COMP 562 Introduction to Machine Learning
Course Project Guidelines

Mahmoud Mostapha*

September 6, 2018

1 Project 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's often your course projects that you end up discussing, so it 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 (regarding 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 Groups: due Monday, Sep. 10, 2018 11:59pm (no late submissions)
- Project Proposal: due Wednesday, Oct. 3, 2018 11:59pm (no late submissions)
- Project Status Report: due Monday, Nov. 12, 2018 11:59pm (no late submissions)
- Project Report and Poster: due Wednesday, Dec. 5, 2018 11:59pm (no late submissions)

1.2 Grading Breakdown

- Project proposal: 10%
- Project Status Report: 10%
- Final Poster: 10%
- Final Report: 70%

1.3 Evaluation Criteria

- Technical quality (i.e., Does the technical material make sense? Are the things tried reasonably? Dies 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?)

*This document was adapted from Eric Eaton class at UPenn.

- 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 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`
- 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`
- C. Guestrin, CMU, 2007:`http://www.cs.cmu.edu/~guestrin/Class/10701/projects.html#datasets`

2.1.3 Projects ideas from NIRAL and UNC CS faculty

- Early Detection of Autism Using Cortical Surface Measures
  - Autism spectrum disorders (ASD) is a group of lifetime developmental disabilities that are denoted by significant social, communication and behavioral challenges. Early detection of ASD could lead to the reduction or the elimination of the manifestation of the disorder through effective early intervention. The goal of this project is to develop a prediction system for the early diagnosis of ASD in high-risk infants using features extracted along the cortical surface (e.g., cortical thickness and surface area). [1]
    (Advisor: Martin Styner, Ph.D., email:styner@cs.unc.edu)
- Automatic Tissue Segmentation from six months MRI Brain Images
  - Accurate segmentation of infant brain magnetic resonance images (MRI) into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) is a key step in studying both normal and abnormal early brain development. In the isointense phase (6-8 months), the intensity range of voxels in GM and WM are largely overlapping, thus leading to the lowest tissue contrast and creating the most significant challenge for tissue segmentation. The goal of this project is to develop a deep learning framework (e.g., U-Net architecture [2]) for the accurate tissue segmentation of isointense infant brain MR images [3].
    (Advisor: Martin Styner, Ph.D., email:styner@cs.unc.edu)
- Analysis of EEG Signals using Deep Learning Techniques
  - Electro-Encephalogram (EEG) signals can be used to help identify deficits in brain communication, either through the power or synchrony of neuronal signals. We are currently looking at several different populations and are interested in identifying ways in which their oscillations or oscillatory patterns differ. To do this, we record 64 EEG channels at 1000Hz over several different tasks. We can then look at the channel or source level EEG activity elicited in response to those tasks. Some questions we may ask are: 1) Does activity vary between two tasks? 2) Does activity vary between the two groups? 3) Can we predict task or group based on observed activity? 4) Which features of the EEG signal make the best predictor for group or task? The objective would be to use deep learning techniques to filter signals and find a correlation between a given task and brain activity. Deep Canonical Correlation Analysis would help accomplish this objective [4].
    (Advisor: Juan Prieto, M.Sc., email:adkuntz@cs.unc.edu)
- Fast Surgical Robot Shape Computation using Machine Learning
  - Concentric tube robots are tentacle-like surgical robots capable of taking complex curved shapes in the human body to enable minimally invasive surgery. Computing their shape ahead of time is necessary for planning the motion of these robots, but is very computationally expensive. Using machine learning techniques, we believe shape computation that is faster than current models is possible. The ability to compute the shape of these robots more quickly would enable better and faster motion plans for these surgical robots.
    (Advisor: Alan David Kuntz, Ph.D., email:juanprietob@gmail.com)

2.1.4 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 Project 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. A LaTeX template for the project proposal is available on the course website and Sakai. 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 Status Report

The project status report is due approximately one month before the final submission, as is intended to make sure 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 two pages long. Please write the status report (and final report) keeping in mind that the intended experienced ML audience (i.e., you should not spend two pages explaining logistic regression.)

Your status report should be in the same LaTeX template as your final report (posted on the course website and Sakai; see the next section for details). Again, The corresponding member of the team should submit the project status report through Sakai by the announced deadline.

6 Project Final Report and Poster

Your final submission will consist of two deliverables: (1) a final project report, and (2) a set of Poster.

6.1 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 the provided LaTeX 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 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, we can extend your work and submit it to an appropriate machine learning conference.

Also, your report must also include a figure that graphically summarizes the main components of your project (e.g., your approach and how it relates to the application, etc.). Such a figure makes your paper much more accessible by providing a visual counterpart to the text. Developing such a short and clear figure can be quite time-consuming; so plan wisely and feel free to consult your advisor during this process or me. I will be reviewing your papers and give you my feedback so you can adjust some of its contents and send back a revised paper that will be posted online on the course website after the class ends. It is preferred to display your work online so other people can read about each other's work. If you are okay with having your final report posted online, be sure you give us explicit permission when you submit, as described below.

6.2 Project Poster

In addition to the final report, you are also required to prepare a poster overview of your project. Think of this poster as a concise presentation of your project, highlighting the problem you worked on, your approach, and your results/contributions. This poster will also be used to present your work to the class on the last day of the course in a poster session. A poster template will be posted for you to use before the poster session date.

The goal is not to cram as much as possible into your poster, but to provide a clear and concise presentation of the main points of your project. The best posters 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 poster. You should be aware that it is quite difficult to present an entire project in such a concise manner while still being clear. Do not leave the poster to the last minute; you will likely need to make several versions of the poster until you narrow them down to the essentials, and so they might take a while.

6.3 Final Submission Instructions

Save your report as a PDF file of 5 pages or less. Save your Poster as an additional 1 page PDF or less, and append them to your report, creating a single PDF of 6 pages or less. Again, The corresponding member of the team should submit the final PDF through Sakai by the announced deadline. Please make sure to state that your team is okay with posting your work online.

References

[1] Heather Cody Hazlett, Hongbin Gu, Brent C Munsell, Sun Hyung Kim, Martin Styner, Jason J Wolff, Jed T Elison, Meghan R Swanson, Hongtu Zhu, Kelly N Botteron, et al. Early brain development in infants at high risk for autism spectrum disorder. *Nature*, 542(7641):348, 2017.

[2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[3] Miccai grand challenge on 6-month infant brain mri segmentation. `http://iseg2017.web.unc.edu/`.

[4] Galen Andrew, Raman Arora, Jeff Bilmes, and Karen Livescu. Deep canonical correlation analysis. In *International Conference on Machine Learning*, pages 1247–1255, 2013.