hw1

TAs

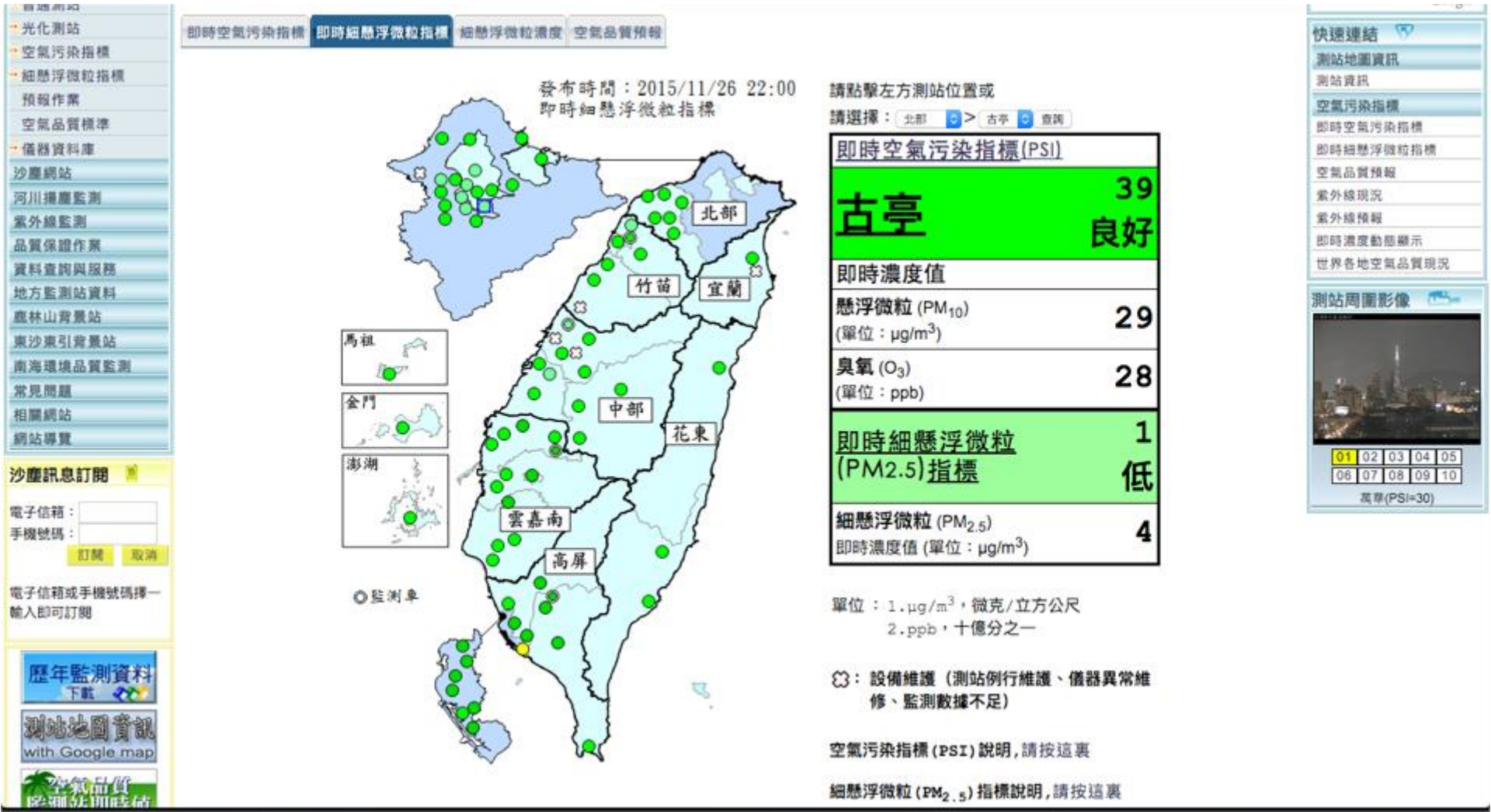
ml2016ta@gmail.com

Outline

---

1. hw1 介紹
2. test data
3. Kaggle
4. 配分，格式規定（限python, C/C++）
5. FAQ
6. Github（請自己看影片）

Task - Predict PM2.5



Data of each station

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	日期	測站	測項	0	1	2	3	4	5	6	7	8	9	10
2	2014/1/1	豐原	AMB_TEMP	14	14	14	13	12	12	12	12	15	17	20
3	2014/1/1	豐原	CH4	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8
4	2014/1/1	豐原	CO	0.51	0.41	0.39	0.37	0.35	0.3	0.37	0.47	0.78	0.74	0.59
5	2014/1/1	豐原	NMHC	0.2	0.15	0.13	0.12	0.11	0.06	0.1	0.13	0.26	0.23	0.2
6	2014/1/1	豐原	NO	0.9	0.6	0.5	1.7	1.8	1.5	1.9	2.2	6.6	7.9	4.2
7	2014/1/1	豐原	NO2	16	9.2	8.2	6.9	6.8	3.8	6.9	7.8	15	21	14
8	2014/1/1	豐原	NOx	17	9.8	8.7	8.6	8.5	5.3	8.8	9.9	22	29	18
9	2014/1/1	豐原	O3	16	30	27	23	24	28	24	22	21	29	44
10	2014/1/1	豐原	PM10	56	50	48	35	25	12	4	2	11	38	56
11	2014/1/1	豐原	PM2.5	26	39	36	35	31	28	25	20	19	30	41
12	2014/1/1	豐原	RAINFALL	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
13	2014/1/1	豐原	RH	77	68	67	74	72	73	74	73	66	56	45
14	2014/1/1	豐原	SO2	1.8	2	1.7	1.6	1.9	1.4	1.5	1.6	5.1	15	4.5
15	2014/1/1	豐原	THC	2	2	2	1.9	1.9	1.8	1.9	1.9	2.1	2	2
16	2014/1/1	豐原	WD_HR	37	80	57	76	110	106	101	104	124	46	241
17	2014/1/1	豐原	WIND_DIRE	35	79	2.4	55	94	116	106	94	232	153	283
18	2014/1/1	豐原	WIND_SPEE	1.4	1.8	1	0.6	1.7	2.5	2.5	2	0.6	0.8	1.6
19	2014/1/1	豐原	WS_HR	0.5	0.9	0.6	0.3	0.6	1.9	2	2	0.5	0.3	0.8
20	2014/1/2	豐原	AMB_TEMP	16	15	15	14	14	15	16	16	17	20	22
21	2014/1/2	豐原	CH4	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.7
22	2014/1/2	豐原	CO	0.26	0.25	0.28	0.27	0.24	0.26	0.34	0.56	0.79	1.05	0.3
23	2014/1/2	豐原	NMHC	0.06	0.05	0.06	0.05	0.05	0.07	0.09	0.19	0.31	0.5	0.13
24	2014/1/2	豐原	NO	1.4	1.1	1.3	1	1.2	1.1	1.6	8.4	17	17	2.2
25	2014/1/2	豐原	NO2	3.7	3.2	3.3	3.1	3.1	4.3	9.4	19	26	36	8.4
26	2014/1/2	豐原	NOx	5	4.3	4.7	4.1	4.3	5.5	11	27	43	53	11
27	2014/1/2	豐原	O3	39	38	39	39	34	31	30	18	17	24	46
28	2014/1/2	豐原	PM10	49	34	31	16	18	8	16	24	37	58	51

Data of each station

本次作業使用豐原站的觀測記錄，分成train set跟test set，train set是豐原站每個月的前20天所有資料。test set則是從豐原站剩下的資料中取樣出來。

train.csv：每個月前20天的完整資料。

test_X.csv：從剩下的10天資料中取樣出連續的10小時為一筆，前九小時的所有觀測數據當作feature，第十小時的PM2.5當作answer。一共取出240筆不重複的test data，請根據feauure預測這240筆的PM2.5。

test_X.csv

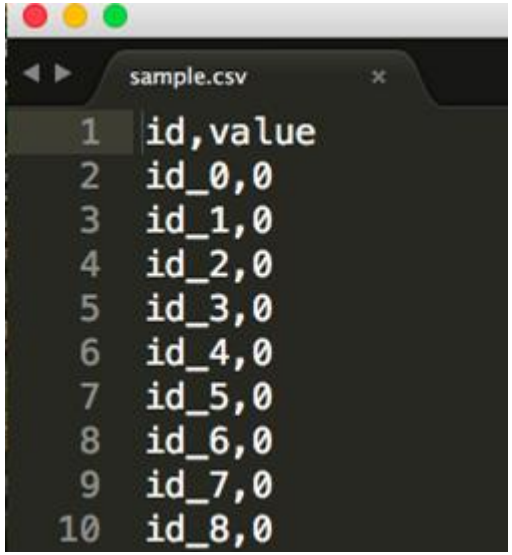
			1	2	3	4	5	6	7	8	9	
	A	B	C	D	E	F	G	H	I	J	K	L
1	id_0	AMB_TEM	15	14	14	13	13	13	13	13	12	
2	id_0	CH4	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8	
3	id_0	CO	0.36	0.35	0.34	0.33	0.33	0.34	0.34	0.37	0.42	
4	id_0	NMHC	0.11	0.09	0.09	0.1	0.1	0.1	0.1	0.11	0.12	
5	id_0	NO	0.6	0.4	0.3	0.3	0.3	0.7	0.8	0.8	0.9	
6	id_0	NO2	9.3	7.1	6.1	5.7	5.5	5.3	5.5	7.1	7.5	
7	id_0	NOx	9.9	7.5	6.4	5.9	5.8	6	6.2	7.8	8.4	
8	id_0	O3	36	44	45	44	44	44	43	40	38	
9	id_0	PM10	51	51	31	40	34	51	42	36	30	
10	id_0	PM2.5	27	13	24	29	41	30	29	27	28	
11	id_0	RAINFALL	NR	NR	NR	NR	NR	NR	NR	NR	NR	
12	id_0	RH	75	71	71	73	74	74	74	74	74	
13	id_0	SO2	1.2	1.2	1.2	1.6	1.5	1.5	1.5	1.6	1.6	
14	id_0	THC	1.9	1.8	1.8	1.9	1.9	1.9	1.9	1.9	1.9	
15	id_0	WD_HR	116	114	112	109	111	104	107	108	104	
16	id_0	WIND_DIR	115	113	105	102	106	106	112	113	106	
17	id_0	WIND_SPE	2.6	2.2	2	1.9	2.4	2.4	2.5	2.8	2	
18	id_0	WS_HR	2.1	2.4	2.2	1.9	2.3	2.3	2.5	2.5	2.3	
19	id_1	AMB_TEM	12	12	12	13	14	15	14	14	13	
20	id_1	CH4	1.8	1.8	1.9	1.9	1.8	1.8	1.8	1.8	1.8	
21	id_1	CO	0.46	0.58	0.64	0.63	0.58	0.52	0.52	0.51	0.49	
22	id_1	NMHC	0.06	0.1	0.12	0.13	0.15	0.12	0.11	0.11	0.14	
23	id_1	NO	0.5	0.7	1.5	1.9	1.8	1.7	1.8	1.8	1.8	

Submission format

預測test set中的240筆PM2.5，上傳至Kaggle。

- 上傳格式為csv
- 第一行必須是 id,value
- 第二行開始，每行分別為id及預測的數值，以逗點分隔。

範例格式:



Kaggle

---

1. https://inclass.kaggle.com/c/ml2016-pm2-5-prediction
2. 請至kaggle創帳號登入
3. 個人進行，不需組隊
4. 隊名: 學號_任意隊名（有修課的同學），旁聽同學請避免學號開頭。
5. 每日上傳上限5次
6. test set的240筆將被分為兩份，120筆public，120筆private
7. Leaderboard上顯示的是public set的分數，比賽之後可選擇兩筆答案作為計算private set的依據。
8. 最後的計分排名將以private set上為準。
9. kaggle deadline：2016/10/14 9:00:00 am (GMT+8)
10. github code & report deadline: 2016/10/14 21:00:00 pm (GMT+8)

作業規定

---

1. 請實作linear regression，方法限定使用Gradient Descent。
2. 請比較不同learning rate之結果。
3. 請比較有無加上regularization之結果。
4. 若想另外嘗試不同方法也可以，但仍須實作linear regression。
5. 不能使用現成function，只能使用numpy, scipy以及處理data會用到的東西 ex:pandas。

繳交格式

- Only Python & C/C++

1. Code in github ML2016/hw1/
2. Report.pdf in github ML2016/hw1/
3. Deadline: 2016/10/14 21:00:00 pm (GMT+8)

Github: ML2016/hw1/ 請包含Report.pdf，linear_regression.sh，kaggle_best.sh，以及所有需要的東西(train.csv, test_X.csv)

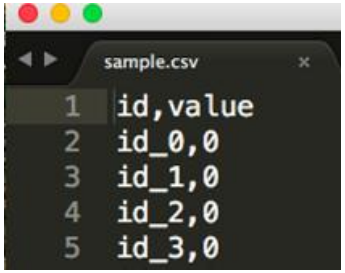
Usage:

./linear_regression.sh 輸出:linear_regression.csv

./kaggle_best.sh 輸出:kaggle_best.csv

Github表單: https://goo.gl/EZGBf8

#請每位修課同學務必填表單，hw0填過也請再填一次。



配分

Kaggle Rank (5%): top10% 5; top20% 4; top50% 3.5;
beyond baseline 3

Report.pdf (5%): PDF, less than 2 pages, not include code.

1. (1%) Linear regression function by Gradient Descent.
2. (1%) Describe your method. 因為我們沒限制你該怎麼做，所以請詳述方法 ex:怎麼取training feature (X,y).
3. (1%) Discussion on regularization.
4. (1%) Discussion on learning rate.
5. (1%) TA depend on your other discussion and detail.

Other policy:

任一script錯誤(0分)，若是格式錯誤，請來找助教修好(kaggle part*0.5)

遲交每24小時(*0.7); 遲交超過48小時不收，有特殊原因請洽助教。

遲交表單: https://goo.gl/DTH8Kp (kaggle無法遲交)

FAQ

如果只有做一個方法是否需要交兩份script?

A: 是。如果你只有做linear regression，kaggle上的分數也是linear regression的話，也麻煩交兩份script。

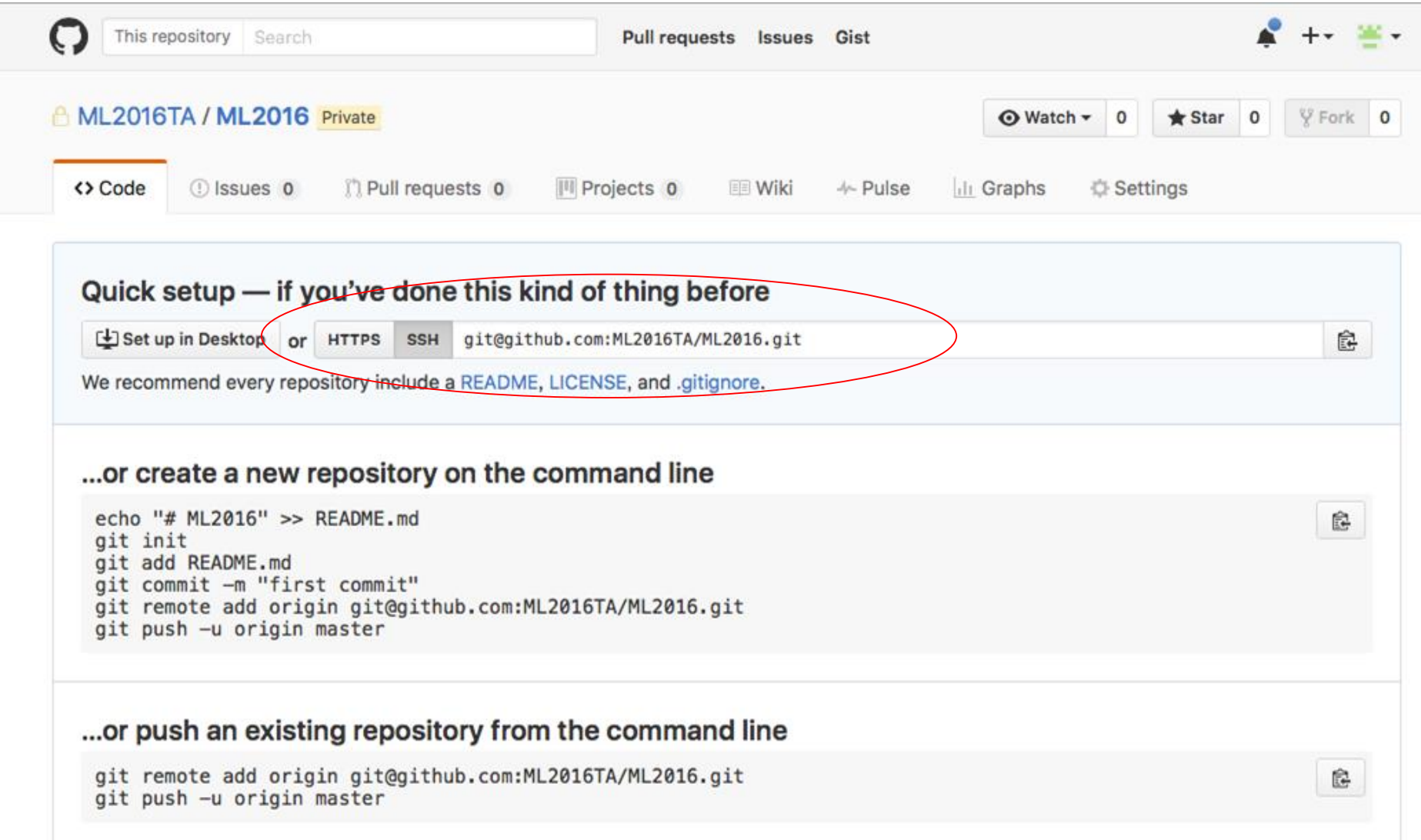
表單填錯怎麼辦?

A: 請直接重填即可，如果有重複的表單我們會以最新的表單為準。

提醒，表單只是要蒐集github repo url，你不需要每次git push都填一次表單，填一次即可。

再次提醒，請每位有選上課的同學都要填表單。

表單: Get Your Git Repo URL?



library限制之疑問?

A: 如果有使用到不知道能不能使用的library，請寫信跟助教確認並簡述用途，基本上如果是用來處理data的library都是可以的。

作業截止時間?

A: 再次提醒，

1. kaggle deadline：2016/10/14 9:00:00 am (GMT+8)
2. github code & report deadline: 2016/10/14 21:00:00 pm (GMT+8)

Github

(詳見hw0說明影片 連結)

開設 github 帳號

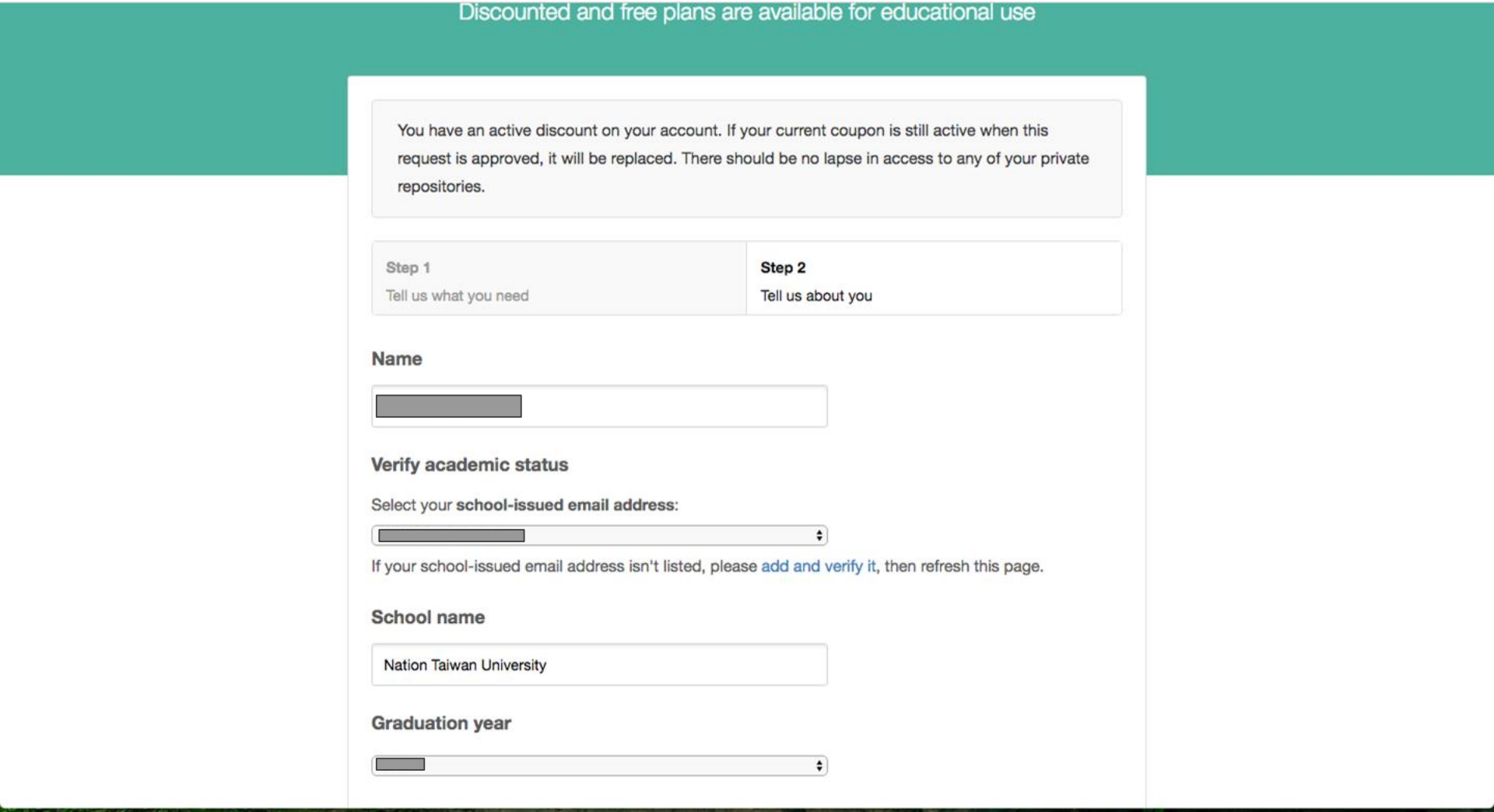
---

1. github: https://github.com/ 使用學校信箱開帳號
    a. 學校信箱可免費使用private功能
    b. 可綁定多個信箱
2. 申請學生版的附加功能
    a. 網址: https://education.github.com/
    b. 點選 Request a discount
    c. 輸入資料，靜候佳音

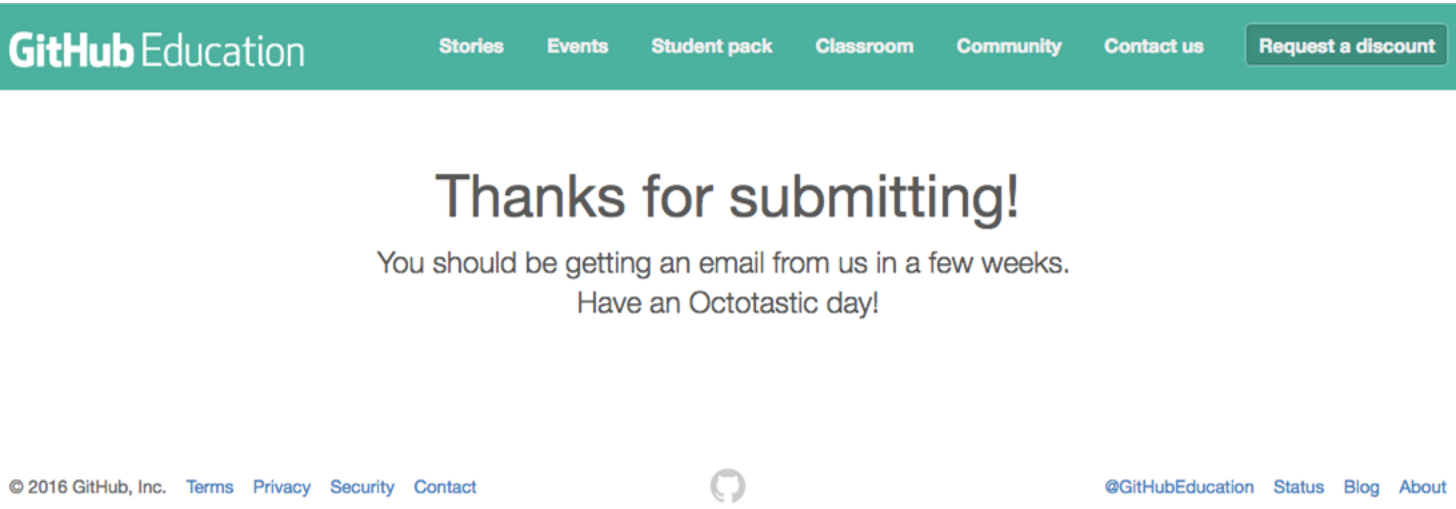
step 1. 進入網址


GitHub Education
Stories
Events
Student pack
Classroom
Community
Contact us
TEACH AND LEARN
BETTER, TOGETHER
Request a discount
STUDENT DEVELOPER PACK
Get the Student Developer Pack
Dozens of free resources from great companies to help students learn.
Get the pack
STORIES

step 2. 填入資料


Discounted and free plans are available for educational use
You have an active discount on your account. If your current coupon is still active when this request is approved, it will be replaced. There should be no lapse in access to any of your private repositories.
Step 1
Tell us what you need
Step 2
Tell us about you
Name
Verify academic status
Select your school-issued email address:
If your school-issued email address isn't listed, please add and verify it, then refresh this page.
School name
Nation Taiwan University
Graduation year

step 3. 靜候佳音


GitHub Education
Stories
Events
Student pack
Classroom
Community
Contact us
Request a discount
Thanks for submitting!
You should be getting an email from us in a few weeks.
Have an Octotastic day!
© 2016 GitHub, Inc.
Terms
Privacy
Security
Contact
@GitHubEducation
Status
Blog
About

作業繳交

---

1. New repository
    a. 請將名稱取為ML2016
    b. 往後所有的作業程式都會在這個路徑下被批改
    c. 權限請設為private
2. 將助教帳號加入存取權
    a. 名稱: ML2016TA

Create a new repository

A repository contains all the files for your project, including the revision history.

Owner **Repository name**

[redacted] / ML2016

Great repository names are short and memorable. Need inspiration? How about **fuzzy-giggle**.

Description (optional)

Public
Anyone can see this repository. You choose who can commit.

Private
You choose who can see and commit to this repository.

Initialize this repository with a README
This will let you immediately clone the repository to your computer. Skip this step if you're importing an existing repository.

Add .gitignore: **None** | Add a license: **None**

Create repository

This repository | Search

Pull requests | Issues | Gist

[redacted] / ML2016 Private

Watch 0 | Star 0 | Fork 0

Code | Issues 0 | Pull requests 0 | Projects 0 | Wiki | Pulse | Graphs | Settings

- Options
- **Collaborators**
- Webhooks
- Integrations & services
- Deploy keys

Collaborators — Push access to the repository

This repository doesn't have any collaborators yet. Use the form below to add a collaborator.

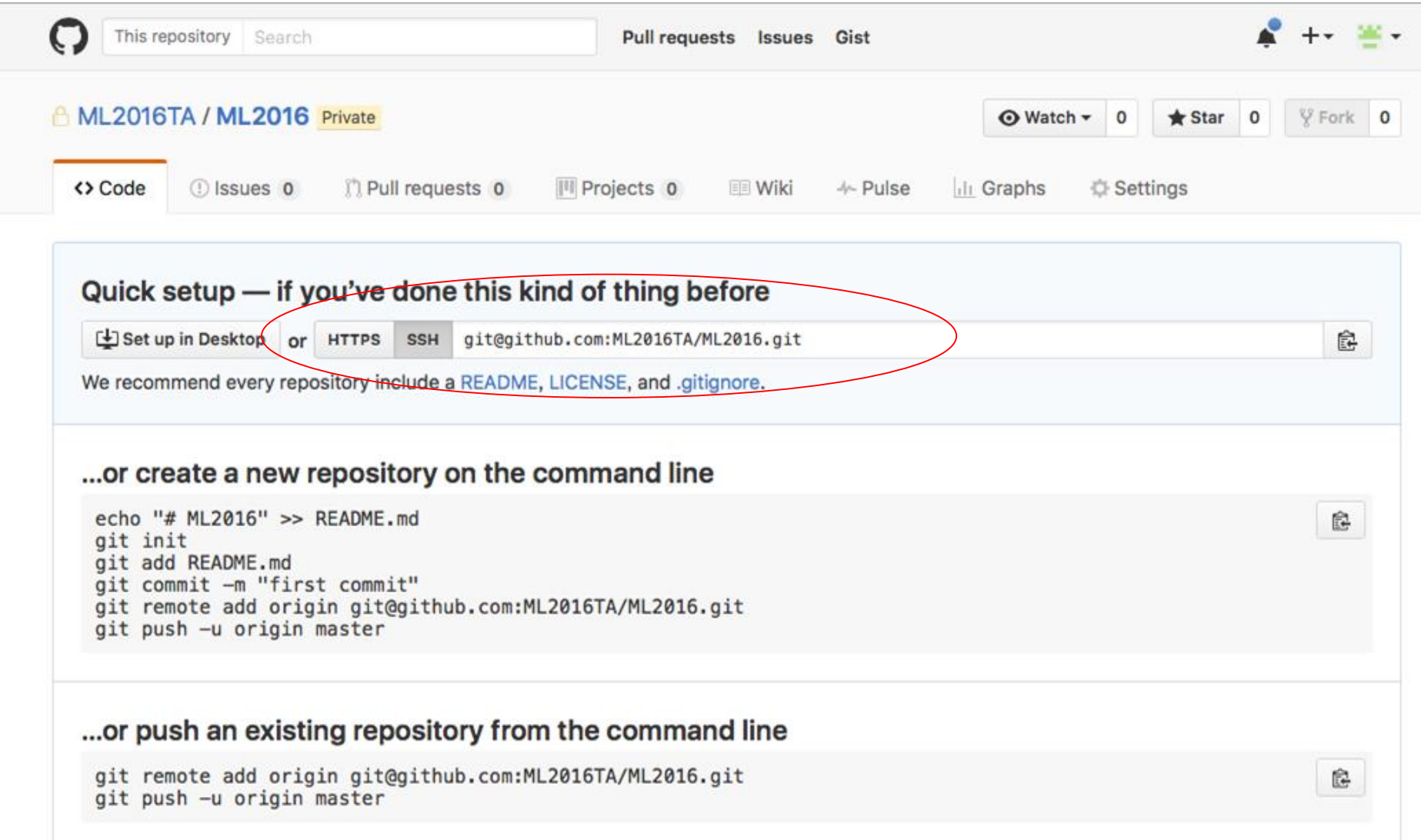
Search by username, full name or email address

ML2016TA

Add collaborator

ML2016TA

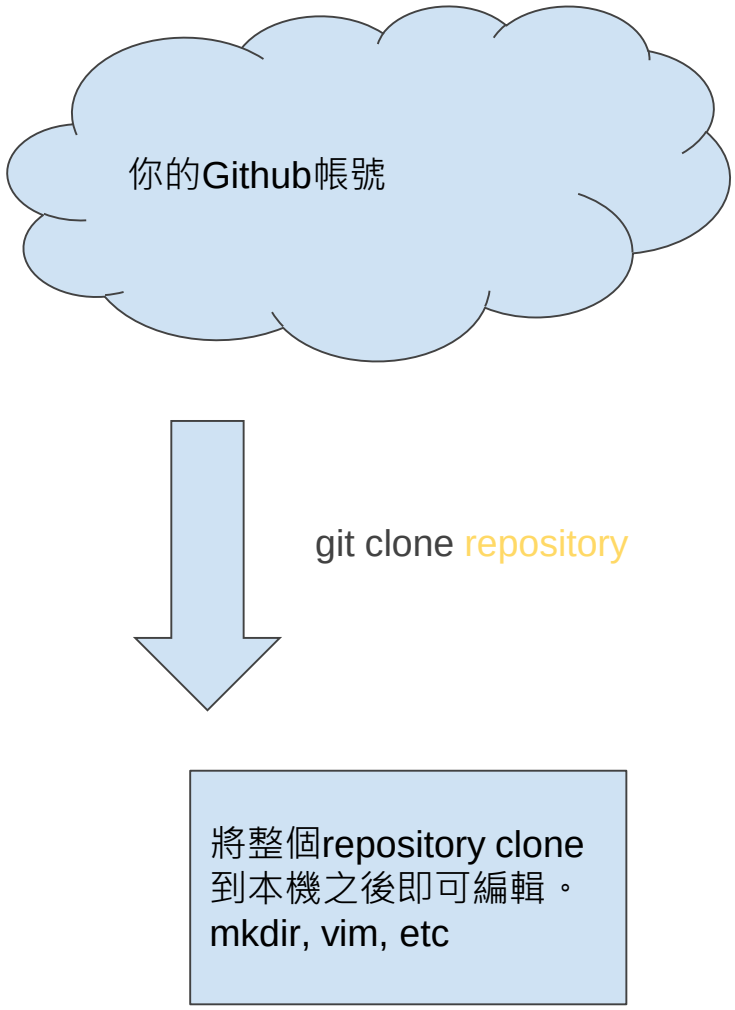
Get Your Git Repo URL



Github

---

1. Open your terminal.
2. git clone *git repo url*
3. cd ML2016
4. mkdir hw0
5. …
6. …



Github

---

1. git add xxx.py
2. git commit
3. git push
4. github網頁確認master是否已更新