NUS IT3011 Project Guidelines

1 Overview

There are two kinds of projects that can be done: Self-Defined Projects and Kaggle Competition Projects.

1.1 Self-Defined Projects

The project is intended to be a limited investigation in an area of machine learning of your choice. The purpose of the project is to enable you to study an area of your interest in greater detail in a practical way. The project can take on many forms, including but not limited to:

a. Projects that explore the application of machine learning ideas to an interesting "real-world" problem.
b. Projects that involve a theoretical or empirical study of aspects of a learning method or model.
c. Projects that do an experimental, comparative study of various machine learning methods.
d. Projects that extend previous work (e.g., UROP or FYP) of a project member[1].

Doing such a project gives you more flexibility and allows you work on something of your liking. However, at the same time, this may potentially require some additional effort (depending on your problem) such as data collection or coming up with suitable baselines or such.

1.2 Kaggle Competition Projects

Alternatively, you can try your hand at Kaggle Competitions. On the website you can find and choose from a number of interesting machine learning competitions. Upon joining a competition, you will be provided with a training and testing sets, and your performance will be measured with specified metrics and ranked with other competitors on the web. You can even win money if you do very well!

The system will score your predictions and inform you of the same. Your scores will be uploaded to a leaderboard for you to compare your performance with your classmates as well as research groups from other institutes.

Note that performance on the different metrics is not the critical factor in your grade on the project. While doing well on the competition will help, you are expected to come up with interesting ideas to solve the task you choose, which is what we will primarily look at.

While the data is easier to obtain for such a project, there is less flexibility and more emphasis on coming up with interesting methods.

2 Project Proposal

The project proposal should outline what you want to do in your project. We will give you feedback on your proposal to make sure the project you are proposing is feasible and appropriate. The proposal should contain the following sections:

Motivation. Explain why this project is interesting and important.

Statement of the Problem/Task. A statement of the problem, issue, or task that you're interested in studying. In particular, try to formulate the key questions (2 to 4 questions is probably a good number) that you will answer in the project.

General Approach. A high-level description of the general approach you will use to address the questions. This should include how you will evaluate and what evidence you are planning to gather (e.g. how you can answer the questions through experiments on data).

Resources. A list of resources you have/need to conduct the project. This includes reading, software, datasets, etc. How are you planning to get these resources?

Schedule. A schedule of work indicating the dates by which you plan to complete components of the project. Make sure the schedule is plausible.

Format: Please use the proposal template available in LumiNUS to prepare your submission. The project proposal should not exceed 1 page. Only one submission per team should be submitted through a single student account (do not submit multiple versions of the project proposal by different team members) on LumiNUS.

3 Interim Consultation

After the midterm period, we expect most student teams to have done some introductory reading and work towards their project goals. The interim project consultation is a mandatory, 15 minute timeslot in which you will have to meet with the lecturer and TAs to discuss your project.

In preparation, you must provide a set of presentation slides to describe your project and feedback that you wish to get from the course staff. Please practice such that you can conclude your presentation within 5-10 minutes so as to make use of staff feedback in the remaining time of your consultation slot.

The slides should consist of a (1) title slide with your team ID; (2) motivation and/or statement of task; (3) key questions to answer; (4) schedule. Teams are advised to make this session useful to you in gathering input or feedback on your project, as well as helping the team assess the quality of your project.

You are encouraged and welcomed to contact the lecturer or TAs to get more feedback on your project, and welcome to use the Project forum on LumiNUS to solicit additional feedback.

A template for the interim consultation slides will be made available on LumiNUS.

[1] If you are working with other faculty or students on the project, please mention that clearly in your project proposal. Note that this type of project sometimes results in uneven member contribution, due to the prior expertise of some members. The team will need to pay greater attention to balancing effort with this project type.

4 Final Project Presentation

Your presentation should contain the following:

- Provide motivation for your project, explaining why it is important and interesting,
- Explain your research questions,
- Provide evidence,
- Draw conclusions.

Presentations are meant for technical presentation of the quality and rigor of the project. It should contain scientific content that helps to defend the decisions made in the project and discuss insights and results related to the theory of the course.

Your team's presentation serves as a primary means of evaluating your project by the staff. You should try to make your presentation as self-contained as possible. You should be able to summarize your project within 5-10 minutes, with sufficient time to answer questions about your project within a 15-minute slot. Recall that the audience for the presentation are your course peers and your instruction staff.

N.B. all student teams' course outputs (presentation slides, final report) are considered as open-source and will likely be archived on the course website. If you would like your project held in confidence, please approach our staff; we will be happy to comply within reason.

The template for the final presentation will be made available on LumiNUS.

Grading Rubrics

Content (main consideration):

- Does the presentation contain a good summary of what you think the team's report should reflect?
- Do you think it is easy to glean the key findings of the team's work as narrated by the team?
- Does it address the motivation / significance of the project, its originality (if a self-defined project), the relevance of the team's work to the project's theme, the quality and persuasiveness of its arguments and linkage with the concepts taught in the course?
- Does it address both macroscopic, dataset-wide level performance (e.g., RMSE, Accuracy, F_1 measures) as well as microscopic, individual instance level performance (careful error analysis with diagnosis)?
- Are references appropriately shown?

Organization

- Is the presentation well organized with respect to accessing the content?
- Does it have a good flow?
- Does it utilize space appropriately for the amount of content that is discussed?

Formatting

- Is the presentation visually well-formatted?
- Are fonts, color, whitespace and visual elements appropriately used?
- Are important highlights easy to find?
- Are the content items proofread and clear?
- Images cleared for use and attributed where necessary?
- Are any links for project code appropriately formatted?

5 Final Project Report

If you submit multiple files (most groups), the .zip file should include your writeup and the source code of any programs you wrote for your project (don't just include a link to your repository). Include other files if you feel they are appropriate, but obviously explain their relevance in a README. All submissions must be made in LumiNUS.

For additional guidance in structuring the report, take a look at the template structure here. Not every project fits into this structure, and you might choose a different structure instead. The most important goals to keep in mind are:

- To motivate your project,
- To make a convincing argument that supports your conclusions,
- To make sure that the reader understands what your project is about and how you came to your conclusions, and
- To make sure that credit is given to all software, literature, *etc.* that helped you in your work.
- Format: Please use the double column format provided. The progress report should not be more than six pages in length (excluding references). Being concise is a good thing, but do not sacrifice clarity and completeness.

The projects will be graded in the same spirit as research papers are assessed (though we do not expect you to do original work at the same level). Here is a list of things that we will be looking for in your Project Report:

- Originality
- Significance (are the questions you are asking interesting)
- Relevance to course
- Quality of arguments (are claims supported, how convincing are the arguments you bring forward)
- Connection to earlier work (scientific literature or lecture materials)
- Clarity of writing (how clearly are goals and achievements presented)
- Scope/Size (in proportion to size of group, member expertise mix)

The template for the final report will be made available on LumiNUS.

6 Grading

The projects will be graded in the same spirit as research papers are assessed (though we do not expect you to do original work at the same level). Your project overall accounts for 30% of your final grade, which is broken down as follows:

- 5% for proposal
- 5% for interim consultation (slides presentation)
- 12% for final presentation
- 8% for final report

Feel free to come and talk to us about the various aspects of your project (in fact we strongly encourage you to) so that we can make sure that you are on the right track. Finally, have fun while doing it; its meant to be something that you are interested in doing!

Good luck!

Credits: Project guidelines based on Cornell's Thorsten Joachims and NUS's Bryan Low.