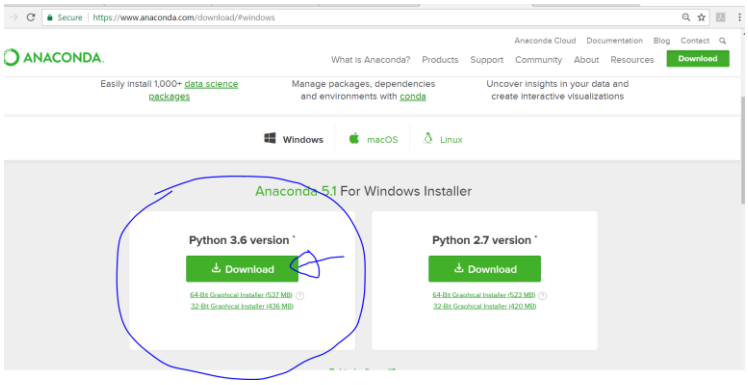
Introduction to Data Mining – Practical 1

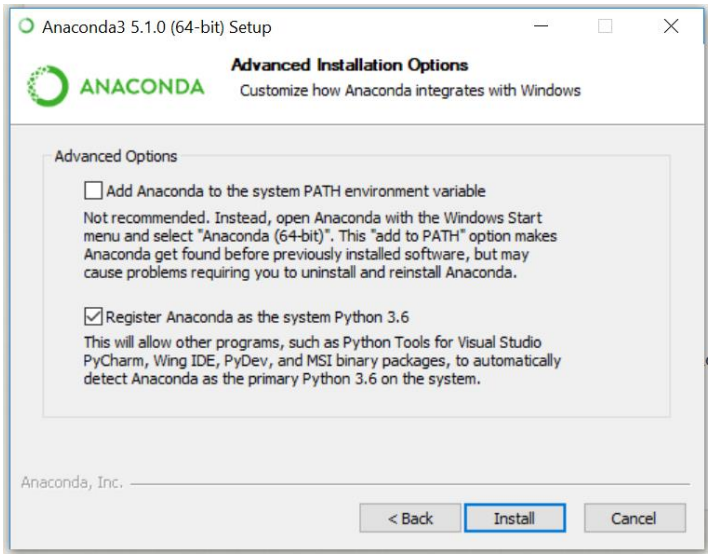
Installing Software

Python

1. Download and install Python 3.6. – we will use the Anaconda distribution.
2. You can download the Anaconda distribution (the Windows 64 bit version is used here – if you need to use a different version check with the tutor) from: https://www.anaconda.com/download



3. Use the default installation options, for example you don't need to add Anaconda to the path during installation:



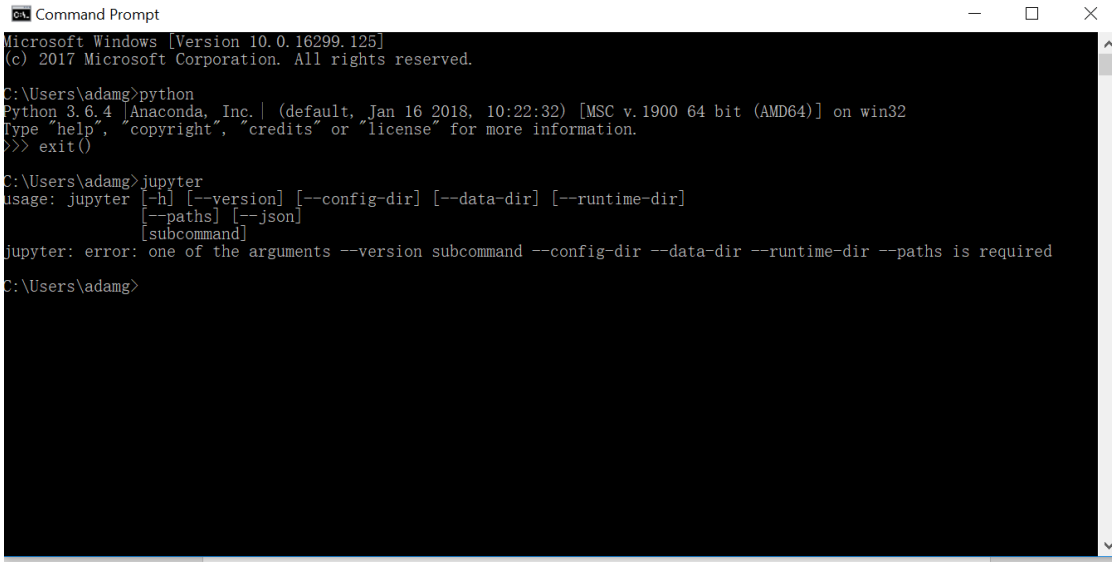
4. When you have installed Anaconda you will have both Python and Jupyter (IPython) on your computer
5. Open a windows command prompt and type "python":

6. Add python and Juypyter to the path variable:

SETX PATH "%PATH%;C:\ProgramData\Anaconda3\Scripts;C:\ProgramData\Anaconda3"

(note: your path might be different if you performed the installation for a user – check under c:/users/your_username)

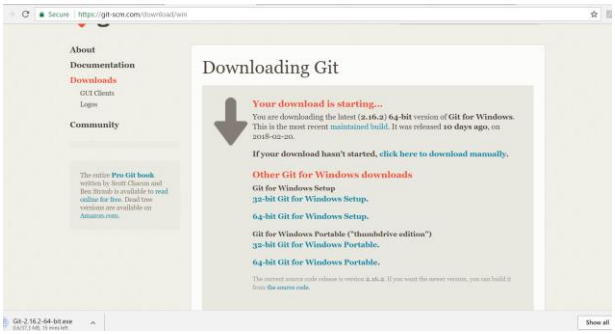
7. Close and reopen the windows command prompt you are able to run Python:



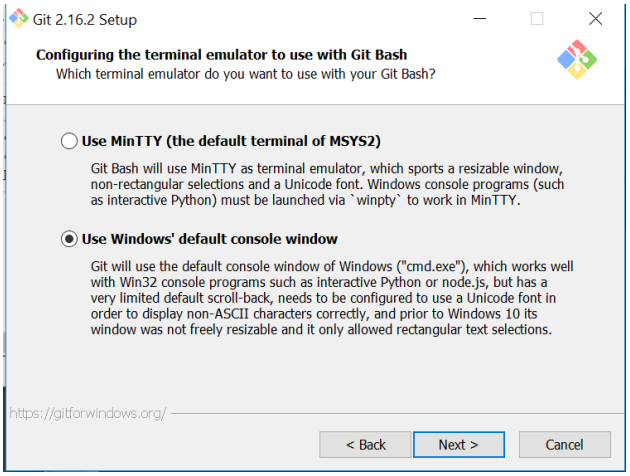
Git

Git is a version control system, it is used in this tutorial to access the assignment description and data. Don't worry if you don't fully understand it, more information will be provided as we progress in the course.

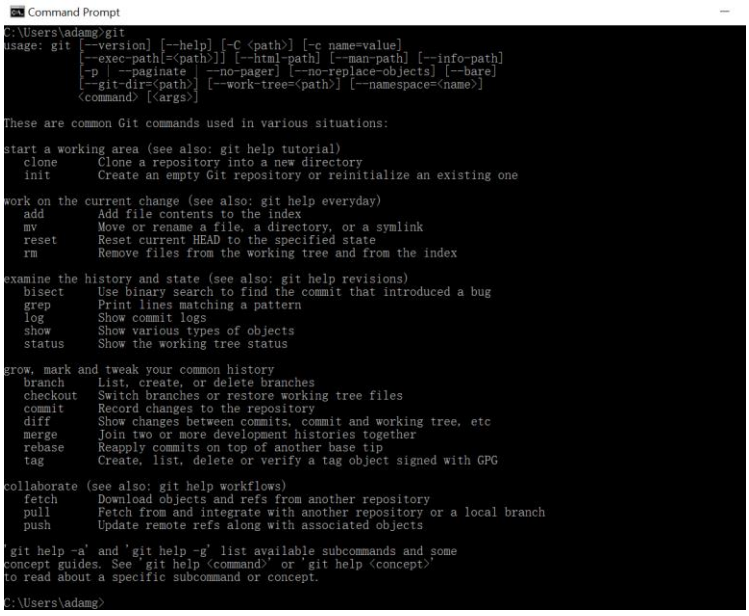
1. For an overview of GIT usage see the guide http://rogerdudler.github.io/git-guide/
2. Download Git from https://git-scm.com/downloads



3. Run the installer to setup git on your machine

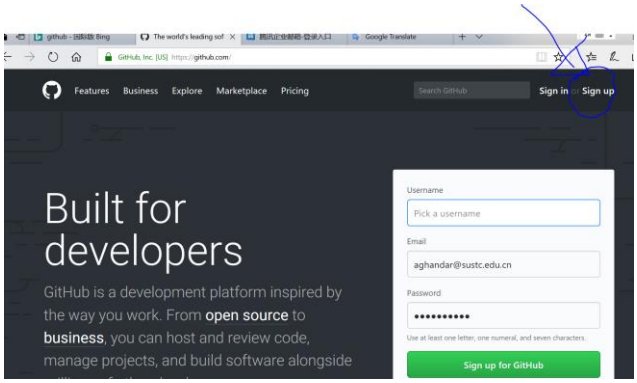


4. Close and reopen the command prompt and ensure you can run git:



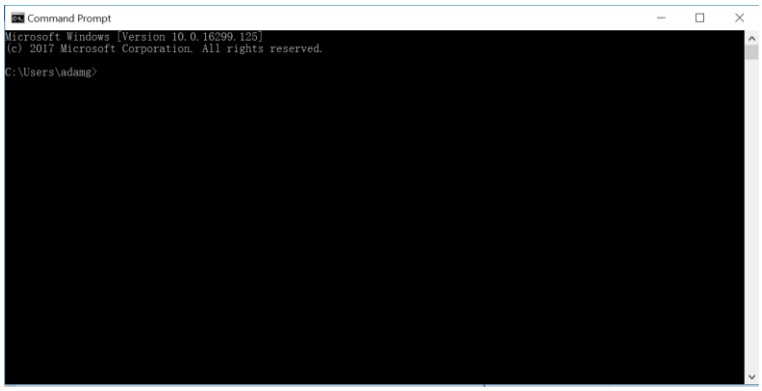
Create a Github account

Navigate to http://github.com and create an account (sign up).

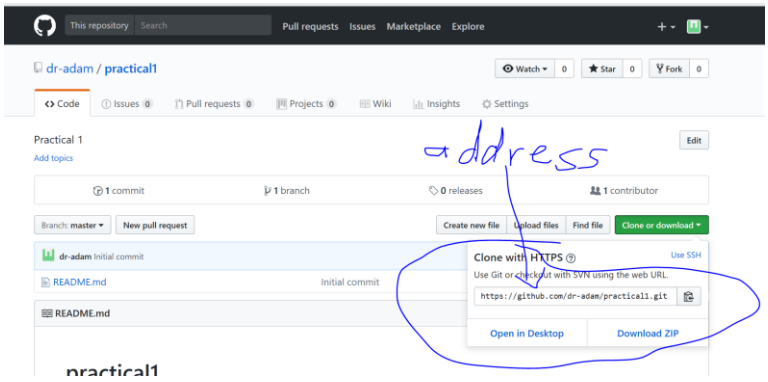


Clone the practical 1 repo to your local machine

1. Open a new cmd prompt.

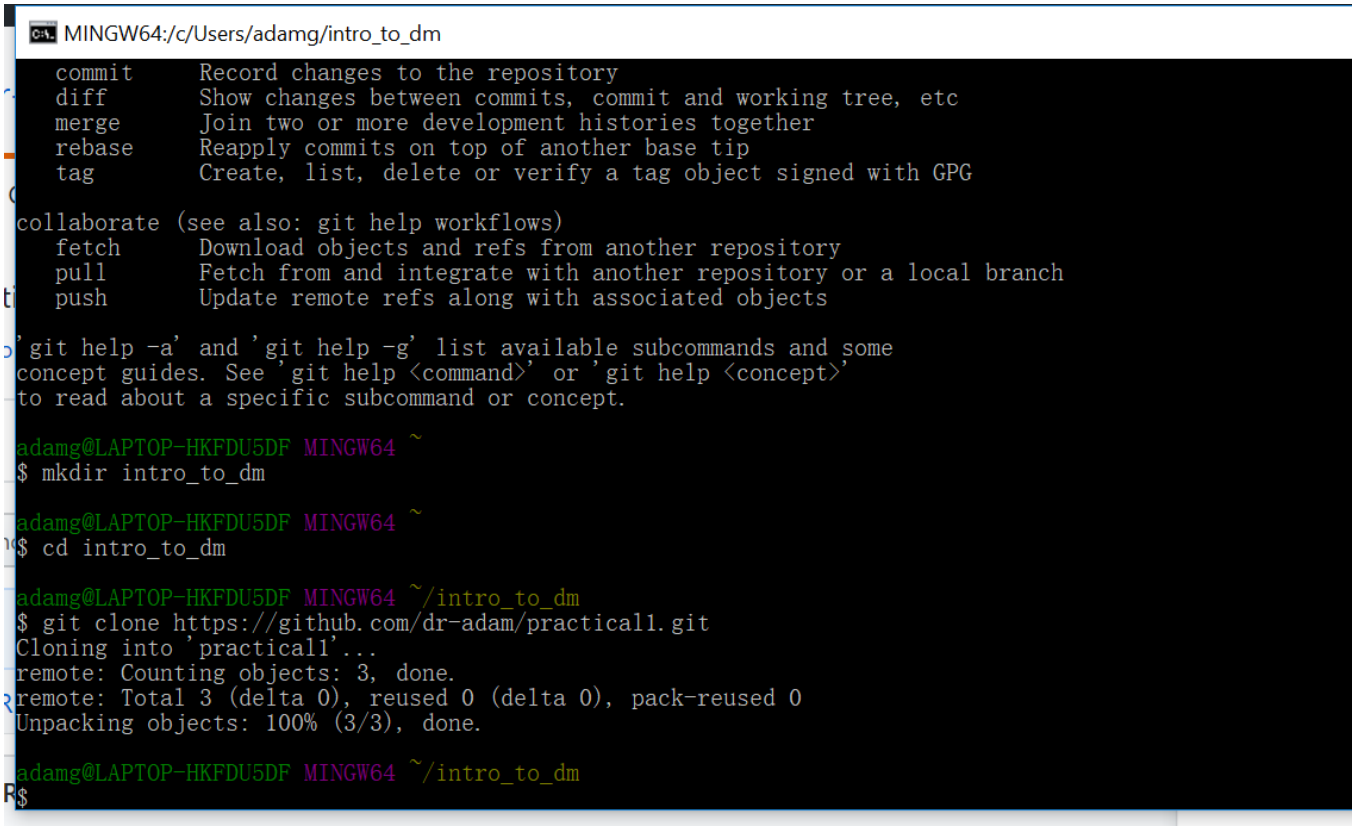


2. Open the assignment 1 repo in a web browser at https://github.com/dr-adam/practical1
3. Copy the address to the clipboard (right click your mouse):



4. Create a clone on your machine (click on "Clone or Download to obtain the address):
    a. Open a cmd prompt and type
        i. mkdir intro_to_dm
        ii. cd intro_to_dm
        iii. git clone https://github.com/dr-adam/practical1.git

b. This will create a new directory called intro_to_dm, move to that directory, and then clone the practical work sheet into the directory on your local machine:



5. Notice that we used the following unix commands:
    a. Change directory: ls
    b. Make new directory: mkdir <directory name>
    c. Change directory: cd
5. Again for an overview of using Git to access and clone a repository (what we just did) see http://rogerdudler.github.io/git-guide/

Running the tutorial

1. Now we will start the tutorial by opening the IPython notebook obtained in the previous section by downloading from the GitHub repo
2. Open a new command line terminal to the directory intro_to_dm
3. Type:
    a. $ jupyter notebook
    b. Then in the jupyter browser window navigate to the notepad prac1.ipynb

```
C:\Users\adamg\intro_to_dm\practical1>dir
 Volume in drive C is Windows
 Volume Serial Number is F0AC-4BD3

 Directory of C:\Users\adamg\intro_to_dm\practical1

03/03/2018  01:47 PM    <DIR>          .
03/03/2018  01:47 PM    <DIR>          ..
02/03/2018  03:50 PM    <DIR>          .ipynb_checkpoints
02/03/2018  04:05 PM           171,524 Introduction to Data Mining practical 1 setup.pdf
02/03/2018  04:49 PM           179,017 Introduction to Data Mining.pdf
03/03/2018  01:47 PM             6,352 prac1.ipynb
02/03/2018  04:07 PM               539 README.md
               4 File(s)        357,432 bytes
               3 Dir(s)  185,920,307,200 bytes free

C:\Users\adamg\intro_to_dm\practical1>jupyter notebook
```