

Math 251: Statistical and Machine Learning Classification

Department of Mathematics and Statistics
San Jose State University

September 12, 2018

Final project: Description and instructions

This course ends with a project that is 30% of your grade and aims to provide you with the culminating experience and lots of additional learning opportunities.

The topic and content of the final project is to be determined between you and the instructor. Below are the specific instructions about the final project (please read them carefully!):

- No groups will be allowed, that is, each of you needs to select and work on a different project. And you cannot simply copy any of the projects from the old Math 285 course.
- You will need to submit a **1-page proposal** that describes what you intend to do in your project, and you must get the approval from the instructor before you start. To maximize the chance of your proposal getting approved, it should be as clear as possible and provide all necessary information for evaluation, such as:
 - The title of your project;
 - Description of the problem
 - Your proposed work with clearly stated goals
 - Data sets you plan to use with urls
 - Reference papers that are relevant to your project;
 - What is significant in your proposed work, e.g., potential applications.

Your proposal is due **Wednesday, October 31**, but you should select a project as early as possible, because projects will be available on a first-propose, first-get basis and you also need enough time to complete your project. Your proposal will be referred to when grading your project presentation and report.

- You will be asked to report your results through a **7-minute oral presentation** in class and meanwhile submit a **report of 5+ pages** that has all the details.
 - Your presentation needs to present a general level summary of your work (i.e., focus on the main ideas but you should still give all necessary specifics, like parameter setting, etc.). It should be clearly structured, logical, and self-sustained. It is advised that you prepare 7 to 10 slides (but not more). We will reserve the last class and also the scheduled final exam time for your presentations.
 - Your report has to be written using your own language (copying from other places is strictly prohibited and will be given a zero point). In addition, it needs to contain a clear structure with the following components: Title, Author, Abstract, Introduction, Your proposed method or study, Experiments, Conclusions (or Discussions), and References. Your report will be due on the scheduled final exam day (December 12, 2018).
 - Your presentation and report will be graded based on clarity, completeness, correctness and originality.
- **Examples of a good project** for this course are the following (not an exhaustive list):
 - **Introducing a new classification algorithm** that is not covered in class. You must describe the new method clearly (with sufficient detail) and demonstrate it on both toy and real data.
 - **Nontrivial improvement of an algorithm** learned in this course. You must demonstrate the performance of your implementation on several data sets, and compare with the old implementation.
 - **A nontrivial application** of a method learned in this course to a large, interesting data set with explorations of different options and parameter settings (you may refer to the course webpage for some well-known databases, or simply google to find your own data set).
 - **An empirical study of several algorithms** using a few data sets to study their strength and weakness and compare their performance.
 - **Proving a nontrivial theoretical result** that has not been done in class.

I hope you will learn a lot through the projects and obtain more experience with real data. Feel free to discuss with me any questions you have about the project.