CS182 - Introduction to Machine Learning, Fall 2024-25
Course Projects

Prof. Ziping Zhao

One of main goals of CS182 is to prepare you to utilize machine learning techniques to solve real-world problems, and this course project provides a good opportunity for you to start in this direction.

I. Specifications

As a part of evaluation of CS182, you are required to complete a course project based on this instruction. You can choose to compare different existing learning algorithms. You can propose methods to improve the performance of specific learning methods. You can also try to develop novel machine learning models and/or algorithms. You can pick an interesting dataset or application, applying one or more well-known machine learning algorithms as baselines, and extending these baselines in creative and innovative ways. The general guidelines are listed as follows.

- This course project can be completed individually or as a team. Each team must not exceed 3 people.
- This course project should be completed based on the methods learned from CS182.
- Projects will be evaluated based on a combination of a project presentation (30%) and a final report (70%).
- There will be a project presentation (no more than 5 minutes). You need prepare a slides for this oral presentation to show your work. Each team member must attend.
- The final submission consists of two major parts: one final report and the source code.
  - **Final writeup**: You are expected to submit a final report summarizing your findings, ideas, contributions, etc. The course report should be written in a uniform format[1] and **no longer than 9 pages** for the body of the report with additional pages for references. Only the hard copy in **pdf**, rather than the source latex code, should be submitted.
  - **Source code**: For the sake of convenience, it is highly recommended to use Python to implement your ideas and algorithms in this project. However, any other programming languages are allowed. It is your responsibility to make sure the source code is executable and contains no bugs. Please submit the code in a separate **zip** file.

II. Structure of the Written Report

In order to make the evaluation of the project as objective as possible, the report should strictly adhere to the following structure with the sections (a penalty will be applied if the report is not organized according to the guideline):

Abstract: 10% grade

1) Introduction of the problem: 10% of grade
2) Models applied: 35% of grade
   a) If your project involves comparing existing methods, select at least 3 methods and provide a detailed overview of each.
   b) If your project aims to improve existing methods or propose novel methods, provide a critical overview of existing work before introducing your new methods.
3) Numerical results: 30% of grade
4) Conclusions: 10% of grade

References: 5% of grade

[1] https://epan.shanghaitech.edu.cn/l/lF3Far

III. Schedule and Submission

1) *Group member & Topic:* By **November 28th, 2024 (CST)**, you need to form your team (if you want) and choose a topic, and report these information to the following link:
https://docs.qq.com/sheet/DYmN4S3dqcnVyV2VD
Note: Every team member should fill in the form.
2) *Presentation:* At **the end of this course**.
3) *Final report:* By **January 6th 11:59pm, 2025 (CST)**, submit your final report using filename

{first_stu_name}–{second_stu_name}–{third_stu_name}–{project_name}.pdf

and a compressed file with source codes and all cited references using filename

{first_stu_name}–{second_stu_name}–{third_stu_name}–{project_name}.zip

to the following link:
https://epan.shanghaitech.edu.cn/l/vFlUQh
Note: Only one team member is supposed to submit the final report.

Please follow the submission deadlines above. They are strict deadlines and there will be penalties for not respecting them. In particular, the final reports late by 1 day will be penalized with 20% of the grade, late by 2 days will be penalized with 40% of the grade, and late by 3 days is most likely a fail in the project evaluation.

IV. Project Topics

Since the project focused on dealing the real-word problem, we recommend some machine learning competitions in Tianchi. There are many avenues that you may pursue for this project, and we encourage you to be brave and creative even if you do not think you will necessarily get "good" results. The project evaluation is not solely based on whether you have solved the problems. We are more interested in seeing your exploration and thought process throughout the project.

Here are some preliminary topics:

- Predicting the usage of O2O coupon: https://tianchi.aliyun.com/competition/entrance/231593
- Predicting the memory failure: https://tianchi.aliyun.com/competition/entrance/532055
- Predicting the interactions of Weibo: https://tianchi.aliyun.com/competition/entrance/231574
- Happiness prediction: https://tianchi.aliyun.com/competition/entrance/231702
- Clustering the automotive: https://tianchi.aliyun.com/competition/entrance/531892
- Heartbeat signal classification: https://tianchi.aliyun.com/competition/entrance/531883
- Repeat purchase prediction of Tmall: https://tianchi.aliyun.com/competition/entrance/231576

There are some guidelines you need to follow:

- In your report, you must include the same evaluation metrics used on Tianchi. Additionally, if other evaluation metrics highlight the advantages of your methods, we recommend including them.
- Registration on Tianchi and participation in its competitions are not mandatory. However, if you do participate in a Tianchi competition, please report your ranking in the "Numerical Results" section.
- You are expected to complete the topic independently, avoiding plagiarism and the use of unauthorized internet resources. We will review the uploaded code, and if any cheating behavior is detected, disciplinary action will be taken in accordance with university rules:
https://oaa.shanghaitech.edu.cn/2015/0706/c4076a31250/page.htm

Note:

- **If you wish to choose your own topic, please first email the instructor to discuss it. The instructor also offers research-oriented topics. If you are interested in gaining research experience during this project period, feel free to contact the instructor.**

- **You can get the bonus points by submitting your project paper to CCF A/B conferences/journals by the end of 16th week. Please provide us with paper and submission proof (email notification) to get it.**

V. After CS182

An excellent CS182 project will be publishable or nearly-publishable piece of work. After completing CS182, if you would like to continue working on your project along this direction as your graduation project, or submit your work to a machine learning conference or journal, please feel free to talk to the instructor.