Tsinghua-Berkeley Shenzhen Institute
Learning from Data
Fall 2019

Final Project

---

1 Project Information

- **Group:** The final project will be done in groups of three. Please tell TA your group composition before **Nov. 24th**. If you can not form a group of three, a group of two will be OK too.
- **Deadline of your report: Dec. 31st, 2019**
  Each group should submit a written report in a single **PDF** document with at least **4 page**. All related codes should also be provided.
- **Presentation:** After you submit your report, the group that has done excellent work will be invited to present in class.
- **Data:** The instructor will provide a list of suggested datasets for students to chose from, but students are encouraged to find their own dataset or topic.
- **How will your work be evaluated:**
  1. Each team should submit by Dec 1st a project proposal (at most 2 pages) (10 points for all on-time submissions)
  2. Each team should make an appointment with the TA or the instructor and go over the project proposal before Dec 6th (10 points for showing up)
  3. Project report (80 points)
  4. The top 12 teams will be able to present in class. Up to 10 extra credits can be given to the top three teams for excellent contribution and good presentation. All students who attend the presentation sessions will receive 2 extra credits.
- **What should be in your project proposal:**

  + Motivation: What problem are you tackling? Is this an application or a theoretical result?

  + Method: What machine learning algorithms and techniques are you planning to apply or improve upon?

  + Intended experiments: What experiments are you planning to run? How do you plan to evaluate your algorithm?

  + A brief literature review for related works.

2 Possible topics

The choice of topic for the class project is up to you so long as it clearly pertains to the course material. It is fine to select a topic that is related to your area of research.

- **Apply a machine learning method to solve a specific problem**
  + A machine learning approach to classifying your incoming mail
  + Predict stock prices based on past price variation
  + Predict how people would rate movies, books, etc.
  + Cluster gene expression data, how to modify existing methods to solve the problem better

- **Develop a machine learning method that has better performance**
  + Algorithm with higher stability
  + Machine learning methods for genomic data, are they effective, what is missing
  + Neural network with higher image recognition capability

- **Theoretical problems**
  + Generalization guarantees for a specific algorithm
  + Convergence/consistency of a specific estimation method

3 Report Guideline

- **A good report should cover such parts:**
  + Abstract
  + Introduction and related work
  + Dataset and Features (if applicable)
  + Methods
  + Experiments/Results/Discussion
  + Conclusion/Future Work
  + References/Bibliography

  Please refer to Stanford CS229 file `http://cs229.stanford.edu/final-report-guidelines.pdf` for more information.

- Your report should be in good format. It is highly recommended that you can organize your work by LaTeX.