Hello all. Please follow this step by step guide for the character segmentation part for the ML project. Please refer to the Troubleshooting section if you face any difficulties.

Step 1 : Downloading the Scripts

- Make a new folder called "ml_char_gen" on your computer
- Go to the github link https://github.com/bitsgf464-ml/fall2017/tree/master/Fall%202017/Data%20Generation
- Save the following scripts in the "ml_char_gen" folder :
  - *data_collection.py*
  - *qt_char_select.py*

Step 2 : Getting your Data

- Go to the doc link https://docs.google.com/a/bits-pilani.ac.in/spreadsheets/d/1LTbQOH4YYD0Pv9HLzW9-OIOj6DL2U0fg8oEwWrlJ_QY/edit?usp=sharing
- Check the line numbers given against your name
- Go to the github link https://github.com/bitsgf464-ml/fall2017/tree/master/Fall%202017/Data%20Generation/lines_of_data
- Download your lines from there and keep them in the same folder as the scripts from Step 1 : "ml_char_gen" folder

Step 3 : Checking Python Dependencies

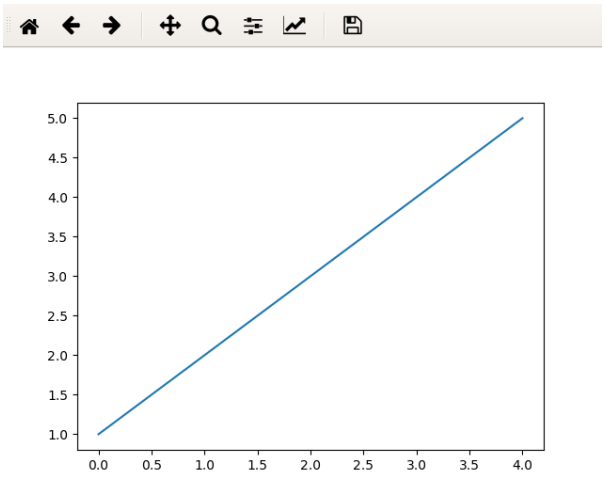
You can install **Anaconda for Python 3** and the dependencies (i.e, the libraries the scripts depend upon to run smoothly) should all be satisfied.

Else,

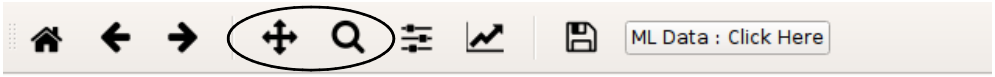
Please install python 3, PyQt5 and matplotlib.
PyQt5 : pip3 install pyqt5
Source : http://pyqt.sourceforge.net/Docs/PyQt5/installation.html

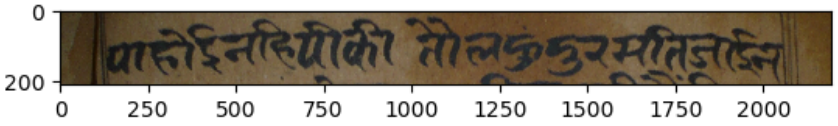
Step 4 : Running the Scripts

- Open your terminal (command prompt) and change directory to your "ml_char_gen" folder directory
- Run the command *python3 data_collection.py page1|0.png*
- The above should be modified to reflect your page and line number and python 3 command. Please refer troubleshooting for further details about python 3 command.
- You will see an initial line plot in matplotlib. This is to configure your matplotlib according to my requirements and show you that the matplotlib library is working well. **Please close the line plot to continue further.**

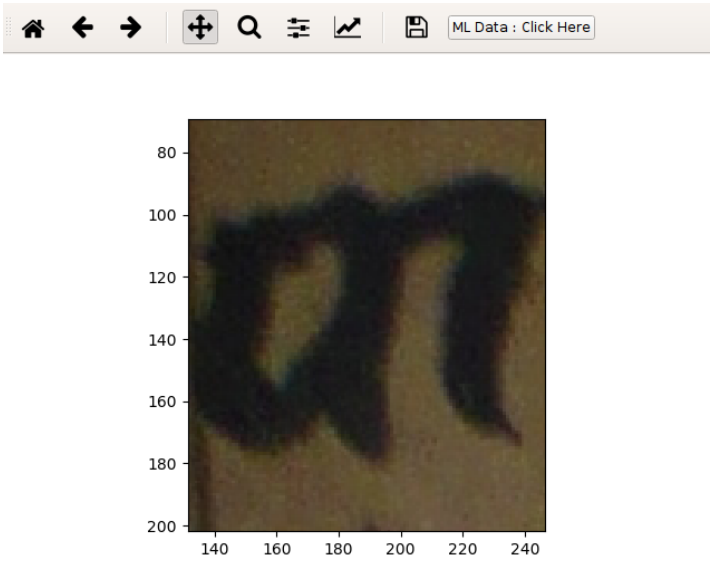


- After you close the line plot, a moment later, the picture of your input line should pop up
- Use the zoom tool and then pan tool to fix your location on the first character

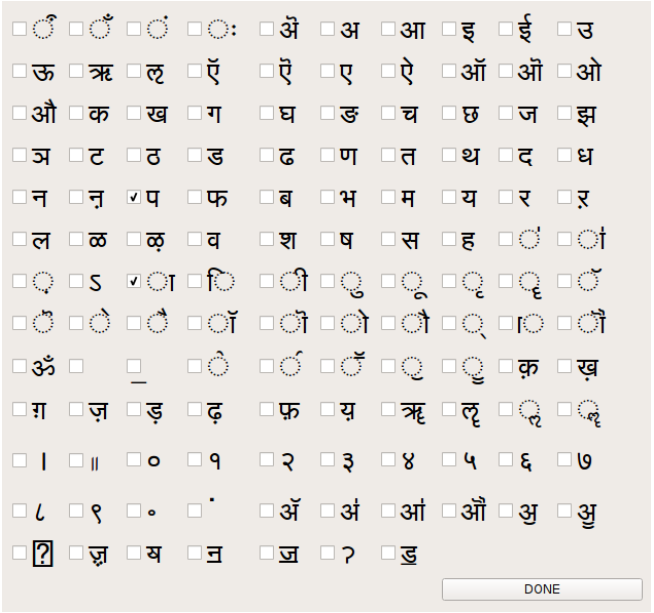




- Now use the ML Data : Click Here button



- In the popup that follows, use the check boxes to select all the different visible characters and then press the Done button



- Continue like this for all the characters in the line
- It is okay for the characters to have some extra blank space on the side but it is not okay to make a box such that there are multiple characters in the box that are not subsequently labeled
- **It is this step that will define the maximum accuracy you can get irrespective of the ML methods you apply**, so please do a good job with this step if you want a high accuracy.

Step 5 : Submitting your Results

- Follow the procedure and label every character from every line assigned to you; then take all the pictures generated and put them all in a folder called "character_segmentation"
- Zip the "character_segmentation" folder (which contains pictures of only the final characters as are generated by the script and nothing else)
- Send this folder across to f20140471@goa.bits-pilani.ac.in as an attachment
- **The subject of the email should be "ML Data Generation" followed by your ID number**

Troubleshooting

1. You may face other dependency issues. They can all be solved by installing the required libraries. The libraries used in the scripts are :
   - Numpy
   - Matplotlib
   - PyQt5
   - Sys
   - Os
   - Time

   Please check if all of these are installed before running the scripts.

2. In step 4 you may see the error "python3 command not found". In such a case, please open the data_collection.py file and modify line 50 to replace "python3" with whatever command your system uses to evoke python 3. Common variations are : python, py3, py, python3.5, python35, etc.