

# Project\*

## for High-Dimensional Data Analysis

### 1 Overview

Your project will consist of five elements.

- Project Proposal : Due October 16
- Project Milestone Check-in : Due November 15
- Project Written Report : Due November 25
- Project Presentation : November 25 - Dec 4

The course project will be an exploration of techniques for managing very high-dimensional data. It will consist of

1. identifying an interesting analysis task involving high-dimensional data;
2. showing some baseline approach incurs a challenge that is hard due to, or increasing with, dimension;
3. proposing a mechanism to manage the dimension-induced difficulties; and
4. demonstrating advantages of your proposed mechanisms.

In the final report and presentation, all 4 steps should be identified and addressed. Each can be step should be supported by evidence. The interestingness can be supported with reductions and citations. The mechanism should be described in enough detail that it could be reproduced. The baseline challenges and proposed mechanism advantages could be supported with mathematics, empirical investigations, or preferably both.

There must be some *novelty* in the project. The novelty could come in the newness of the proposed mechanism, but that is not the only place it could come. You could use an existing mechanism, but apply it to a new problem, or demonstrate its advantage in a new way (e.g., a new form of empirical evaluation). The novelty should be commented on explicitly.

This is an individual project. You can propose and make a case for a team that builds on a common shared infrastructure, but this needs to be approved by the instructor, and all components will be graded individually.

### 2 Project Proposal (10 points)

#### Due October 16

Prepare a **100 to 200 word** outline of your plan. The submission is as a Canvas quiz. It should cover your plan for each of the 4 steps above. It should be clear where some plan for novelty lies.

The proposal can be resubmitted. It is not complete until it is approved by the instructor.

Once the proposal is approved, then a student can sign up for a presentation slot.

Talking to the instructor after class or in office hours is a great way to get ideas.

### 3 Project Milestone Check-in (10 points)

#### Due November 25

This milestone should demonstrate *substantial* progress on the project. This can be materialized in several ways including:

- extensive data collection and preparation
- data transformation into high dimensions
- early experiments
- a lemma / important formulation towards mathematical result

The milestone can be accomplished by either:

- meeting with the TA (for say 5-10 minutes) to discuss your progress [**This is preferred**]
- a roughly 1 page pdf outlining your progress uploaded to canvas.

### 4 Project Report (45 points)

#### Due November 25

Student presenting in the first week (Nov 25 and Nov 27) are permitted to submit a full draft by Nov 25, and submit a final draft the next week on Dec 2, with no penalty.

Students should submit a report that is at most 4 pages of core content (at least 1 inch margins, 11 point font). There can be an unlimited number of appendix which the instructor can read at their discretion. Any critical figures or plots should go in the core 4 pages.

The content should cover, and will be graded on, how well it addresses:

1. identifies an interesting analysis task involving high-dimensional data;
2. provides and describes a baseline approach which incurs a challenge that is hard due to, or increasing with, dimension;
3. proposes and describes a mechanism to manage the dimension-induced difficulties;
4. demonstrates advantages of your proposed mechanisms; and
5. has a novel aspect.

Each aspect should be clearly labeled and easy to find in grading. They should also be backed up with evidence and work accomplished in the project.

### 5 Project Presentation (35 points)

#### Due November 25, 27, December 2 or 4

Each student will get a 10 minute presentation slot. They should aim for a clear, precise, and interesting/fun overview of their project. It should hit on the same key aspects as outlined in the report.