

General Regulations

Machine Learning 2020 Spring

ML 助教群

March 5, 2020

Outline

1. Environment
2. Kaggle
3. GitHub
4. Grading
5. Contact TAs

Environment

助教強烈建議使用 Linux 或 macOS 系統來完成作業，Windows 系統常常特立獨行，很容易出問題。

如果先前沒有安裝系統的經驗，我們建議安裝較能快速上手的 Ubuntu，[Ubuntu 正體中文 Wiki](#) 裡面有相當詳細的介紹及教學。

所有的作業都必須以 Python 3.6.8 完成，可以使用的套件如[連結](#)所示，根據不同的作業，某些套件（或函數）會被禁止使用。

若有使用其他套件的需求，請事先來信詢問並說明理由，經助教同意後才可使用。

Pyenv 是一個管理 Python 版本的工具，使用者可以在系統中安裝不同且彼此獨立的 Python，維持系統原生 Python 不受套件安裝移除污染。我們強烈建議同學使用 Pyenv 來管理並維護作業的環境，若出現任何問題，可以直接將環境移除重新來過。

Pyenv - Installation

Pyenv 的安裝非常簡單：

```
git clone https://github.com/pyenv/pyenv.git ~/.pyenv
echo 'export PYENV_ROOT="$HOME/.pyenv"' >> ~/.bash_profile
echo 'export PATH="$PYENV_ROOT/bin:$PATH"' >> ~/.bash_profile
echo -e 'if command -v pyenv 1>/dev/null 2>&1; then\neval "$(pyenv init -)"\nfi ' >> ~/.bash_profile
exec "$SHELL"
```

其中 .bash_profile 會依環境而異：bash 為 .bashrc，zsh 為 .zshrc。

常用的 pyenv 指令如下：

- `pyenv versions`
列出目前系統中所有安裝的 Python
- `pyenv version`
顯示目前預設的 Python 版本
- `pyenv global <python_version>`
設定預設 Python 版本
- `pyenv install <python_version>`
安裝特定版本的 Python
- `pyenv uninstall <python_version>`
移除特定版本的 Python

Pyenv - Package Installation

使用 pyenv 後，仍然可以透過 pip 管理套件，所有 pip 操作的對象都是預設的 Python 版本。

```
>> pyenv versions
  system
* 3.6.8 (set by /home/Minori/.pyenv/version)
>> pip install numpy
```










→ numpy 會被安裝在 Python 3.6.8 上。

Kaggle

Kaggle 是一個龐大的資料科學社群，上面有各種資料分析的競賽。本學期的部份作業也會使用到 Kaggle 上的 in-class competition。

- Home
- Compete**
- Data
- Notebooks
- Discuss
- Courses
- More

All Competitions

Active	Completed	InClass	All Categories	Reward
 Deepfake Detection Challenge Identify videos with facial or voice manipulations Featured • a month to go • Code Competition • 1775 Teams \$1,000,000				
 Google Cloud & NCAA® ML Competition 2020-NCAAM Apply Machine Learning to NCAA® March Madness® Featured • a month to go • 114 Teams \$25,000				
 Google Cloud & NCAA® ML Competition 2020-NCAAW Apply Machine Learning to NCAA® March Madness® Featured • a month to go • 74 Teams \$25,000				
 DS4G: Environmental Insights Explorer Exploring alternatives for emissions factor calculations Analytics • a month to go \$25,000				
 Google Cloud & NCAA® March Madness Analytics Uncover the madness of March Madness® Analytics • 2 months to go \$25,000				
 Abstraction and Reasoning Challenge Create an AI capable of solving reasoning tasks it has never seen before Research • 3 months to go • Code Competition • 151 Teams \$20,000				
 Bengali AI Handwritten Grapheme Classification Classify the components of handwritten Bengali Research • a month to go • Code Competition • 1421 Teams \$10,000				
 Real or Not? NLP with Disaster Tweets Predict which Tweets are about real disasters and which ones are not Getting Started • a month to go • 3549 Teams \$10,000				
 Digit Recognizer Learn computer vision fundamentals with the famous MNIST data Getting Started • Ongoing • 2388 Teams Knowledge				

Registration and Name Format

同學們可以自由選擇使用 Google 或是其他網域的信箱註冊帳號，但 Display Name 必須遵守以下格式：

< 小寫學號 >_< 任意名稱 >

舉例來說：

- ✓ r08921062_Minori 好可愛
- ✓ r08921062_
- × R08921062_
- × r08921062
- × r08921062 三玖天下第一

旁聽的同學則請避免使用學號開頭的名稱。

助教在 Kaggle 比賽結束時會根據名稱對應學號及程式重現結果來給予分數，名稱錯誤者將拿不到這部份的分數，請務必多加留意。

Public/Private Leaderboard

在 Kaggle 的比賽中，測試資料會被切成 public 及 private 兩個部份。同學們在比賽時限之前只能夠看到 public leaderboard，之後才能知道 private leaderboard 的排名。

#	Team Name	Notebook	Team Members	Score	Entries	Last
1	b06902021_rm -f trained_model			0.77550	38	9mo
2	b05901176_(\f*\n*)			0.77400	28	9mo
3	b05901063_QQ			0.77380	23	9mo
4	r07522839_劉承岳			0.77130	11	9mo
5	b06902030_5/14資訊之夜			0.77020	30	9mo
6	b04901147_系統隊長			0.76920	24	9mo
Your Best Entry ↑ Your submission scored 0.76920, which is not an improvement of your best score. Keep trying!						
7	r07943150_ML競賽 輕鬆自在			0.76830	35	9mo
8	r07943156_慈母守中綠濶子遊野...			0.76770	19	9mo
9	b05901075_Five			0.76740	29	9mo

每個人一天至多可以上傳 5 次預測結果，並且在時限前可以勾選任意兩筆紀錄作為最後用於 private leaderboard 的 submission。

result.csv 9 months ago by b04901147_系統隊長 add submission details	0.76940	0.76800	<input type="checkbox"/>
result.csv 9 months ago by b04901147_系統隊長 add submission details	0.76610	0.76920	✓
result.csv 9 months ago by b04901147_系統隊長 add submission details	0.76300	0.76690	✓

注意：請不要貪心多開帳號來突破上傳限制，違者除該次作業 0 分外，最嚴重會被 Kaggle **封鎖帳號**。

GitHub

GitHub 是透過 Git 進行版本控制的平台，本學期所有作業的繳交都是透過 GitHub Classroom 來完成。

以下僅介紹幾個 Git 的基本操作，其餘的部份請同學自行上網學習。

- `git init`
初始化目前的資料夾
- `git clone <url>`
將網路上的資料 "clone" 到本地端
- `git add <files>`
將檔案加入追蹤
- `git rm --cached <file>`
取消追蹤檔案

- `git status`
查看目前的工作環境狀態
- `git commit -m <description>`
將追蹤的檔案新增到一個 commit
- `git remote add origin <url>`
將本地端對應到遠端網址
- `git push -u origin <branch>`
將 commit 同步到 GitHub

同學首次使用 GitHub Classroom 時必須進行帳號及修課名單的配對。

- 點選自己的姓名及學號

Join the classroom:

NTU-speech-lab-classroom-ML2020

To join the GitHub Classroom for this course, please select yourself from the list below to associate your GitHub account with your school's identifier (i.e., your name, ID, or email).

Can't find your name? [Skip to the next step](#) →

Identifiers	
b04901147 白銀御行	>
b05902127 石上優	>
r08921062 四宮輝夜	>
r08944041 藤原千花	>

- 按下 "Accept this assignment" 來「收下」這份作業

NTU-speech-lab-classroom-ML2020

Accept the assignment —

hw0

Once you accept this assignment, you will be granted access to the `hw0-cyhuang-tw` repository in the [NTU-speech-lab](#) organization on GitHub.

Accept this assignment


- 看到這個畫面就表示你的專屬 repository 已經建立好了



You're ready to go!

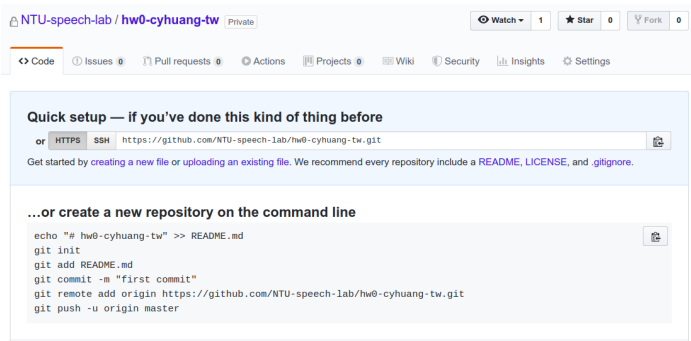
You accepted the assignment, **hw0**. Your assignment repository has been created:

 <https://github.com/NTU-speech-lab/hw0-cyhuang-tw>

 Your assignment is due by **Mar 6, 2020, 23:00 TST**

Note: You may receive an email invitation to join [NTU-speech-lab](#) on your behalf. No further action is necessary.

- 接下來上傳作業的方式就跟一般的 GitHub repository 相同



NTU-speech-lab / hw0-cyhuang-tw Private

Watch 1 Star 0 Fork 0

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

Quick setup — if you've done this kind of thing before

or **HTTPS** **SSH** `https://github.com/NTU-speech-lab/hw0-cyhuang-tw.git`

Get started by [creating a new file](#) or [uploading an existing file](#). We recommend every repository include a [README](#), [LICENSE](#), and [.gitignore](#).

...or create a new repository on the command line

```
echo "# hw0-cyhuang-tw" >> README.md
git init
git add README.md
git commit -m "first commit"
git remote add origin https://github.com/NTU-speech-lab/hw0-cyhuang-tw.git
git push -u origin master
```

Grading

每項作業滿分皆為 10 分，評分項目根據 Kaggle 競賽的有無分為兩種：

- 有 Kaggle 競賽
 - (1%) public simple baseline
 - (1%) public strong baseline
 - (1%) private simple baseline
 - (1%) private strong baseline
 - (6%) report
- 無 Kaggle 競賽
 - (10 %) report

Kaggle private leaderboard 排名前五的同學，若願意在 FB 社團分享自己的作法，則可以另外獲得 1 分的 bonus。

學期成績以分數最高的前十個作業計算，因此不需要完成所有的作業。

在作業繳交截止日期後，同學仍有三天的緩衝期限可以補交。
將 GitHub repository 更新後，請填寫遲交表單，助教才會去下載你的檔案來批改。
遲交的分數計算如下：

$$\text{score}_{\text{final}}(hr) = \begin{cases} \text{score}_{\text{original}} \times 0.985^{hr} & , hr \leq 72 \\ 0 & , hr > 72 \end{cases}$$

遲交時間以表單時間為主，請勿在上傳檔案前就填寫表單，助教會嚴加檢查。

作業批改完成後，若同學因為 script 錯誤導致被扣分，可以向助教請求補救。

開放時段原則為各作業公告成績後兩週內，補救時不得修改與 training/testing 演算法相關的程式碼，且補救後相關部份的成績會打八折。

任何作弊都是不被允許的，行為包含但不限於：

- 抄襲同學（包含以前的修課學生）的 code 或 report
- 上網搜尋或手動標記資料的 label
- 在 Kaggle 註冊多個帳號來上傳檔案
- 將資料上傳到先前的 Kaggle 競賽

若對作業或課程規定有任何問題，你可以

- 將疑問張貼於 FB 社團，你的問題也可能是別人的問題。
- 寫信至 ntu-ml-2020spring-ta@googlegroups.com，並請記得於標題以 [hwX] 註明作業編號。
 - 舉例來說，作業三的問題就必須用 [hw3] 作為信件標題開頭。
- 在每週的 TA 時間當面詢問。

