

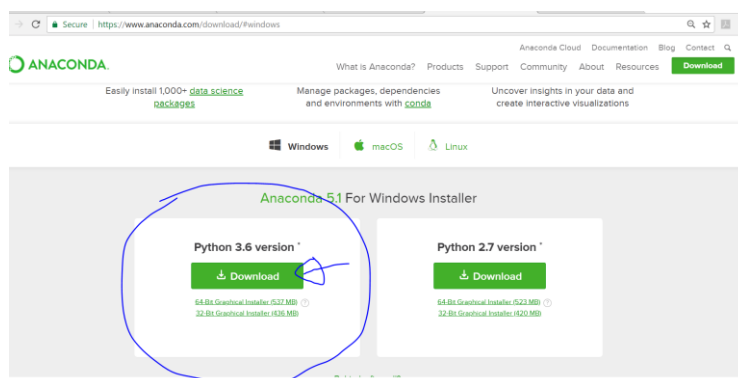
## Introduction to Data Mining – Practical 1

### Installing Software

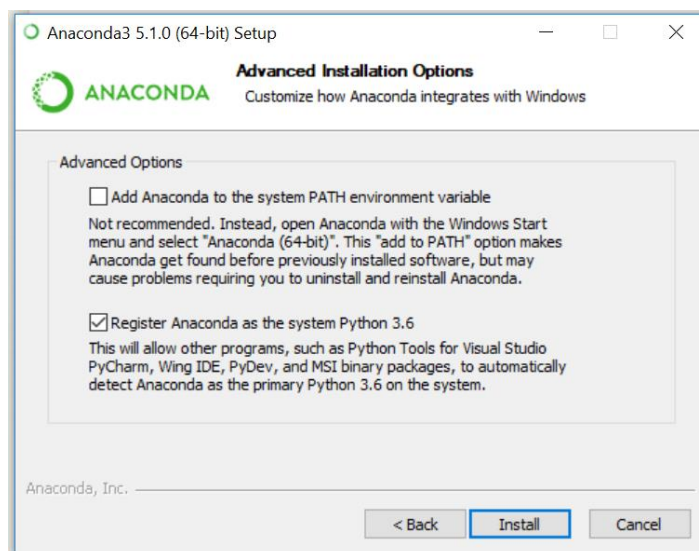
#### Python

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:

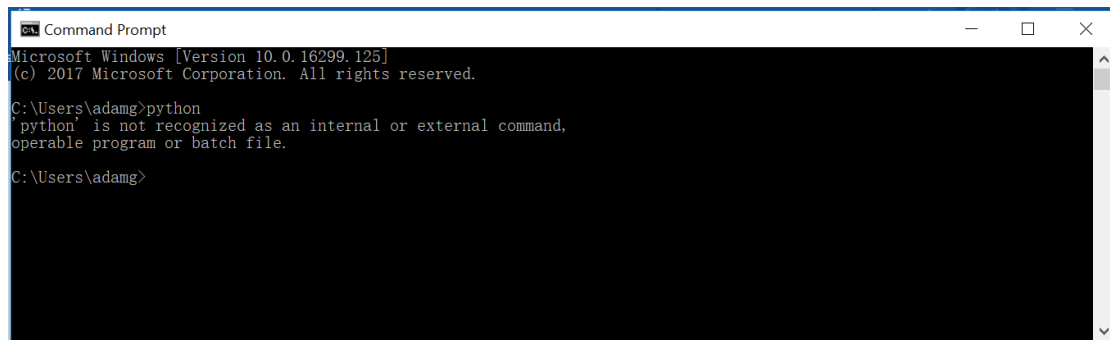
<https://www.anaconda.com/download>



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":



```
Command Prompt
Microsoft Windows [Version 10.0.16299.125]
(c) 2017 Microsoft Corporation. All rights reserved.

C:\Users\adamg>python
'python' is not recognized as an internal or external command,
operable program or batch file.

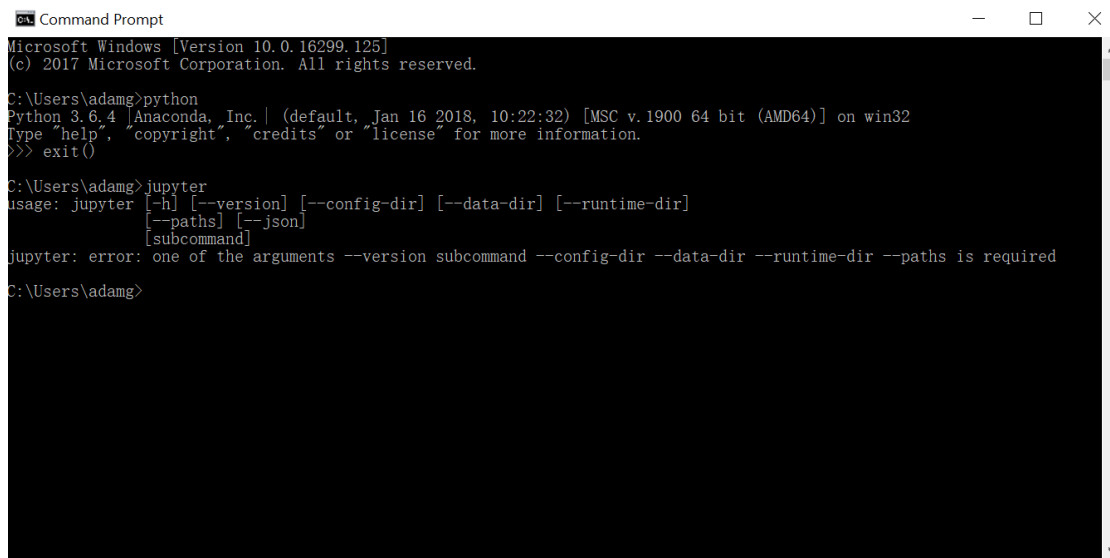
C:\Users\adamg>
```

6. Add python and Jupyter 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:



```
Command Prompt
Microsoft Windows [Version 10.0.16299.125]
(c) 2017 Microsoft Corporation. All rights reserved.

C:\Users\adamg>python
Python 3.6.4 |Anaconda, Inc.| (default, Jan 16 2018, 10:22:32) [MSC v.1900 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> exit()

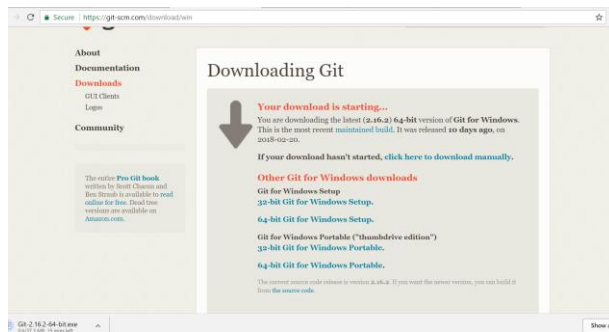
C:\Users\adamg>jupyter
usage: jupyter [-h] [--version] [--config-dir] [--data-dir] [--runtime-dir]
              [--paths] [--json]
              [subcommand]
jupyter: error: one of the arguments --version subcommand --config-dir --data-dir --runtime-dir --paths is required

C:\Users\adamg>
```

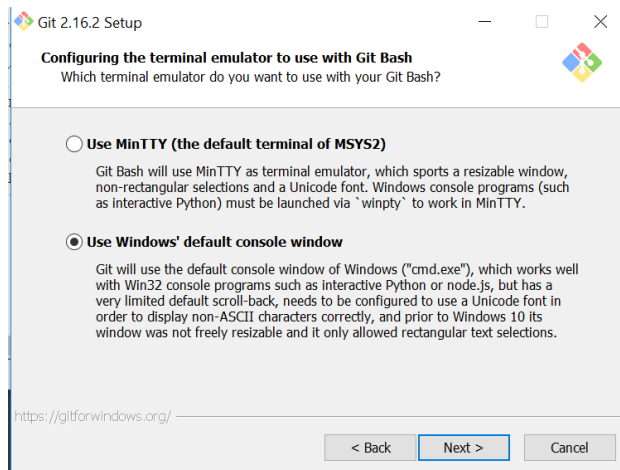
## Git

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:

```

C:\Users\adam>git
usage: git [--version] [--help] [-C <path>] [-c name=value]
       [--exec-path[=<path>]] [--html-path] [--man-path] [--info-path]
       [-p | --paginate] [-no-pager] [--no-replace-objects] [--bare]
       [--git-dir=<path>] [--work-tree=<path>] [--namespace=<name>]
       <command> [<args>]

These are common Git commands used in various situations:

start a working area (see also: git help tutorial)
  clone Clone a repository into a new directory
  init Create an empty Git repository or reinitialize an existing one

work on the current change (see also: git help everyday)
  add Add file contents to the index
  mv Move or rename a file, a directory, or a symlink
  reset Reset current HEAD to the specified state
  rm Remove files from the working tree and from the index

examine the history and state (see also: git help revisions)
  bisect Use binary search to find the commit that introduced a bug
  grep Print lines matching a pattern
  log Show commit logs
  show Show various types of objects
  status Show the working tree status

grow, mark and tweak your common history
  branch List, create, or delete branches
  checkout Switch branches or restore working tree files
  commit Record changes to the repository
  diff Show changes between commits, commit and working tree, etc
  merge Join two or more development histories together
  rebase Reapply commits on top of another base tip
  tag Create, list, delete or verify a tag object signed with GPG

collaborate (see also: git help workflows)
  fetch Download objects and refs from another repository
  pull Fetch from and integrate with another repository or a local branch
  push Update remote refs along with associated objects

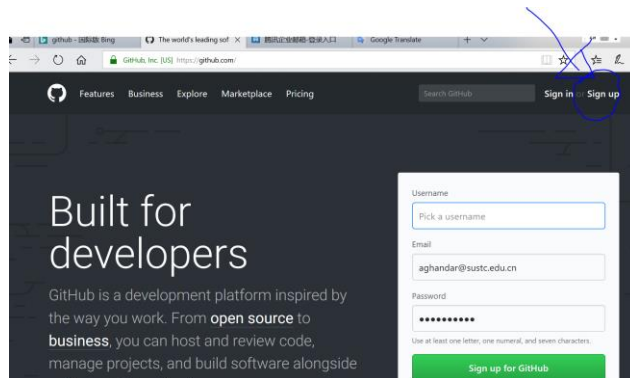
'git help -a' and 'git help -g' list available subcommands and some
concept guides. See 'git help <command>' or 'git help <concept>'
to read about a specific subcommand or concept.

C:\Users\adam>

```

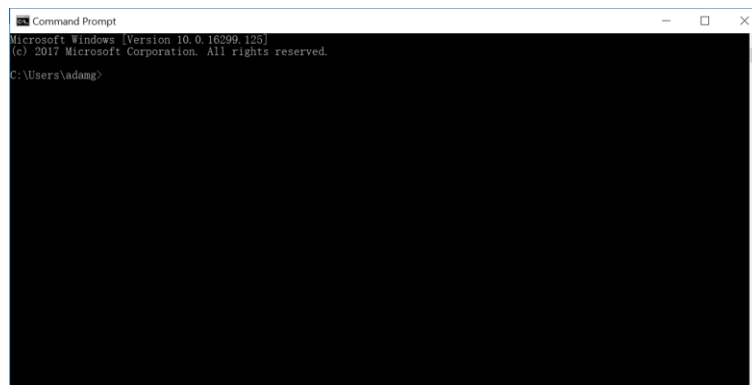
## Create a Github account

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

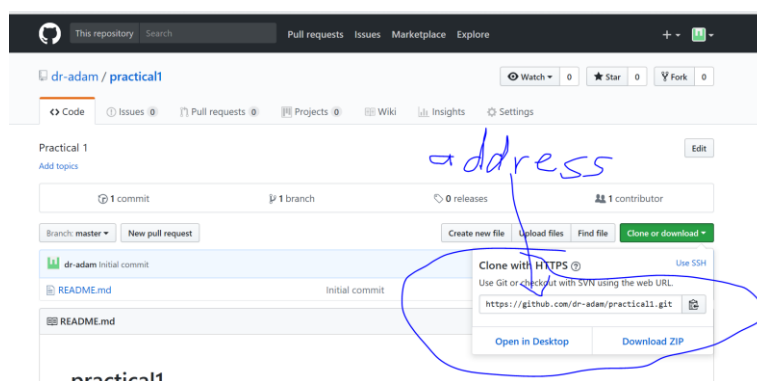


## Clone the practical 1 repo to your local machine

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:

```
CA: MINGW64:/c/Users/adamg/intro_to_dm
commit      Record changes to the repository
diff        Show changes between commits, commit and working tree, etc
merge       Join two or more development histories together
rebase      Reapply commits on top of another base tip
tag         Create, list, delete or verify a tag object signed with GPG

collaborate (see also: git help workflows)
fetch       Download objects and refs from another repository
pull        Fetch from and integrate with another repository or a local branch
push        Update remote refs along with associated objects

'git help -a' and 'git help -g' list available subcommands and some
concept guides. See 'git help <command>' or 'git help <concept>'
to read about a specific subcommand or concept.

adamg@LAPTOP-HKF5DU5DF MINGW64 ~
$ mkdir intro_to_dm

adamg@LAPTOP-HKF5DU5DF MINGW64 ~
$ cd intro_to_dm

adamg@LAPTOP-HKF5DU5DF MINGW64 ~/intro_to_dm
$ git clone https://github.com/dr-adam/practical1.git
Cloning into 'practical1'...
remote: Counting objects: 3, done.
remote: Total 3 (delta 0), reused 0 (delta 0), pack-reused 0
Unpacking objects: 100% (3/3), done.

adamg@LAPTOP-HKF5DU5DF MINGW64 ~/intro_to_dm
$
```

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
```

Home x

localhost:8888/tree

jupyter Logout

Files Running Clusters

Select items to perform actions on them. Upload New

	Name	Last Modified
<input checked="" type="checkbox"/>	prac1.ipynb	5 minutes ago
<input type="checkbox"/>	Introduction to Data Mining practical 1 setup.pdf	a day ago
<input type="checkbox"/>	Introduction to Data Mining.pdf	21 hours ago
<input type="checkbox"/>	README.md	a day ago