# CS74/174, Winter 2015
# Machine Learning and Statistical Data Analysis
# Term Project

January 5, 2015

## 1 Important Dates

- January 22, 2015: project proposal due
- February 17, 2015: project milestone due
- March 10, 2015: final project presentation
- March 15, 2015: final write-up due

## 2 General Guidelines

The goal of the project is to allow you to explore in some depth a topic of your interest within the area of machine learning. The project can be done in teams of two or three students. Individual projects are highly discouraged. Please feel free to use the forum that we have set up in Blackboard to find project partners and to discuss ideas. There are two main types of project you can do:

- **Application project**: in this type of project you apply machine learning methods to a problem that interests you. Keep in mind that not all problems are suitable to be addressed using a learning approach. Think of an application problem which cannot be solved well by hand programmed solutions. Furthermore, examples must be available to train a learning system. Don't pick a task that is too hard. Try to come up with a creative application of machine learning to a feasible problem. Once you have defined the task, research the machine learning literature in order to find an interesting approach that can be used to solve the chosen problem. If you are unsure about the method to use, talk to the instructor.
- **Algorithm project**: in this type of project you design a novel variant of an existing learning algorithm. Pick a learning technique that you like, study it thoroughly, identify its shortcomings, and attempt to address these limitations using a new principled strategy. Be aware: this is a challenging kind of project. Chances are that the most straightforward variations of common algorithms have already been investigated. Thus, it is quite important that you complete an accurate literature review in order to understand if the idea you are proposing is truly new. You can evaluate your algorithm using standard machine learning benchmarks, such as the data sets available at the UC Irvine Machine Learning Repository. Your work will

be graded primarily on the basis of how well you motivate your proposed solution: run experiments illustrating the shortcomings of the existing method and suggesting the benefits of your proposed modification. You won't be penalized if your proposed method does not yield the expected results as long as you analyze and describe clearly the reasons of the failure.

You are allowed to use external software for *parts* of your work. The project grade will be based on the novelty of your solution/application but also on the amount of new code written by you to implement the idea. Thus, using external software written entirely by someone else would not be acceptable. Similarly, merely applying code developed for the homework assignments to a new set of examples is not sufficient. Keep in mind that you will implement pretty much all the algorithms listed on the course schedule page for the homework assignments, so you should select some method outside this list. If you use external software, please report clearly which code is your own and which one was taken from external sources.

If you have difficulty coming up with an idea for the project, come to the office hours of the instructor: we will brainstorm together and attempt to find possible projects related to your specific interests. For inspiration you may also want to look at the proceedings of ICML and NIPS, the two main machine learning conferences. It is also instructive to look at last year's projects, which are listed on the course web page.

## 3 Project submission

The project submission involves three distinct components: proposal, milestone, and final. For each of these you will need to prepare a web document. We suggest that you include all the files in a folder, and then submit the entire compressed folder via blackboard. Please use relative links in your HTML documents, as your web pages will be copied to our server and made available off the course webpage. This will allow other CS74/174 students to look at the description of your project and learn from it.

- **Proposal**: the write-up of your project proposal should provide a detailed description of the work that you intend to complete. This document should address the following questions:

  1. What problem do you want to solve? Describe at a high level the application and the learning problem involved.
  2. What are suitable methods for this problem? Research the machine learning literature and find a set of techniques that may be applied to solve this task. Include references to these methods in your proposal.
  3. What data sets do you plan to use? Include pointers to training data that you will use in your project.
  4. What do you expect to accomplish by the milestone due date?

  Note that while item 1. is binding (i.e., you cannot change your project topic after the proposal), items 2. and 3. are *not*: if after submitting the proposal you find new algorithms or databases that are more suitable to your task, you are free to use them without penalty. However, I want to see from your proposal that you have thought carefully about what you want to accomplish and how to do it.
  *Spotlight presentation*: the same day when the proposal is due, you will give an in-class brief

presentation of your project proposal. You will have **two minutes** to present **one slide** describing your project plan. Be aware that summarizing a project idea in such a short amount of time is not an easy task. Please prepare well for this: rehearse your presentation a few times in order to make sure you can provide an overview of your project in the alloted two minutes. This time limit will be strictly enforced and if you don't prepare carefully there is the risk that your presentation will be cut short. Email your slide to us by 5 pm the day before your presentation. Late submission will not be accepted.
*Proposal write-up and spotlight presentation will count equally, and contribute in total to 15% of your project grade.*

- **Milestone**: the milestone submission is a web document providing an account of the accomplishments you have achieved so far. The purpose of the milestone submission is to encourage you to start working early toward your project goals. It is crucial that you make good progress early in order to complete a successful project. We expect the milestone submission to not differ much from the final one, as this would indicate you are approaching the completion of your project. You should comment honestly on whether you are on-track and if you have reached the milestone goals that you had set in the proposal. Please provide a brief description of the algorithms you have implemented and include visualizations of the results already obtained. It is not necessary to submit code.
  *Milestone presentation*: you will have 5 minutes to present an overview of what you have accomplished up to this date. You should view the milestone presentation as an opportunity to gather feedback that may help you resolve issues that have impeded the progress of your project. For this purpose, please prepare a few slides describing in detail what you have implemented, the experiments that you have run, what has worked and what has not. Discuss also what you still need to do in order to complete the project.
  *Milestone write-up and presentation will count equally, and contribute in total to 25% of your project grade.*

- **Final**: the final write-up should be an expanded and polished version of the milestone submission. You should specify all the algorithms that you have implemented, even those that possibly have not worked well. You should write your document for a technical audience. Therefore, do not write lengthy descriptions of basic learning methods. However, you do need to cite the articles related to all the algorithms that you have implemented. Furthermore, if you have made modifications to the methods, please discuss your variations in detail and provide a justification for them. The most important part of the final project report is the discussion of the results: include a thorough description of your experiments and use well captioned figures and plots to summarize your results.
  *Poster presentation*: there will be a poster presentation of the final projects at the last lecture. We will provide easels and boards to display the posters. You should be prepared to give a brief (3-4 minutes) presentation of your work, using the poster as an aid to explain your project. We will review each poster carefully and ask questions related to the work. We will also invite other faculty members to come to the presentation.
  *Final write-up and poster presentation will count equally and, together, they will be worth 60% of your project grade.*

We are very much looking forward to seeing your project. Have fun and good luck!