Final Project Instructions
CMPSCI 689 - Fall 2017

Project Proposal: Due Oct 25th, 2017 by 11:59pm. 5% of final grade.
Final Report: Due Dec 19, 2017 by 11:59pm. 35% of final grade.

Late days cannot be used for either the project proposal or the final report.

General Guidelines:

We expect to see a creative, well-researched, and well-executed project. Projects should be one of three different types:

1. **Implementation projects** should select a recent and non-trivial machine learning model and re-implement existing learning, prediction, and/or inference algorithms in a novel computational context (i.e., on a smart phone, wearable device, embedded microcontroller, within a browser, in a multi-core/parallel/distributed computing environment, or in a language where existing implementations do not exist).
2. **Algorithms projects** should select an existing machine learning model and investigate new learning, prediction, and/or inference algorithms for it based on different optimization or approximation methods. The proposed algorithms should be well-motivated according to some criteria (i.e., potential for reduced computational complexity, improved speed-accuracy trade-offs, etc.).
3. **Modeling projects** should start from an existing machine learning model and propose extensions or modifications that are well-motivated by the needs of a particular application (i.e., dealing with different types of outputs, handling missing data, better representing latent structures, improved regularization, etc.). Projects of this type will also need to develop learning, prediction, and/or inference algorithms and are better suited for advanced students or teams.

All projects will require implementing models and algorithms and conducting appropriate experiments to validate what you develop and to compare it to existing baseline and state-of-the-art methods (where existing state-of-the-art implementations are available). This means you will need access to appropriate data sets and need to take your computational resources into account when choosing a topic. All projects require writing up your results in conference paper format along with a related work section. Note that straight application projects do not meet the above requirements.

Projects can be completed in groups of up to three students. The expectation is that groups of 2 or 3 will do 2 or 3 times more work than a single student could accomplish on their own. Students that are involved in research can tie this project to their existing work, but all of the work done for the course project must be new, and whatever you submit for the course project must be entirely your (or your group's) own work. The project proposal requires that you declare any overlap with your current or past projects.

Specific Requirements for Project Proposals:

Your project proposal is a short description of what you plan to work on for your final course project. You should read the specific requirements for the final project report before starting to draft your proposal. You may change your mind about some of the details as you go, but you should contact us if you decide to completely change your project topic. The maximum page limit for the proposal is **1 page** for the proposal, **1 page** to explain any overlap between this project and any other past or existing projects you or your team are involved in, a **1 page** for collaboration plan for groups of two or more, and **1 page** for references. You will submit your proposal on Gradescope. Students working in groups should submit their proposal as a group on Gradescope.

Your proposal should include:

1. **Title:** Select an informative title for your project
2. **Author(s):** List the names of all group members (your name if working alone).
3. **Problem:** Provide a clear description of what novel problem your project will address, and the motivation for addressing this problem. Include context information to situate your project relative to the machine learning literature.
4. **Methodology:** A sketch of the methodology you plan to apply including what mathematical or optimization frameworks you will leverage, what code libraries you will use or implement, and/or what hardware platform you will target.
5. **Related work:** Briefly describe the most closely related work within the machine learning literature to what you are proposing to study.
6. **Data Sets:** A brief description of the data sets you plan to use in your experiments including a link to the data sets if possible.
7. **Experiments:** A description of what experiments you will perform to validate your approach, study its properties, and compare to other existing alternatives.
8. **Overlap Statement:** If your project will build on past work performed by you or your group, or if it overlaps with other current work you or your group are doing for a job, personal project, research project, or in another course, explain any overlap. If your project does not overlap with past or current work at all, please state that explicitly.
9. **Collaboration Plan:** For groups of 2 or 3, provide an additional summary of which group members will do what work on the project (up to one additional page). There must be a clear separation of tasks among group members and the work must be equally distributed across the group.
10. **References:** Provide a list of references to support your assessment of related work. You can provide up to one additional page of references. References should be from the primary machine learning literature, and should be listed in a standard format.

Grading Notes: For full points, you must provide all requested information. You **may not** indicate that you will make decisions about items like data sets or experiments later. You must provide an initial plan that you can refine as you go. This is to ensure that course staff can provide proper feedback.

Specific Requirements for Final Projects:

Your final project should follow standard machine learning paper structure including the following sections. The number of pages per section listed below should be taken as a rough guide, but there is a firm upper limit of **8/16/20 pages** excluding references for groups of 1/2/3 students. Your report should be prepared in NIPS conference format (see https://nips.cc/Conferences/2014/PaperInformation/StyleFiles). Students working in groups should submit their proposal as a group on Gradescope.

1. **Title:** Select an informative title for your project.
2. **Author(s):** List the names of all group members (your name if working alone).
3. **Abstract (100-200 words):** A brief summary of your problem, approach, and results.
4. **Introduction (0.5-1 pages):** An introduction describing the problem your project is solving, a discussion of the motivation for selecting this problem, a brief statement about what is novel in your work relative to the related literature, and a summary of your methodology, experiments, and results.
5. **Related Work (1-2 pages):** A related work section describing at least 5 pieces of prior research from the primary machine learning literature related to the problem your project addresses. Your selections should include a mix of both well-cited and more recent articles when possible. Do your best to identify the current state-of-the-art methods related to your work. You should provide mathematical and algorithm details for at least one model that you will compare to in your experiments. You can use the machine learning literature search tool here to identify relevant articles: https://people.cs.umass.edu/~marlin/mlls.html to help identify articles.
6. **Methodology (1-6 pages):**
   - For implementation projects, this section will describe the models and/or algorithms that you decided to implement including mathematical descriptions, a description of the hardware/software platform or programming language you decided to implement the models and/or algorithms in, your design methodology or implementation architecture, and any libraries or existing resources you leveraged for your implementation. Implementation projects should be well-motivated.
   - For algorithms projects, this section should describe the model that you will work with including mathematical descriptions. It should include derivations and/or proofs for the algorithms you are proposing, pseudo-code descriptions, and any required implementation-level details.
   - For models projects, this section should describe your proposed model including a mathematical description (and where possible, a graphical depiction), as well as derivations for the associated algorithms that you are proposing for learning, inference, and/or prediction.
7. **Data Set(s) (0.5-1 Pages):** Describe your data including where it was obtained from. Describe the number of data cases, the number of features, what the features represent and what their data types are, etc. For data sets with large numbers of features, you should provide a summary of the features and not an exhaustive list-

ing (include a reference to a published paper or website that describe the data in more detail for large data sets if possible). Explain why you picked each data set.

8. **Experiments (2-3 Pages):** You must perform experiments to validate your results. Implementation projects should validate the correctness of the proposed implementations against existing reference implementations, and study performance characteristics like learning and prediction times, for example. Algorithms projects should compare aspects of performance (accuracy, convergence time, speed-accuracy trade-offs, etc.) relative to existing baseline and state-of-the-art algorithms. Modeling projects should conduct experiments similar to algorithms projects. This section needs to clearly describe the experiments you conducted including what data sets you used, how they were processed, what hyper-parameter settings you investigated, how free hyper-parameters were selected, what performance metrics you used, etc. You should also discuss the rationale for performing the selected set of experiments and what your hypotheses were for the outcomes.
9. **Results (2-5 Pages):** In this section, you will describe and discuss the results of your experiments using suitable figures and tables (your report must contain at least 2/4/6 results figures or tables for groups of 1/2/3).
10. **Discussion and Conclusions (0.5-1 Pages):** In this section, you will discuss the overall outcomes of your project. How do your results relate to what has been reported in the literature previously? What seemed to work well and what didn't? Did you run into any particular problems? What else would you have done if you had more time? What should others learn from your results?
11. **References:** Provide a list of references to support your assessment of related work. You can provide up to one additional page of references.

Grading Notes: In addition to the technical requirements for final projects, 10% of the grade will be allocated to clarity. This covers items including clarity of exposition, providing clearly legible figures and tables (at print resolution), and including appropriate labels, titles, and legends on figures and tables.

General Project Advice:

• **Getting started.** One approach to get started is to identify an application you are interested in working on. Begin by looking at recent machine learning papers addressing this application using the provided literature search tool. Look for places where you could try modifications or extensions of existing models or algorithms to solve a problem, remove a limitation, extend functionality, etc. Many papers will include a "future directions" section at the end detailing next steps that could be taken to build on the paper.

• **Be selective.** Make sure you have a solid motivation for looking at a particular problem before you start thinking about solutions. Don't attempt to address a problem that is irrelevant, ill-defined or unsolvable. Don't pick a problem that has already been clearly solved in the literature. Don't pick a problem where you don't have access to the data or computation needed to validate your solution.

• **Be modest.** For implementation projects, it's probably not a good idea to target a platform you have no experience with. For model and algorithms projects, start with a small, but novel twist on something that exists. It will take you longer than you think to write the code and run the experiments to assess the usefulness of your modifications.

• **Be careful.** Don't make basic mistakes like testing on your training data, setting parameters by cheating, comparing unfairly against other methods, using undefined symbols in equations, etc. Do sensible crosschecks like running your algorithms several times and making sure it gives sensible outputs on simple cases.

• **Be honest.** You are not being marked on how good your experimental results are. What matters is that you try something creative (but sensible), clearly describe the problem, your method and experiments, and what the results were.

• **Use time well.** Don't wait until the end of the semester to start seriously working on your project. Don't underestimate the time needed to understand new methods, get code running, conduct experiments, produce figures and tables, and write a clear report.

• **Choose something you can build on.** MS/PhD students should consider projects that can be tied back into their research when complete. MS/PhD students in AI areas in particular should be looking to tie their project into a future publication.

• **Have fun!** If you pick something you think is cool, that will make getting it to work more interesting and you'll be more excited to write-up your results.