CS5242 : Neural Networks and Deep Learning

Project

Semester 1 2021/22

Xavier Bresson

https://twitter.com/xbresson

Department of Computer Science
National University of Singapore (NUS)

Outline

- Project Goals
- Group Selection
- Project Schedule
- Project Development
- Marking Scheme
- Misc

Outline

- Project Goals
- Group Selection
- Project Schedule
- Project Development
- Marking Scheme
- Misc

Goals

- Project will focus on :
  - Theoretical knowledge received in this module.
  - Practical skills with data acquisition, exploration, exploitation, analysis.
  - Teamwork with management of tasks.
  - Concise and clear communication with oral presentation.

- Project counts for 30% of the final grade.

Goals

- Project specifics :
  - Collect and prepare a small dataset of images or/and text documents.
    - This dataset(s) must be new (i.e. not downloaded from e.g. Kaggle or GitHub).
  - Implement MLP, CNNs, RNNs and ANNs on this dataset(s).
    - MLP as baseline, CNNs/RNNs/ANNs are expected to improve MLP.
  - Propose an improvement for each class of neural networks.
    - Motivation, description, equation, implementation, result, discussion.
  - Demonstrate initiatives
    - Develop own scrapper, dynamic visualization, discover new data insights, etc.
  - Deliveries
    - Python notebook with project report and code.
    - Video presentation and slides.

Project Philosophy

- This project focuses on
  - The understanding of the fundamental concepts of deep learning techniques,
  - The practical skills required to develop a data analysis project.
- It is not about learning to use GitHub codes.
- It is not about winning a Kaggle competition.
- It is not about ten lines of Keras' code to run deep learning techniques.
- It is not about running long experiments with the best possible GPUs.
  - Google Colab, Google Cloud, and your computer/laptop are enough.
- It is not about getting 90% of accuracy.
- It is about how to design from scratch, debug, understand and train neural networks.
- It is about to understand why it works and why it does not.

Outline

- Project Goals
- **Group Selection**
- Project Schedule
- Project Development
- Marking Scheme
- Misc

Groups

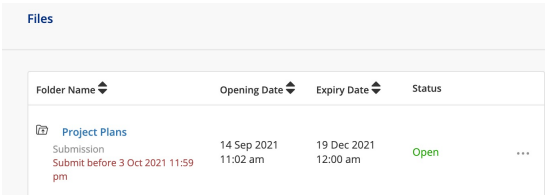
- The project is group-based with a size of at most 3.
  - It can be composed of 1 or 2 or 3 members, preferably 2-3 but not more than 3.
  - Use LumiNUS=>Forum=>Project Groups if you are looking for teammate.
- Choose your group wisely
  - Each teammate must contribute *equally* to the project.
    - Avoid conflict and make a short written "contract" at the beginning of the project regarding the commitment of each team member to the project.
    - Clearly, each member has different skills and it is fine to be weaker in maths, coding, etc. However, it is not fine to let the other teammates to do most of the work.
  - The notebook and the presentation must indicate clearly the contributions of each teammate.
  - Each teammate will introduce her/his contribution to the project in the video presentation.
- Group selection :
  - Please, enter your group in this document
    - https://drive.google.com/drive/folders/1fnr59Eb_AIMXdT2i3UBKv8CVzhv7DBI3
  - Deadline : Sept 26th 11:59pm
  - Penalty : You lose 25% of the grade if you submit your group selection after the deadline.

Outline

- Project Goals
- Group Selection
- **Project Schedule**
- Project Development
- Marking Scheme
- Misc

Schedule

- Week 6 and Recess Week : Group selection
- Recess Week and Week 7 : Project plan writing
  - Write a project plan.
    - Clear and concise one-page description of the project
    - Project motivation, description, proposed solution, project milestones.
    - Write down the tasks for each team member.
  - Submit the project plan by Oct 3rd 11:59pm to LumiNUS=>Files=>Project Plans
    - File name must be "project_plan_groupIDXX.pdf", (for example project_plan_group38.pdf).
  - Project plan will be graded and counts for 10% of the project grade (100% project = 10% plan + 90% delivery)
    - Penalty : You lose the 10% if you submit the project plan after the deadline.
    - Grade release : LumiNUS=>Gradebook on Oct 11th.

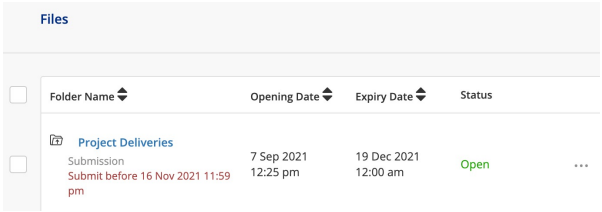


Schedule

- Week 8 – Week 14 : Project development
  - Work on the project every week.
  - Do not wait for the last week to start working on the project!
  - If you have any questions:
    - First, ask the TAs assigned to your group.
      - The TAs assigned to your group will be announced after the group selection is completed.
    - Ask me during the live lectures.

Schedule

- Week 14 : Project deliveries
  - A working/reproducible python notebook which includes the project description/report (with Markdown).
  - Presentation slides and video.
  - Deadline : Tuesday Nov 16th 2021 11:59pm.
  - Upload notebook, slides and video here:
    - Create a zip file with your notebook, slides and video.
      - Use the format "project_groupIDXX.zip" (for example project_group38.zip).
    - LumiNUS=>Files=>Project Deliveries
      - Note that the maximum upload file size on LumiNUS is 500MB.
  - Penalty : Final grade is zero if notebook, slides and video are not delivered in LumiNUS at the time of the deadline (unless medical certificate).



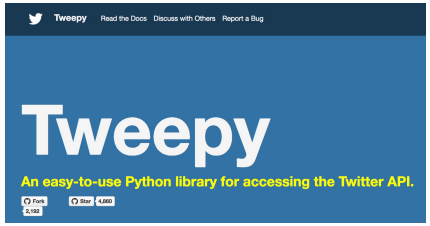
Outline

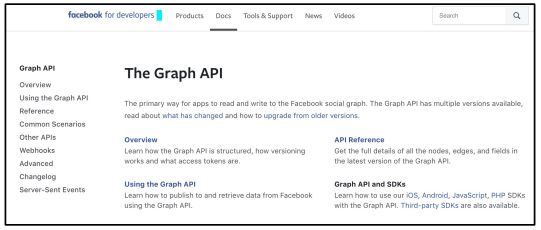
- Project Goals
- Group Selection
- Project Schedule
- **Project Development**
- Marking Scheme
- Misc

Project development

- Step 1: Identify a data problem to solve.
  - Use your own field of expertise or your interests.

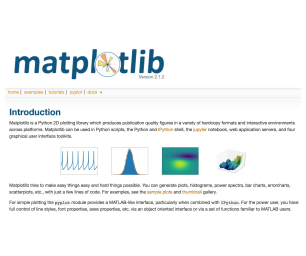
- Step 2: Data collection
  - Develop your data scraper.
  - Use API
    - Twitter: http://www.tweepy.org
    - Facebook: https://developers.facebook.com/docs/graph-api
    - Kaggle: https://github.com/Kaggle/kaggle-api

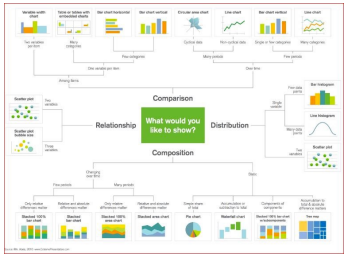


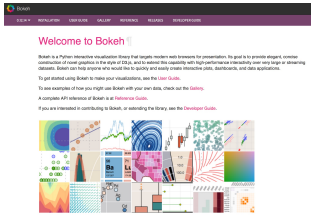


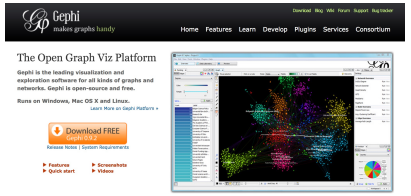
Project development

- Step 3: Data exploration (analyze your data, get insights)
  - Use statistics
  - Use visualization libraries:
    - Matplotlib: https://matplotlib.org
    - Bokeh: https://bokeh.pydata.org
    - Graphlab: https://gephi.org





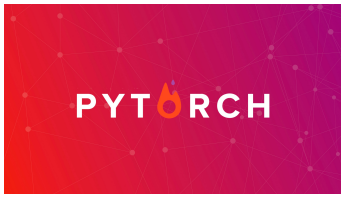




- Step 4: Pre-processing
  - Data cleaning (missing features)
  - Data normalization (unbalanced scaling)
  - Important and consuming step to prepare data as clean as possible for analysis

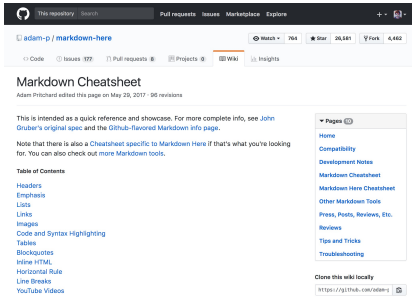
Project development

- Step 5: Data analysis with deep learning
  - Apply deep learning to solve your data problem :
    - Regression, classification, etc
  - Compare different models



- Step 6: Numerical results
  - Analysis, interpretation, conclusion

- Step 7: Report
  - Use Python Notebook and Markdown: https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet
  - Future of scientific reports:
    - Code + description + analysis merged into a single document.
    - Code is reproducible, transferable to a new dataset, can be extended with new ideas.



Project development

- Step 8: Video presentation
  - The project presentation must present concisely a standard data science project:
    - Project motivation and description, data acquisition, data exploration, pre-processing, proposed deep learning solutions, analysis of results, future development.
  - Each person must present her/his contribution to the project.
    - You will receive grade zero if you do not present your contribution.
  - Use slides (one slide is 1-2min).
  - The length of the presentation is 8min.
    - The time is strict, not more than 8min (we will stop the video after 8min).
    - Each team member must present what she/he contributed to the project.
    - Each member has 2-3min if your group size is 3, and 4min/member if your group size is 2.
  - Convince us you understood what you did !

Outline

- Project Goals
- Group Selection
- Project Schedule
- Project Development
- **Marking Scheme**
- Misc

Marking scheme

- Project plan counts for 10%.
- A project that completes all steps 1-8 correctly gets 60%.
- Anything that demonstrates initiatives will receive up to 30% additional points.

Outline

- Project Goals
- Group Selection
- Project Schedule
- Project Development
- Marking Scheme
- Misc

GPU

- The project is not about running long experiments with the best possible GPUs.
- Google Colab (free GPU with limitations), Google Cloud (600hr of free GPU), and your computer/laptop are enough.
  - Read lecture "lecture02_part2_pytorch.pdf" for GPU setup.

TAs

- Ask the TA in charge of your group if you have any additional questions.

Questions?