Discipline of
Computer Science & Engineering

Assignment 1 and Project
Course: **Deep Learning and Advance AI** Date: September 7, 2021

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 Tenserflow and Pytorch work for 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 CSC004P5E, you will learn about a wide range of deep learning applications. Part of the learning will be online, 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 by far 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 do feel free to propose ambitious things that you're excited about. (Just be sure to ask us for help if you're uncertain how to best get started.) Alternatively, if you're already working on a research or industry project that deep learning might apply to, then you may already have a great project idea.

Project Hints

A very good project will be a publishable or nearly-publishable piece of work. There are many good conferences from where you can chose your topics from few examples are ICML, ICLR, WWW, ICDM, AAAI, IJCAI, NACL, ACL etc. Once you have identified a topic of interest, it can be useful to look up existing research on relevant topics by searching related keywords on an academic search engine such as: http://scholar.google.com. Another important aspect of designing your project is to identify one or several datasets suitable for your topic of interest. If that data needs considerable pre-processing to suit your task, or that 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 expect 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 use the same please contact TA′s 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 pre processing tasks so you can also avail 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 choose 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, also try using the technique on another application, or do some analysis of how each component of the model contributes to final performance.

Project Deliverables

This section contains the 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 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 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?)

In order to highlight these components, it is important you present a solid discussion regarding the learnings from the development of your method, and summarizing how your work compares to existing approaches.

Proposal

Deadline: 10^{th} October, Sunday 11:59 PM
First, make sure to submit the following **Google form** so that we can match you to a TA mentor. In the form you will have to provide your project title, team members and relevant research area(s).

In the proposal, below your project title, include the project category. The category can be one of:

- Computer Vision
- Natural Language Processing
- Generative Modeling
- Speech Recognition
- Reinforcement Learning
- Healthcare
- Others (Please specify!)

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)?

one good project proposal example I find out from stanford **Click here** if you want to look. As for latex template we will provide you the same.

Assignment Problem 1

Deadline: 20^{th} September Monday 11:59 PM.

A web crawler:

A web crawler is a program that starts with a list of links/URLs/addresses to web pages, and works to traverse the pages reachable by following links starting at the starting pages. The best-known web crawlers are the Googlebots (and their friends the bingbots, etc.), used by Google to collect the data used by their search engine.
For this assignment, you will write a small web crawler to collect all 3 types of data Image,text and videos. And store that data into some specific files so you can use them for your further Assignments.

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 solution well enough in order to reconstruct it by him/herself. In addition, each student should submit his/her own code and mention anyone he/she 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, it is an honor code violation to post your assignment solutions online, such as on a public git repository.

IIT Jammu Honor Code for CS Courses