C S 487/519 Applied Machine Learning I
Open Machine Learning Project

1 Objective

In this *team* project homework, you are required to form a team with 3-4 students and apply machine learning knowledge to solve a real-world problem.

2 Requirements

Design a problem that comes from real-world applications and design solutions to solve the problem. The problem can be totally new (from your own investigation) or from other sources (e.g., a Kaggle competition problem, a problem defined in a recent research article). The problem can also be part of a problem that you are currently working on (e.g., towards your PhD dissertation, or Master's project, or Master's thesis).

- **Motivation**: Please clearly describe the applications/motivations of the problem.
- **Problem**: Clearly define the problem.
- **Solution**: Design reasonable solutions to solve the problem by utilizing the machine learning knowledge that you have learned and making use of other related machine learning tools.
- **Data**: Obtain proper datasets to test your solution. You can use self-created datasets or publicly available datasets. If you utilize existing datasets, they had better be reasonably big in size (e.g., with more than 10K instances). If you create your own datasets (e.g., by writing script to crawl data), your dataset may not be huge in size (e.g., with several hundreds of instances).
- **Analysis**: Properly analyze the performance of your solution.

3 Submission instructions

ONE member from the team can be representative to submit the project.

Stages.

- **Stage 1 - Team Formation**: Form a team. Please submit **team.txt** file with student names on Canvas.
- **Stage 2 - Project Proposal**: Formulate your problem. Submit your report via Canvas with the motivations and problem definitions. Note that this report can be updated in later stages.
- **Stage 3 - Midterm Report**: Add your proposed solution to your report, continue to refine your motivations and problem definition, and update your report; Start to write code to do basic analysis or to get the data. Continue to refine your report; finish half of the code (approximately); have some preliminary results from at least one dataset. In **stage3**, submit the updated report, code, small datasets (if created by you), or links to your large datasets on Canvas.

- **Stage 4 - Final Report**: Finish your solution and your algorithm analysis; finish your report. Submit your final product on Canvas with the following information:

  - team.txt
  - A readme file readme.txt with (1) how the code base is organized, (2) the commands to run your code, and (3) data set information: if you use existing datasets to test your solutions, the readme file should include the link to the existing datasets.
  - Code base containing your code for crawling the data (if there is any), preprocessing the data, solving the problem, analyzing your solutions, plotting analysis figures, etc.
    - Your Python code should be written for Python version 3.5.2 or higher.
    - Please properly organize your Python code (e.g., create proper classes, modules).
    - If your team created your own datasets, put the datasets in the repository.
  - A 3-5 page report, **report.pdf**, to include the above content (motivation, problem definition, solution explanation, data description, result analysis).

Format requirement of the report.

- For LaTeX users: please use sample-sigconf.tex from the ACM article template (https://www.acm.org/binaries/content/assets/publications/consolidated-tex-template/acmart-master.zip); Additional information about formatting and style files is available online at: https://www.acm.org/publications/proceedings-template.

- For word users: Margin at each of the top, bottom, left, and right sides is 1.0 inch; double columns; The font size is 10pt; font type is Times New Roman. Single line space.