BNU-HKBU United International College
DS 4023: Machine Learning
Spring 2020 Course Project

Description:
This is an **INDIVIDUAL** project, which aims at applying machine learning algorithms (including but not limited to those covered in our lectures) to solve real-world tasks or conducting machine learning research.

Project Topics:
Each student will need to select a topic. There are two mainstreams of topics:

- Application project: Pick an application that interests you, and explore how best to apply learning algorithms to solve it.
- Algorithmic project. Pick a problem or family of problems, and develop a new learning algorithm, or a novel variant of an existing algorithm, to solve it.

You are suggested to pick something that you can get excited and passionate about, e.g., either an application area that you're interested in, or pick some subfield of machine learning that you want to explore more. Alternatively, if you're already working on a research or project that machine learning might apply to, then you may already have a great project idea.

Note: You may decide to work on Deep learning project, and since our course mainly discusses basic concepts for machine learning, please make sure that you use materials you learned in the class as well if you choose Deep learning project.

- For example, you might set up logistic regression and SVM baselines, or do some data analysis using the unsupervised methods covered in class.
- Also, training deep learning models can be very time consuming, so make sure you have the necessary computing resources.
- Deep learning model not be considered as more superior than other machine learning models if you just repeat a model that is designed by others.

To undertake the project, the following steps are essential:
1. Select one topic.
2. Survey on existing research on relevant topics by searching related keywords on an academic search engine such as: http://scholar.google.com.
3. Collect, read, and analyze relevant materials /data.
    - An important aspect of designing your project is to identify one or several datasets suitable for your topic of interest. Get the benchmark datasets and validate your learning algorithms on the benchmark datasets is preferred. We don't want you to spend much time collecting raw data.
    - If you choose to use preprepared datasets (e.g. from Kaggle), we encourage you to do some data exploration and analysis to get familiar with the problem.
4. Design and implement learning algorithms and validate the proposed algorithms on benchmark/collected dataset.

- We expect a solid methodology, comprehensive validation and detail discussion of the experimental results.
- Replicating the results in a paper can be a good way to learn. However, instead of just replicating a paper, also try using the technique on another application, or do some analysis of how each component of the model contributes to final performance.

5. Produce a report and give the presentation.

A very good project report will be a publishable or nearly-publishable piece of written work. You may read some recent papers and follow the writing styles. Refer to the "previous projects" from http://cs229.stanford.edu/projects.html for reference .

Submission Requirement:

Upon completion, each student must submit the following materials:

1. Project report, your report should contain but not limited to the following content:
   a) Motivation and background of the topic
   b) Related works and existing techniques of the topic
   c) Methodology
   d) Experimental study and result analysis
   e) Future work and conclusion
2. Link and description to the Dataset and the implementation code.

Assessment:

In general, projects will be evaluated based on:

- Significance. (Did the authors choose an interesting or a "real" problem to work on, or only a small "toy" problem? Is this work likely to be useful and/or have impact?)
- The technical quality of the work. (i.e., Does the technical material make sense? Are the things tried reasonable? Are the proposed algorithms or applications clever and interesting? Do the student convey novel insight about the problem and/or algorithms?)
- The novelty of the work. (Is this project applying a common technique to a well-studied problem, or is the problem or method relatively unexplored?)

The assessments are:

1. Report: 30%
2. Presentation: 20%
   a) Each student will have 6 minutes for presentation and 4 minutes for QA.
3. Code: 20%