

BNU-HKBU United International College
DS 4023: Machine Learning
Autumn 2024 Course Project

Description:

This is a **group** project (3-5 group members), 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 group 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.

(some projects will also combine elements of applications, algorithms and theory)

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. For inspiration, you might also look at some recent machine learning research papers. You can find research papers of top conferences in this website <https://aipapers-top.github.io/>, in which NeurIPS, ICML, ICLR are conferences most relevant to machine learning. 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.

To undertake the project, the following steps are essential:

1. Select one topic. (week 7-8)
2. Survey on existing research on relevant topics by searching related keywords on an academic search engine such as: <http://scholar.google.com> (remember to **narrow the topic** to a feasible and suitable scope) (week 7-8)
3. Collect, read, and analyze relevant materials /data. (week 9)
 - 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/framework and validate the proposed algorithms (baselines) on benchmark/collected dataset. (week 10-11)
 - 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 the report and slides. (week 12)

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. For example, IEEE format is common for research papers <https://www.ieee.org/conferences/publishing/templates.html>, <https://www.overleaf.com/latex/templates/ieee-conference-template/grfzhnncsfqn>.

Here are some example project works of Stanford University, for your

reference: https://docs.google.com/spreadsheets/d/1GIsGZKozmTxqzmXLt7FBmVRCHVJvgd_1DD-f9f5PwkU/edit#gid=0, <https://cs229.stanford.edu/proj2021spr/>.

However, you should **ensure your own topic** because these topics are relatively old (several years ago).

Submission Requirement:

Upon completion, each group 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 (can use the content in the survey)
 - c) Methodology
 - d) Experimental study and result analysis
 - e) Future work and conclusion
2. Link and description to the dataset and the implementation code.
3. Slides for presentation.

Submission deadline:

1. Slides for presentation: Dec. 1, 2024
2. Report and code: Dec. 20, 2024

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 (good performance), or is the problem or method relatively unexplored?)

The assessments are:

1. Report: 30%
2. Presentation: 50% (in the [week 13-14](#), the detailed arrangement will be determined)
3. Code: 20%