Final project: Description and instructions

Math 285 Selected Topics in High Dimensional Data Modeling
San José State University
Fall 2015

This course ends with a project that is 40% of your grade, to provide you with the culminating experience. The topic and content of the final project is to be determined between you and the instructor. Below is a description of how the project runs (please read them carefully!):

- The class will be divided into groups of size 1 or 2 (but not more); that is, you may choose to **work on your own or with a partner** that you find from this class.
- Each group needs to **submit a 1-page proposal** that describes what you intend to do in your project, and must get the approval from the instructor. To maximize the chance of your proposal getting approved, it should be as clear as possible while providing all necessary information for evaluation, such as:
  - Who is/are in your group;
  - The title of your project;
  - What problem you intend to work on;
  - Any existing algorithms or known results;
  - The main reference paper(s) your project will be based on;
  - What is new/significant in your proposed work;
  - What are the applications of your proposed work;
  - How you plan to report your findings (see next bullet);
  - Any other relevant information.

  Your proposal is due **Tuesday, November 24**, but you should select a project as early as possible, because projects will be available on a first-propose, first-get basis and you need enough time to work on your project. Your proposal will be graded together with your final presentation or report (see below).
- You may report your results through a **15-minute presentation in class** or in **a report that is at least 5 pages long** (in standard format, e.g., regular font size and margins):
  - If you are the only person in the group, you can choose either of the two options. However, if you work with a partner, then your group needs to to do both (i.e., one person presents and the other person writes).
  - We will reserve the last few classes and also the final exam day (Thursday, December 10, 0715-0930) for your in-class presentations (depending on how many groups select this option in the end).
  - Your report has to be written using your own language (copying from other papers is strictly prohibited and will be given zero point). In addition, it needs to contain a clear structure with the following parts: Title, Author(s), Abstract, Introduction, Your proposed method (or review of a previous method), Experiments, Conclusions, and References. You may refer to

the paper at `http://w.american.edu/cas/sampta/papers/a138-chen.pdf` for structure and formatting.
  - Your presentation or/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 method** that was not covered in class, such as Locally Linear Embedding (LLE) and Independent Component Analysis (ICA) for dimensionality reduction. You must describe the method clearly (with sufficient detail) and demonstrate it on both toy and real data.
  - **Nontrivial improvement of an algorithm** learned in this course, e.g., how to select the number of clusters in spectral clustering (for a large range of data sets), how to speed up the nearest neighbor search used in kNN-based methods. You must demonstrate the performance of your implementation on several data sets, possibly comparing with the old implementation.
  - **A novel and nontrivial application** of a method learned in this course to a (large!) real, interesting data set (you may refer to the course webpage for some well-known databases, or simply google to find your own data set). An example is successful application of spectral clustering to large social network data.
  - **An empirical study of several algorithms** using a few data sets to study their strength and weakness and compare their performance. A specific example would be to compare different dimensionality reduction algorithms on the handwritten digits.
  - **Proving a nontrivial theoretical result** that has not been done in class or directly available in a paper.

I hope you will learn a lot though the projects (and in-class presentations) and obtain more experience with real data. Feel free to discuss with me any questions you have about the project.