Department of Electrical, Computer, and Software Engineering
University of Auckland
COMPSYS 302 S1 2020 – Python Project (Weighting 45%)

Background

Nowadays, ubiquitous artificial intelligence algorithms affect every aspect of our life. From searching recommendation systems, to self-driving cars, the foundation of these applications is Neural Networks. Although neural networks are easy to implement, their behaviour is hard to analyse, and the training of neural networks sometimes can be tricky. Moreover, in order to implement different functions, different neural networks shall be selected. This project will focus on designing and developing your own neural networks to fulfil different purposes. You will be asked to find your own research topic and implement the algorithm that can solve it well.

The first step of this project would be searching for the topic you are interested in. Different projects will have different features. **The project you selected should use labelled images as input data.** It will be a classification project with image input. A general **but not limited** framework is using the Convolutional Neural Network to reduce dimensionality first, then using fully connected layers to do classification. Please prepare **at least 4 different algorithms** for your project. **An evaluation of the algorithms you used is needed at the end of this project.** The potential datasets are listed below. Nevertheless, you are free to choose any other supervised learning projects other than that. For instance, you can try facial recognition or aerial imaging.

- ImageNet (http://www.image-net.org/)
- CIFAR-10 (https://www.cs.toronto.edu/~kriz/cifar.html)
- MNIST (http://yann.lecun.com/exdb/mnist/)
- More: https://en.wikipedia.org/wiki/List_of_datasets_for_machine-learning_research

The project will be implemented by PyTorch in Python3. You may need an IDE like Eclipse for better coding experience. A python project structure will be adapted, which including:

- models <folder to put your model>
- 'requirement.txt' <pip install all your dependencies>
- 'setup.py' <packaging>
- 'LICENSE'
- 'README.rst' <something about your project>
- '__init__.py' <something you can share globally>

Pytorch with CUDA acceleration is optional. With CUDA, the training time will be significantly reduced. However, it needs a large space on your computer and some more time for configuration. Moreover, please choose your dataset wisely so it will not be too big to train.

Students can find **one** partner to implement this project together. Two students will cooperate by using Github. Both students should contribute to project design and implementation. The use of Github will also be assessed as well. As you may know, there are lots of codes you can find online. However, we would suggest you **write your own code from scratch**.

Objective

The overall goal of this project is to get familiar with machine learning or supervised learning. Primary objectives are presented here. Student proposals can be discussed with the lecturer and TAs.

At a bare minimum, the requirements are:

1. The project must run in **python3** on Windows computer without us having to download additional packages or configure anything manually. Please design to allow us to run the training and evaluation with few lines of command.
    a. Write your instruction in README,
    b. Run the application,
    c. We will check the learning curve to make sure it is learning.
2. The project must:
    a. Classify images into different classes.
    b. Generate an evaluation graph for learning.
    c. Correctly using the concept of the training set, validation set and test set.

The purpose of this project is to learn supervised learning techniques. There are several libraries that will be learnt: Python, PyTorch, Numpy, Matplotlib, Pandas et al. The project is completed in groups**.**

Versioning

The use of the Git versioning system via Github is compulsory and will be monitored. Steady progress from all groups is expected throughout the semester, and you should not just commit the day before everything is due**. Students should create one private repository per group with the name "2020-Python-groupnumber"**.

Assessment

This project is completed and assessed in a group.

1. **Design Document---Proposal (10%) – Due 8pm Friday 27 March (Week 4)**

Please write a proposal about the dataset you select and why you select it. Moreover, what feature does this dataset present? It would be a good idea that you can talk about their applications. You may want to propose different learning algorithms. Or what do you think you can improve performance?

Other than algorithms and dataset, in the proposal, you also need to include the Gantt Chart for your project plan. How do you want to develop your project? Also, how would you evaluate your algorithms? You need to have a clear plan of what you are trying to do.

2. **Final Project Demo (15%) – Demo in Lab 12pm Wednesday 29 April (Week 7)**

On the final demo day (in addition to submitting the code), there will be **individual** interviews with each student, where each student individually presents their project, and discusses the novelty of the project.

3. **Question & Answer (5%) Same day with Demo**

Possible questions are listed below:

a. Dose larger network means better performance
b. Dose larger learning rate means better performance
c. Does your dataset balance? If it is not, plot it out each class and propose a solution.
d. Why do you need to divide your dataset into training, validating, testing?

4. **Final Project Report (10%) – Due 8pm Friday 1 May (Week 7)**

Your design should be documented in a report which is in the form of a real paper report. There are a few sections which are listed below you need to cover. Please remember to cover the five questions listed below, which will be marked as well. The length of the final report should be controlled under 2000 words. You don't have to write a whole thesis with it.

- (1%) Landing/welcome/index page, Format
  - Abstract to project/system
  - Category

- (2%) Introduction
  - Why did you choose this dataset?
  - Why would you use this algorithm?
  - Any interesting finding?

- (1%) (Background and Literature Review)
  - Background research about Deep Neural Network
  - Dataset selection, state of art results.

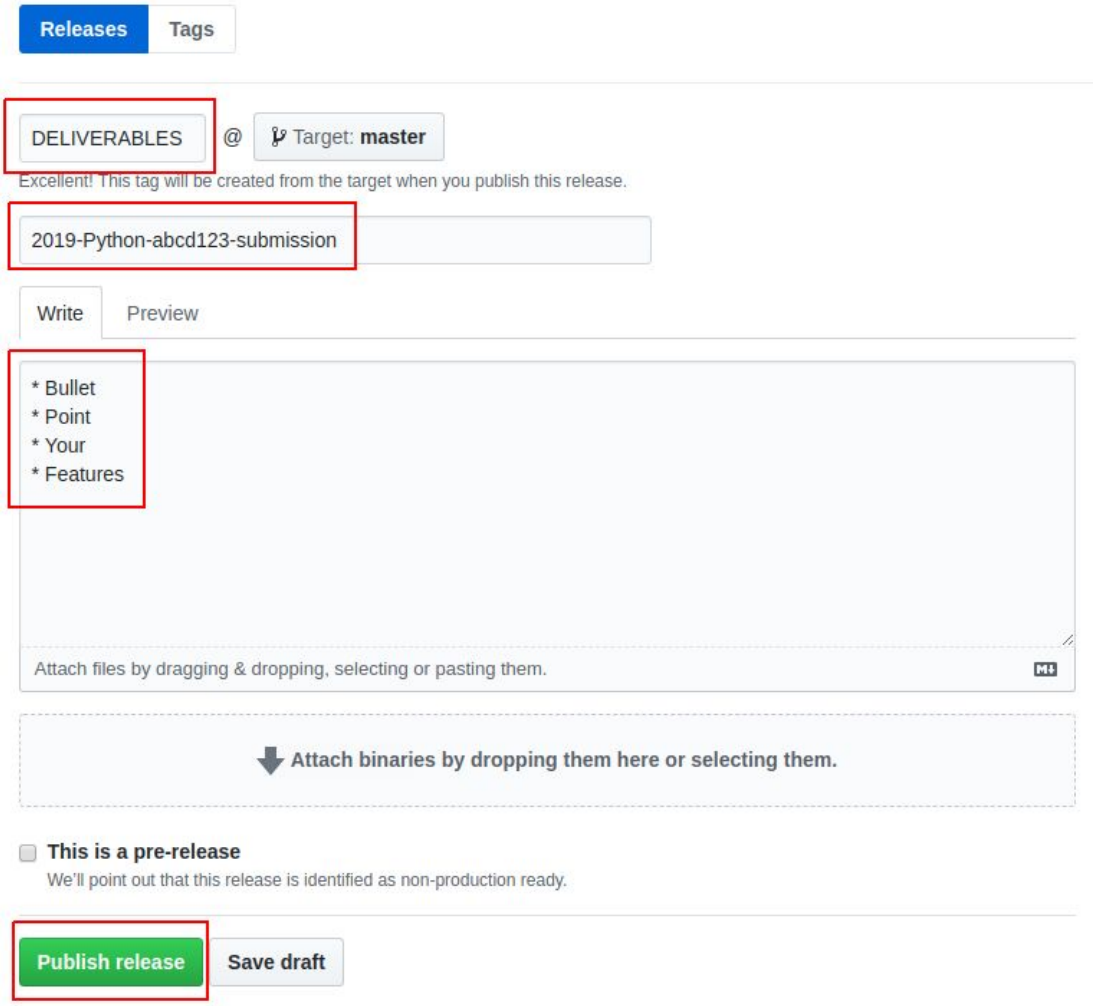
- (2%) (Methodology)
  - How you designed your neural network. Why do you think it is good?
  - Where do you think the learning ability is from and why?
  - A good illustration of your system.

- (3%) (Evaluation)
  - Please illustrate the learning process. It should be a curve.
  - The statistical learning result, precision rate, recall rate, f1.
  - Evaluation should be done with the Test dataset.

- (1%) (Result and Future Work)
  - Briefly describe what is your result, and do you think it is reasonable.
  - Anything interesting?
  - Draw your conclusion.

5. **Project Code (5%) – Due 8pm Friday 1 May (Week 7)**

Submission is via Github – create a release tag on the appropriate commit with DELIVERABLES as the tag name, and 2020-Python-GroupXX-submission as the title.


Releases
Tags
DELIVERABLES
@
Target: master
Excellent! This tag will be created from the target when you publish this release.
2019-Python-abcd123-submission
Write
Preview
* Bullet
* Point
* Your
* Features
Attach files by dragging & dropping, selecting or pasting them.
Attach binaries by dropping them here or selecting them.
This is a pre-release
We'll point out that this release is identified as non-production ready.
Publish release
Save draft

Final Note

In this project, you are expected to do your own planning and learning. You'll be responsible for finding out how to do things yourselves. There will be content delivered in lectures to help speed this process up at the beginning, but ultimately it is your responsibility to find out what you will need. If you get stuck and need help, please ask a TA or the lecturer. **Do not leave things to the last minute**. The easiest way to fail this project is to try to start it in the last week (or night) before it's due.

Do not mistake the minimum requirements as the only requirements; if you want to do well in this course, you will need to design your project carefully and manage your time well. Do not hesitate to contact the lecturers and TAs if you are in doubt! They are available to help with concepts, troubleshooting, and debugging, but they will not write your project.

Poorly documented or implemented code, while otherwise functionally correct, may not get you full marks for the project overall. Someone else has to be able to work on your code later! Please develop consistent and good coding practices.

Due to the nature of this project, a certain level of collaboration is required within the class. As a result, it is strongly recommended that you interact with other students in the labs. Additionally, TAs will be there, and you can discuss requirements with them as well.

Tips

- Select a reasonable dataset, try not to select very big one (not a Gigabytes ones)
- While training, start with a small network and low learning rate.
- CUDA acceleration is optional, it will boost training speed a lot, but it is also a bit pain to install.
- After the project, you will be able to do some online competitions. It will benefit your career in the future.
- https://www.kaggle.com/