Instructions for participants -- Installing software and data

3-day workshop 'Big Data, Statistics and Machine Learning', Oct 2017

These instructions are available:

- as an ML Instructions PDF on **https://goo.gl/FDgxWS** < Reminder: DO NOT USE IN THE NGD environment>
- or as a Notes document http://goto/Notes://DomCO1510/CA2573310019894E//58A709BFCA3B0123CA2581B500829737

A good idea to download the PDF onto your laptop environment in a place that is easy to find. Is also available as a Lotus Notes document in case you prefer to print out (suggest colour).

Contents:

REMEMBER -- DOWNLOADS MUST BE ON YOUR ABS PHYSICAL LAPTOP environment –

DO NOT download into the NGD virtual environment

The laptop environment is usually a yellow background, (the NGD is usually a blue background -- NOT this environment).

General notes:

- Everything will be either downloaded from a software website (freeware) or from the googledrive **https://goo.gl/FDgxWS**
- Use a mouse for this set-up, will probably not need a mouse for the training course itself.
- Wireless may be slow for some of the downloads.
- Check the space available on your ABS Laptop's C: drive -- you should have plenty of space but worth checking (can see this in File Explorer).
- For copy and paste may need to right-click to select copy/paste (rather than Ctrl+c / Ctrl+v).
- <Should be able to do this using internet at work.>

#1 Download 7-zip, Notepad++ & Chrome ONTO YOUR PHYSICAL laptop environment

Check to see whether your ABS laptop has a text editor, zipping/unzipping and spreadsheet software.

If not we suggest **7-zip Portable** , **Notepad++ Portable** -- https://portableapps.com/apps/

Will need to download & install. (You may prefer non-portable, which is also available on a different site. LibreOffice Portable is also available but non-essential.)

How to use 7-zip portable: To unzip files using the portable version, need to open up 7-zip first and access the file from within 7-zip. (You may prefer the non-portable version or different zipping software).
Remember to always open the .zip file before extracting its contents.

Will probably end up here -- C:\Users\Your Name\Downloads\

Also -- **Chrome** is the recommended browser for anaconda, so check to see if this is already on your laptop. Make this your standard browser on the laptop. (*Open a Chrome window. Open settings menu via the '3 verticle dots' button on the top right. Scroll down to the 'Default browser' and click on the MAKE DEFAULT button and select Google Chrome.)*

#2 Download and set up -- R, RStudio (and packages) ONTO YOUR PHYSICAL laptop environ.
Note: needs to be the 64bit version for Windows. There are free open source versions available (as per below instructions).
Software Prerequisites:
#2.1 Download R: **https://mirror.aarnet.edu.au/pub/CRAN/** -- select **Download R for Windows** and then select **install R for the first time**.
Install R on laptop (64bit). When installing, 4 checkboxes will appear -- Uncheck the option for 32bit.
(Will probably get a message "Do you want to allow the following program from an unknown publisher to make changes to this computer? R-3.4.2-win.exe ")
(HINT: If you have any issues with this website, there are other mirrors to select from -- try the CSIRO mirror in on this site https://cran.r-project/mirrors.html)

#2.2 Download
RStudio: **https://www.rstudio.com/products/rstudio/download/#download**
Install RStudio (i.e. run for the first time)

#2.3 R packages to be installed with this R command: Enter line by line from R (check successful before running the next line – i.e. no Warning message).
(Hint: You may need to select a CRAN Mirror again the first line) --
install.packages("googledrive")
install.packages("tidyverse")
install.packages("rpart")
install.packages("knitr")
install.packages("rpart.plot")
install.packages("randomForest")
install.packages("caret")
install.packages("e1071")
install.packages("cluster")
install.packages("purrr")
install.packages("VIM")
install.packages("recommenderlab")
install.packages("lubridate")
install.packages("rmarkdown")

#3 Download and set up -- Anaconda3 (python and packages) ONTO YOUR PHYSICAL laptop environment
(Hint: This step takes roughly half an hour).

#3.1 Create a new directory on your laptop C: drive called Training -- **C:\Training**

#3.2 Download ANACONDA3 version 4.2.0 (64bit Windows) -- **https://repo.continuum.io/archive/Anaconda3-4.2.0-Windows-x86_64.exe**
Important: It has to be Anaconda3 NOT 2, and it needs to be this particular version as this is the version that has been tested - DO NOT download the latest version.
Important: If you do happen to download the wrong version, then please follow the 'Uninstall' instructions at the end of this document -- Using a different process can cause problems.

#3.3 Find the directory where anaconda has been saved to (probably your Downloads directory C:\Users\Your Name\Downloads) and double-click the Anaconda executable file called "**Anaconda3-4.2.0-Windows-x86_64**".
Anaconda will start installing -- make sure to change the default location to be **C:\Training\Anaconda3** (IMPORTANT: type it in rather than using browse) -- and at the end of the installation process Untick the box "Learn more about Anaconda Cloud" -- (installation takes about ~10min).
(Note: It HAS TO BE IN THIS DIRECTORY -- NOT THE ORIGINAL DEFAULT DIRECTORY -- WILL HAVE TO UNINSTALL AND START AGAIN.)

#3.4 Check that is has installed by opening **Anaconda Navigator** --
(Hint: as for all programs, can be found by pressing the Windows button on the bottom left corner of your screen, then pressing the downwards arrow that appears nearby at the bottom of the screen. You may need to scroll across the page to find all the Anaconda3 items grouped together to get to Anaconda Navigator - click once to open).
It may take 1-2 minutes to start up the first time - a green circular logo will appear in the centre of the screen.
A dialog box will open up once installed (possibly in the background or on one of your other screens) -- find it and Untick the box "Yes ... I'd like to etc." and press the button "K, and don't ask me again".
Once this has opened, close Anaconda Navigator before moving to the next steps.
IF YOU FIND THE DIALOG BOX CLOSES STRAIGHT AWAY THEN JUST MAKE A NOTE OF IT AND PLEASE MOVE ON TO THE NEXT STEPS.

#3.5 Download the **Tensorflow** folder from **https://goo.gl/FDgxWS**
Important: Download the whole folder all together -- hover over Tensorflow, right-click and select Download (rather than each file individually) - this will download the entire zip file. Once downloaded, find the folder on your laptop and do the following:

- use 7-Zip to open **.conda.zip** and extract into **C:\Users\Your Name**
- use 7-Zip to open **tensorflow.zip** and extract into **C:\Training\Anaconda3\envs** (Hint: this may take up to 9 minutes)
  You should have a directory called **tensorflow** under **C:\Training\Anaconda3\envs** which contains a number of other directories & files (67 items) . If not, make sure you have extracted the zip correctly. You shouldn't have **C:\Training\Anaconda3\envs\tensorflow\tensorflow.**
- copy the shortcut files called **IPython (tensorflow)** and **Jupyter Notebook (tensorflow)** into the following directory ... including the (64-bit) part on the end ... (also, remember to substitute your name):
  C:\Users\Your Name\AppData\Roaming\Microsoft\Windows\Start Menu\Programs\Anaconda3 (64-bit)

#3.6 Open **Jupyter Notebook (tensorflow)** (the same way we opened Anaconda Navigator in #3.4) *NOTE: NOT the one that just says Jupyter Notebook.* Both a command line window and a browser window should open (it takes a minute).
(Hint: The command line window will open up blank and then after a minute text will appear and the browser window will open. If you are using multiple screens they may open up on different screens so try to find both of them. If the browser window doesn't open, then close down the command line window (Ctrl+c Ctrl+c then press the X in the top right corner of the command line window -- Note if the Ctrl+c Ctrl+c doesn't work then can just press the X to shutdown). Then try opening up Jupyter Notebook again.)
In the browser window, press the **New** button in the top right corner (Underneath the Logout button) -- and select the Notebook called **Python [conda env:tensorflow]**
(Hint: A new browser window tab will open up, and may initially show an error message which will disappear in a minute.)
As a quick check - once the window has loaded up, type in the command **import tensorflow as tf** and press the play/pause button (small triangle button). If successful nothing will happen -- If unsuccessful there will be an error message.

---------- OPTIONAL CHECK ----------
As another quick check - enter the following lines pressing Enter after each line:
import tensorflow as tf
hello = tf.constant('Hello, TensorFlow!')
sess = tf.Session()
print(sess.run(hello))

Then press the play/pause button again.
If the installation has worked correctly, the system should output the following (we don't know where the "b" comes from though ☺): bHello, TensorFlow!

Close Jupyter Notebook -- both the command line window *(Ctrl+C Ctrl+C -- it can take a minute. Note if the Ctrl+c Ctrl+c doesn't work then just press the X to shutdown)*as well as the browser windows.
(Hint: As before, to close the command line window use Ctrl+C Ctrl+C and press X -- it can take a minute. Note if the Ctrl+c Ctrl+c doesn't work then just press the X to shutdown).

(If decide to use the command line version of Anaconda rather than Jupyter Notebook ... see instructions below.)

For info: Some help documentation on anaconda is available within the program -- Re-open Anaconda Navigator (see instructions further above), a Documentation button is on the bottom right.

#4a Download the data/models ON YOUR PHYSICAL laptop environment

After installing software prerequesits...
#4b.1 In the googledrive **https://goo.gl/FDgxWS** you will see a number of folders called **Prac1** through to **Prac6**.
Using the below instructions we are going to download these folders to C:\Downloads and then unzip each Prac1 to Prac6 folder underneath C:\Training.

- Check to see which of these folders have files in them (It may just be Prac1 to Prac3 to start with).
- For each of the folders that contain files -- download each of these folders to C:\Downloads (or wherever you can find it again). Make sure to download each folder in it's entirety.
- Open up explorer and go to C:\Downloads and find the new folders. (*HINT: The name of the folders will now include a date-time stamp – e.g. look something like* ***Prac2-20171019etcetc-001****.*)
- Open each folder and copy the file within into **C:\Training** (Use drag-n-drop).
  e.g. Open up folder ***Prac2-20171019etcetc****, drag-n-drop the folder called* ***Prac2*** *into the C:\Training\ and repeat for the remaining folders.*

Check that the folders look correct – e.g. C:\Training\Prac1

#4b Test Practical Session 1 - Test Run

- Check the C:\Training\Prac1 folder contains the following*:*
  - *two files ABS_ML1.Rmd and fetch_data.R*
  - *'data' directory with RTAData_MM_raw.csv*
  - *'figs' directory with m4_times.jpeg*
- Open RStudio -- there are 3 main windows -- in the bottom-right-window press **Files** and the prac1 files should already be visible (If not, navigate to the C:\Training\Prac1 directory) -- select **ABS_ML1.Rmd**
- ABS_ML1.Rmd will open as text/code in the large left hand side window. Some code will already be 'highlighted' ... press the green side-ways pointing arrow to the right to run the code.
  <If you have time extra time -- read through, executing code chunks as you reach them –probably part of the first session of the workshop.>

#5 Download and test the python file quick_test ONTO YOUR PHYSICAL laptop environment

#5.1 Open the **Tensorflow** folder in **https://goo.gl/FDgxWS** and download the file **quick_test.py** and the entire **MNIST_data** folder.
Depending on whether Chrome is your standard then one of the following approaches will work:
(i) Hover over the MNIST_data folder and select DOWNLOAD from the menu that appears.
ALTERNATIVELY
(ii) Hover over the small picture of the MNIST_data folder – it should become a down arrow – press directly on the down arrow to download
ALTERNATIVELY
(iii) Select the MNIST_data folder to open up the folder, and press the DOWNLOAD ALL button on the top-right – Save to a directory of your choice (e.g. C:\Downloads).

#5.2 Go to C:\downloads (or wherever you downloaded the files to).

- Drag and drop the file **quick_test.py** into **C:\Training** directory.
- Open the folder **MNIST_data-2017etcetc** <not actually etcetc>– and drag-n-drop the folder **MNIST_data** straight into the **C:\Training** directory.

#5.3 Re-open **Jupyter Notebook (tensorflow)** (as outlined in step #3.4 – remember it may take a minute for both the command line window and the browser windows to open).

The browser window should automatically start in the Files tab showing C:\Training (and you should see the Anaconda3, MNIST_data and quick_test.py at least – If you don't then please follow the below workaround).
UPDATE: If the browser doesn't open at C:\Training then will also need to do the following:

- Going into the 'Jupyter Notebook (tensorflow)' shortcut's properties dialog. (This is done by finding the Jupyter Notebook (tensorflow) as outlined in step #3.4; hovering over and right-click to bring up a menu, select 'Open file location'. This will bring up an explorer window and 'Jupyter Notebook (tensorflow) should already be ticked/highlighted – right click and select properties).
- Going to the Target field in the Shortcut tab,
- Replacing the "%USERPROFILE%" part of the text with "C:\Training".

(The files on the googledrive will be changed shortly to ensure that it is C:\Training – but just in case it isn't ready by the time you are running the installation, then please use this workaround.)

#5.4 Click on the item **quick_test.py** which will open up some test code into a new window – this is just an editor window (can't run from here).
Will need to copy and paste all of the code from quick_test.py into a new code window (follow below instructions to open a new code window):

Back in the original browser window, press the **New** button in the top right corner (Underneath the Logout button) -- and select the Notebook called **Python [conda env:tensorflow]**
Hint: A new browser window tab will open up, and may initially show an error message which will disappear in a minute. You should then see a single line showing
In [ ] : _____
Copy and paste all of the code from quick_test.py into the above single line In [ ]: _____.
Once copied, ensure the curser is somewhere in that code box and press the run-pause button *(it looks like a rightward-pointing triangle with a vertical line – next to the stop button*) – It will take around 20seconds.

If successful, you will see something along the lines of the below (it will be slightly different):

```
Extracting MNIST_data/train-images-idx3-ubyte.gz
Extracting MNIST_data/train-labels-idx1-ubyte.gz
Extracting MNIST_data/t10k-images-idx3-ubyte.gz
Extracting MNIST_data/t10k-labels-idx1-ubyte.gz
step 0, training accuracy 0.51
step 100, training accuracy 0.94
step 200, training accuracy 0.9
step 300, training accuracy 0.91
step 400, training accuracy 0.93
step 500, training accuracy 0.94
step 600, training accuracy 0.92
step 700, training accuracy 0.94
step 800, training accuracy 0.95
step 900, training accuracy 0.93
0.919
```

If decide to use the command line version of Anaconda (ie. rather than Jupyter Notebook).

Run **Anaconda Prompt** (As before, this can be found by pressing the windows button on the bottom right corner of your screen, then pressing the downwards arrow that appears nearby at the bottom of the page. You may need to scroll across the page to find all the Anaconda3 items grouped together to get to Anaconda Prompt - click once to open.)
A command line interface called Anaconda Prompt will open up, showing the environment and the directory in this format **<environment> directory** For example **<C:\Training\Anaconda3> C:\Users\Your Name**
We want to change the environment -- so type in: **activate tensorflow**
The command line should now say**:** **<tensorflow> C:\Users\Your Name**

A quick check -- In the Anaconda Prompt command line, type in **python**
The command line will now begin with >>>
Enter the following lines -- one at a time -- and press Enter after each line (wait until the >>> on the next line appears before typing in the next line):
import tensorflow as tf *<This may take a minute until the >>> appears on the next line. >*
hello = tf.constant('Hello, TensorFlow!')
sess = tf.Session() *<Don't worry about the messages after pressing Enter for this line of code. >*
print(sess.run(hello))
If the installation has worked correctly, the system should output the following:
Hello, TensorFlow!

To get out of the python environment again -- press **Ctrl** + **z** then press **Enter**.
Close the Anaconda Prompt command line window.

UNINSTALL Instructions
If you have problems with the install then you might have to uninstall anaconda and reinstall it again.
Open the Control Panel (Start Menu then navigate over to the right)
Then select “Programs and Features”
Scroll down until you find “Python 3.5.2 (Anaconda3 4.2.0 64-bit)”
Select it, then click Uninstall on the bar above.
Once this finishes (will take a few minutes), you can reinstall using the instructions above.