CELLSTRAT

Machine Learning Certification Program

Presenter :- Vivek Singhal, AI Data Scientist

PRE-WORKSHOP INSTRUCTIONS

You can do ANACONDA based install as in the associated Installation guide or you can do python command line based install (PIP/PIP3) in Command Prompt (CMD) of the laptop.

Pre-preparation for hands-on programming tutorials

Please complete these steps BEFORE the class.

1) Install Python, Sckikit-Learn and TensorFlow on your laptop :- (TensorFlow is NOT NEEDED for ML Certification)

   For a windows install, your laptop must be i5 / i7 intel processor or equivalent AMD processor.

   Fundamentally we need Python, Scikit-Learn, Matpotlib, Numpy and TensorFlow installed. Programming editor is your choice. More details below. Don't forget to update Windows PATH variable so these installs can be detected at run time.

   Online instructions to install TensorFlow :
   https://www.tensorflow.org/install

   a. Make sure that Python 3.5.x is installed and is in path first (Tensorflow will not work with Python2.x neither with Python 3.6). In Windows check echo %PATH% on windows command prompt and ensure that Python 3.5.x is found first. You can do "which python" to check which version of python is used. Where to install Python 3.5.2 from :
      https://www.python.org/downloads/release/python-352/
      i. AMD64 version Python will work with Intel64 also.
   b. Follow detailed instructions on the website
      https://www.tensorflow.org/install/ for installation.
   c. You will be installing TensorFlow and Python library both (unless your laptop already has these installed).
   d. For laptop Windows install, unless you have compliant Nvidia GPU processors, **we recommend you set up CPU version of TensorFlow**. If you chose to install GPU version, you must meet the pre-requisites listed on the website.
   e. "native pip" based installed is preferred over Anaconda install.

f. It will also ask you to install Python 3.5.x from python.org (https://www.python.org/downloads/release/python-352/). Go ahead and install this. Do select on "Add Python 3.5 to PATH". Also Do select TCL / TKINTER under Optional modules of Python, when installing Python – TKINTER is the GUI interface for Python.
g. During Python install DO CLICK on the option to install TCL / TKINTER. If forgot in first pass, and python already installed, then click on Python install file AGAIN and JUST click on TCL / Kinter only to install and not the overall Python again.
h. Make sure your Windows %PATH% variable is able to point to Python install. If not, update Windows environment variable %PATH% (under Control Panel->System->Advanced->Environment Variables).
i. If you installed GPU version, then Nvidia CUDA (cudnn.dll) and it's bin folder must be in the path. Also Nvidia CUPTI library must be in the path eg C:\Program Files\NVIDIA GPU Computing Toolkit\CUDA\v8.0\extras\CUPTI\libx64.
j. ALSO :- Please ensure that you have matplotlib and numpy also installed.. People can use which ever editor they are comfortable with. PyCharm or Atom IDE (development environments) can be used (your choice). Other options are Jupyter, Eclipse with Python, Visual Studio with Python and some others etc. Some people may also prefer command line programming using "vi" editor or DOS edit, but we recommend using one or the other IDE).
    i. **We will be using PyCharm and Jupyter from our end.**
k. For some tutorials like Multinomial Regression (OPTIONAL) which uses stats_models api, download and install Cython from http://www.lfd.uci.edu/~gohlke/pythonlibs/#cython

    i. Cython-0.27-cp36-cp36m-win32.whl
    ii. OR
    iii. Cython-0.27-cp36-cp36m-win_amd64.whl
    iv. In the download folder run
        A) pip install Cython
l. Then
    i. pip install -U statsmodels
m. Install OpenCV 3.4 with this command :-

```
python -m pip install opencv-contrib-python (for details refer here :- https://www.scivision.co/install-opencv-python-windows/)
```

2) Troubleshooting :-
    a. If this kind of error comes while installing TensorFlow or pip upgrade :- "PermissionError: [WinError 5] Access is denied: 'c:\\program files\\python35\\Li b\\site-packages\\backports'" or a similar error related to permissions, then on CMD prompt in windows, right click and "Run as Administrator" and then install TensorFlow.

b. If this kind of error comes :- On Windows, TensorFlow reports either or both of the following errors after executing an import tensorflow statement:

   No module named "_pywrap_tensorflow"
   DLL load failed.

   Then be sure you run the Visual C++ build tools file (google it to find VC++ build tools or Visual Studio Build tools)
   vs_buildtools__1020184259.1499279260
   OR
   VC++ Build tools ie MSVCP140.DLL by checking your system/path - if not get it here https://www.microsoft.com/en-us/download/details.aspx?id=48145

c. If scipy or numpy installation gives trouble :-
   i. You can install scipy and numpy using their wheels.

   ii. First install wheel package if it's already not there...

   iii. pip install wheel
   iv. Just select the package you want from http://www.lfd.uci.edu/~gohlke/pythonlibs/#scipy

   v. Example: if you're running python3.5 32 bit on Windows choose scipy-0.18.1-cp35-cp35m-win_amd64.whl then it will automatically download.

   vi. Then go to the command line and change the directory to the downloads folder and install the above wheel using pip.

   vii. Example:

   viii. cd C:\Users\[user]\Downloads
   ix. pip install scipy-0.18.1-cp35-cp35m-win_amd64.whl OR pip3 install scipy-0.18.1-cp35-cp35m-win_amd64.whl
   x. Same way one can install numpy+mkl from the folder http://www.lfd.uci.edu/~gohlke/pythonlibs/#numpy.
      A) pip3 install "numpy-1.13.1+mkl-cp35-cp35m-win_amd64.whl"

d. For some tutorials, one may have to install pydotplus
   i. pip3 install pydotplus

e. For some graphs in tutorials, one may need to install GraphViz and add it to the path :
   i. http://www.graphviz.org/Download_windows.php

f. For the spam detection in SMS tutorial :
   i. `$ pip install -U textblob`
   ii. `$ python -m textblob.download_corpora`

This will install TextBlob and download the necessary NLTK corpora. If you need to change the default download directory set the NLTK_DATA environment variable.

g. For the program on IMDB Sentiment Analysis, install Python Progress Indicator or pyprind.
   i. pip install pyprind

h. For Scikit-Learn, install version 0.19.1 (0.19.0 is giving error on fetch_california_housing() function).

i. File "C:\Python27\lib\site-packages\scipy__init__.py", line 61, in <module>
   from numpy._distributor_init import NUMPY_MKL # requires numpy+mkl

   ⇨ To solve this just upgrade scipy
   1. pip3 install --upgrade scipy (to upgrade scipy to v1.0.0)

3) Learning

Some knowledge or review of basic Machine Learning techniques is useful (but not required) eg https://en.wikipedia.org/wiki/Machine_learning.

This workshop will cover these techniques in sufficient detail.

4) Programming Tutorials

Review the tutorials in the attached documents and code files titled :-

- Programming tutorials given in the folder titled “Core ML Tutorials”
- Classification
- Decision Trees
- Spam Detection in SMS
- TensorFlow
  - hello world.py
  - Linear Regression (lr.py)
  - MNIST For ML Beginners (mnist_softmax.py)
  - Deep MNIST for Experts (mnist_deep.py)
  - MNSIT using three layer ANN (with tensorboard summaries) (fully_connected_feed.py and mnist.py)
  - TensorBoard - Visualizing Learning (mnsit_with_summaries.py)

you can try to run some of these programs before the workshop to get a hang of it. The word docs sent will explain what the programs do.

To run a python program :

1) Open a command prompt (eg Windows Command prompt) and navigate to the area where the python code file is. Type :

   >>which python // to check whether it is picking the correct version of Python install

   >>python "hello world.py"

   >>python lr.py

2) Alternatively you can run the Python program in the IDE that you installed eg PyCharm or Atom. Open the code file in an IDE like PyCharm and run it there. You will have to attach Python interpreter which you installed earlier to the PyCharm program by using the option "Attach Python interpreter"

Working with Jupyter and github cloning :

To clone the code of Geron book "Hands on Machine Learning with Scikit-Learn and TensorFlow" from source code repository github :-

```
$ cd $HOME  # or any other development directory you prefer
$ git clone https://github.com/ageron/handson-ml.git
$ cd handson-ml
```

To clone the code for "Python Machine Learning" book by Sebastian Raschka :-

git clone https://github.com/rasbt/python-machine-learning-book-2nd-edition.git

Starting Jupyter :

If you want to use the Jupyter extensions (optional, they are mainly useful to have nice tables of contents), you first need to install them:

```
$ jupyter contrib nbextension install --user
```

Then you can activate an extension, such as the Table of Contents (2) extension:

```
$ jupyter nbextension enable toc2/main
```

Okay! You can now start Jupyter, simply type:

```
$ jupyter notebook
```

This should open up your browser, and you should see Jupyter's tree view, with the contents of the current directory. If your browser does not open automatically, visit localhost:8888. Click on `index.ipynb` to get started!
Note: you can also visit http://localhost:8888/nbextensions to activate and configure Jupyter extensions.