Project
Course: **Natural Language Processing(NLP) 2025**   Date: March 18, 2025

Project:

To get yourself familiar with both Tensorflow and PyTorch, this is a Stanford **github** repository with project code examples, including a computer vision and a natural language processing example (both in Tensorflow and Pytorch). You can visit this repository to see how both TensorFlow and PyTorch work for the same problem. The code can be reused in your projects, but the examples presented are not complex enough to meet the expectations of a project.

Project Topics:

In NLP, you will learn about a wide range of methods and applications which we use in everyday life while interacting with any Smart machine. Part of the learning will be during in-class lectures and when completing assignments, but you will really experience hands-on work in your final project. We would like you to choose wisely a project that fits your interests. One that would be both motivating and technically challenging.

Most students do one of three kinds of projects:

- **Application project**. This is the most common: 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.

- **Theoretical project**. Prove some interesting/non-trivial properties of a new or an existing learning algorithm. (This is often quite difficult, and so very few, if any, projects will be purely theoretical.) Some projects will also combine elements of applications and algorithms.

Many fantastic class projects come from students picking either an application area that they're interested in or picking some sub-field of machine learning that they want to explore more. So, pick something that you can get excited and passionate about! Be brave rather than timid, and feel free to propose ambitious things you're excited about. (Just be sure to ask us for help if you're uncertain how to get best started.)

Project Hints

An excellent project will be a publishable or nearly-publishable piece of work. There are many good conferences from where you can choose your topics a few examples are ICML, ICLR, WWW, ICDM, AAAI, IJCAI, NACL, ACL, USENIX Security, etc. Once you have identified a topic of interest, it can be helpful to look up existing research on relevant topics by searching related keywords on an academic search engine such as: http://scholar.google.com. Another critical aspect of designing your project is identifying one or several datasets suitable for your topic of interest. If that data needs considerable preprocessing to suit your task, or if you intend to collect the needed data yourself, keep in mind that this is only one part of the expected project work but can often take considerable time. We still hope a solid methodology and discussion of results, so pace your project accordingly.

- **Computation power**. At IIT Jammu, we have GPU and HPC with us, and if you want to use the same, please get in touch with TA to get GPU access for your projects. Alternatively, Google Cloud and Microsoft Azure offer free academic units, which you can apply to. Also, we have Kaggel and Google cloud to do some preprocessing tasks, so you can also avail of that facility.

- **Preprocessed datasets**. While we don't want you to have to spend much time collecting raw data, the process of inspecting and visualizing the data, trying out different types of preprocessing, and doing error analysis is often an important part of machine learning. Hence if you want to use preprepared datasets (e.g., from Kaggle, the UCI machine learning repository, etc.), we encourage you to do some data exploration and analysis to get familiar with the problem.

- **Replicating results**. Replicating the results in a paper can be a good way to learn. However, we ask that instead of just replicating a paper, we also try using the technique on another application or do some analysis of how each component of the model contributes to the final performance.

Project Deliverable

This section contains detailed instructions for the different parts of your project.

Evaluation: We will not be disclosing the breakdown of the weightage that the final project is worth amongst the different parts, but the Presentation and final report will combine to be the majority of the grade. Projects will be evaluated based on the following:

- 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 authors convey novel insight about the problem and/or algorithms?)

- 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 an impact?)

- 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?)

To highlight these components, you must present a solid discussion regarding the learnings from the development of your method and summarize how your work compares to existing approaches.

Proposal

Deadline: 30th March, Monday 11:59 PM
First, make sure to submit the following **Google form To Be Filled**. In the form, you will have to provide your project title, team members, and relevant research area(s).

Your project proposal should include the following information:

- What is the problem that you will be investigating? Why is it interesting?

- What are the challenges of this project?

- What dataset are you using? How do you plan to collect it?

- What method or algorithm are you proposing? If there are existing implementations, will you use them, and how? How do you plan to improve or modify such implementations?

- What reading will you examine to provide context and background? If relevant, what papers do you refer to?

- How will you evaluate your results? Qualitatively, what kind of results do you expect (e.g., plots or figures)? Quantitatively, what kind of analysis will you use to evaluate and/or compare your results (e.g., what performance metrics or statistical tests)?

Topic for project is **EMBEDDING** and look for paper in these conference but paper should not be older than 2 year.

- ACL

- COLING

- NIPS

- EMNLP

- NAACL

- AAAI

As for the latex template, you can use any template for the report present in the overleaf template.

Honor code

We strongly encourage students to form study groups. Students may discuss and work on programming assignments and quizzes in groups. However, each student must write down the solutions independently and without referring to written notes from the joint session. In other words, each student must understand the answer well enough to reconstruct it by themself. In addition, each student should submit their code and mention anyone they collaborated with. It is also an honor code violation to copy, refer to, or look at written or code solutions from a previous year, including but not limited to: official solutions from a previous year, solutions posted online, and solutions you or someone else may have written up in a previous year. Furthermore, posting your assignment solutions online is an honor code violation, such as on a public git repository.

IIT Jammu Honor Code for CS Courses