportunity to do so. Alternatively, if you are already working on a research project that machine learning might apply to, then work out how to apply to learn to it will often make an excellent project topic. Similarly, if you currently work in the industry and have an application on which machine learning might help, that could also do a great project. Also, it is encouraged to engage with other UNC CS faculty for research ideas that they might be willing to be your advisor for the project. I will list some examples of that; however, you are not limited to what is listed here.

### 2.1.1 Ideas from online ML competitions

- Kaggle competitions:<https://www.kaggle.com/competitions>
- A submission to the online competition is not required for this class.

### 2.1.2 Ideas from similar courses at other universities

- Stanford, 2004–2017:<http://cs229.stanford.edu/projects.html>
- M. Balcan, CMU, 2018:<http://www.cs.cmu.edu/~ninamf/courses/401sp18/projects.html>
- D. Rosenberg NYU, 2018:<https://davidrosenberg.github.io/ml2018/#people>
- E. Xing, CMU, 2015:<http://www.cs.cmu.edu/~epxing/Class/10701/project.html>
- A. McGovern, OU:[http://www.cs.ou.edu/~amy/courses/cs5033\\_fall2014/index.html](http://www.cs.ou.edu/~amy/courses/cs5033_fall2014/index.html)
- T. Mitchell, CMU, 2011:<http://www.cs.cmu.edu/~aarti/Class/10601/proj.shtml>
- T. Mitchell, CMU, 2009:[http://www.cs.cmu.edu/~tom/10601\\_sp09/project.html](http://www.cs.cmu.edu/~tom/10601_sp09/project.html)
- C. Guestrin, CMU, 2007:<http://www.cs.cmu.edu/~guestrin/Class/10701/projects.html#datasets>

### 2.1.3 Look through papers from recent machine learning conferences

- 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/>
- Neural Information Processing Systems:<http://papers.nips.cc/>

### 3 Forming a group

Projects must be completed in teams of four students, and it is advised to form groups as soon as possible. You may not complete the project solo or in a team of less than four, unless one or more of your project partners drop the class. You also can work with an advisor from the UNC CS faculty on to address a particular research problem. You are required to select a team member from your group to act as a correspondence member who will be emailing me a list of the team members (and advisor if existing) and will also be responsible for submitting the project proposal, project status report, etc.

### 4 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. The corresponding member of the team should submit the project proposal through Sakai by the announced deadline. 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 plan should demonstrate that you've started to think through the various issues involved with your project and present a compelling argument in support of it. If you are not sure exactly what the proposal should include, contact your advisor or me to discuss that in more details. When writing 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, relevant, and that you can complete it successfully. And, you must do all of that in only one page.

### 5 Project final report

Your final project report can be at **most four pages long** (include all text, appendices, figures, and anything else), with one additional page that can contain nothing but references, and must be written in L<sup>A</sup>T<sub>E</sub>X. 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 explain your results/conclusions/impact of your project.** It should include enough detail such that someone else can reproduce your method and results. **You are also required to provide a link to a GitHub repository where your code is stored.** You may look at previous projects or papers from the list in section 2 to get an idea of what should be included in your project report. 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. Keep in mind if you have an exciting and novel idea for this project, you can extend your work and submit it to an appropriate machine learning conference.

### 6 Final submission instructions

Save your report as a PDF file of 5 pages or less. Again, The corresponding member of the team should submit the final PDF and the GitHub link through Sakai by the announced deadline.