ISyE 6740 – Computational Statistics – Fall 2020

Project Guideline

Each project is done by a team of **1-3 students**. Please form group, pick a topic and design the methodology all by yourself. You are welcome to discuss your project ideas with the professor and TAs during the office hours. Follow suggestions below:

Step 1: identify a problem that is related to our class (data analysis, machine learning related)

It can be analyzing some real data, such as:

(1) Predict used oil price by several factors; fitting the model using your own method.
(2) Test whether or not there is a difference in the daily coffee consumption by engineering student versus non-engineering student; find the confidence interval by bootstrap.
(3) Build some classification methods to predict presidential election outcome.
(4) Finding the most important factor that leads to two people become friends on facebook, using variable selection method.
(5) Find whether or not there has been an increasing trend in crimes in local neighborhood.

It can also be anything from your current research project. The dataset you use does not necessarily have to be really large. But you should come up with a novel way of using data.

Alternatively, you may also identify a theoretical problem and solve it (such proving convergence, consistency, etc.)

Step 2: Go ahead to collect data yourself, find data from Internet by crawling data, or identify an existing dataset from some source.

If you project is purely theoretical and does not involve any data component, you do not need to find data. However, we would encourage even simple example of your algorithm/theory applied on real data to demonstrate that your algorithm/theory is working.

Step 3: analyze data and fit models using methods we did in class (or even outside of our class). Careful design an evaluation method. Interpret your results, whether or not it makes sense or are according to what you expect. Do you have a large enough sample size? Is your finding statistically significant? Do you have a representative sample or its biased? How would you improve your result? Can you use it for anything real?

Evaluation:

If you have a team, please clearly specify in final report what part/tasks each team member is responsible for. Make sure to (at least) include the aspects required in the final report. Professor will evaluate the project according to the metric of evaluating an academic paper: whether or

not it is interesting (novel), carefully design, and rigorously implemented and carefully evaluated.

The final report should have minimum 4 pages, **using the word template provided** (if you perfer to use Latex, please make sure follow a similar format in terms of font size, page margin, using single-spacing and single column.

Please submit the final report though Canvas.

Grading: The project will be graded by two different TAs (randomly assigned, may be different from your TA group assignment). Each TA will give a rating and the grade is the average of the two ratings. One grade is given for each project (all team members receive the same grade).

The project will be evaluated according to the following metric:
- **20%** on creativity and innovation (make it very clear in your writing why you do this, why the problem is relevant, for example, it is not yet-another-classification-of-handwritten-digits project).
- **40%** on implementation (the methods is reasonable, the results are reasonable).
- **40%** on writing (writing clearly, write like an academic paper as much as possible).

And finally, be creative, be interesting, and be bold.