Lecture 6.5

Optional Interlude: Cloud Computing

Option 1 -- Google Colaboratory
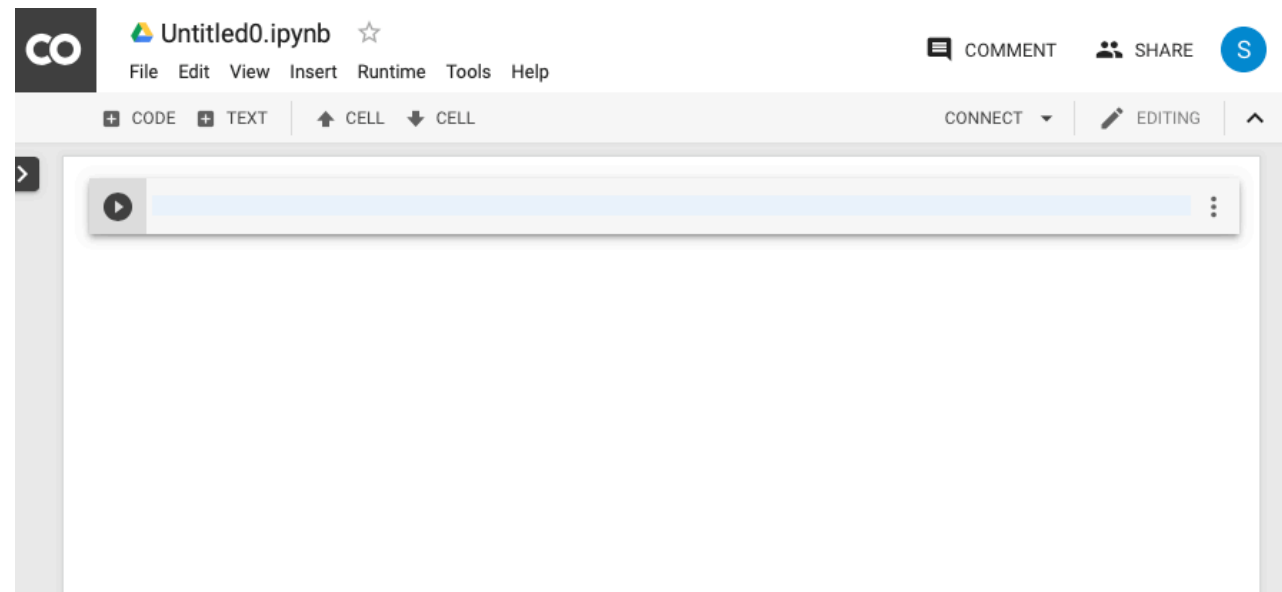
STAT 453: Deep Learning, Spring 2020

Sebastian Raschka

http://stat.wisc.edu/~sraschka/teaching/stat453-ss2020/

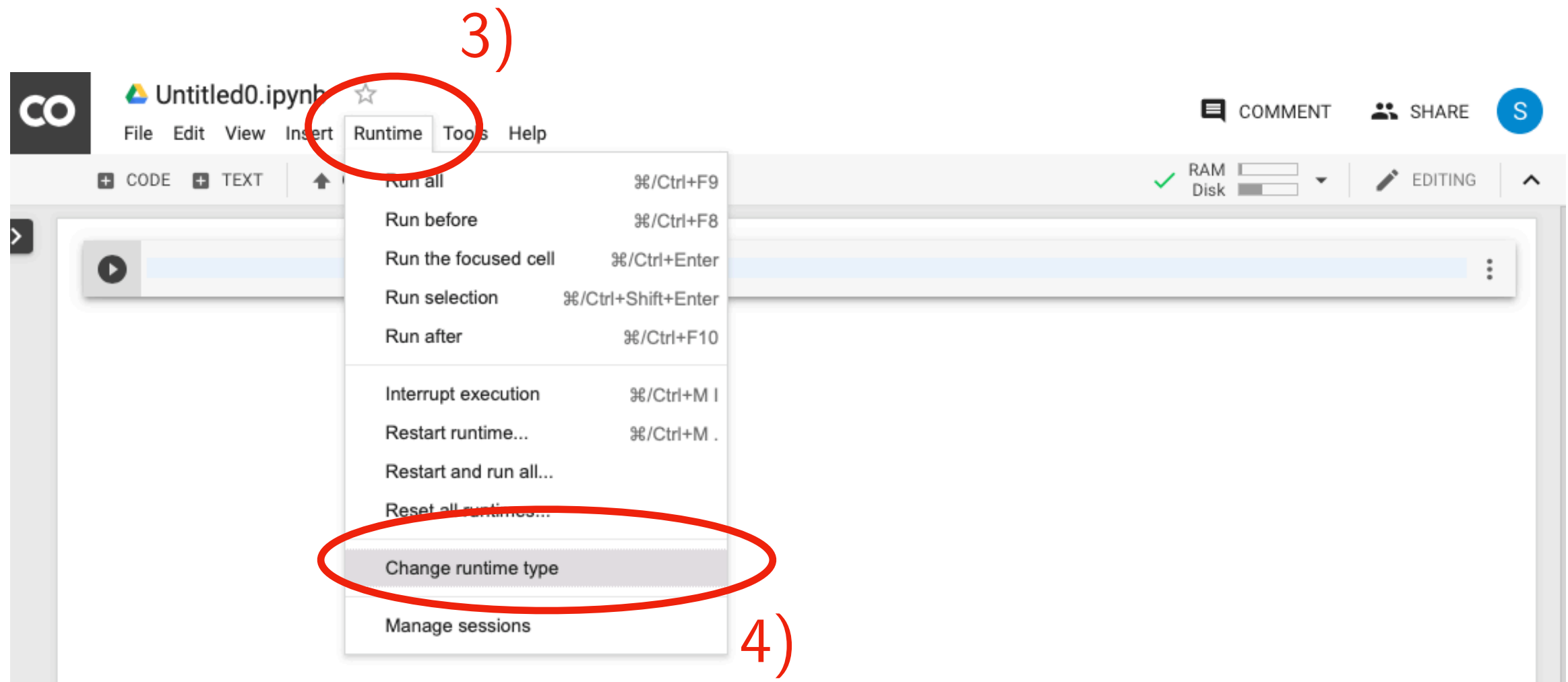# Option 1: Google Colab

https://colab.research.google.com

- Free Google-flavored Jupyter Notebooks in the Cloud

- For each notebook, they spin up a custom (Linux-based) computing instance

- Computations limited to ~12 h though; you won't lose your notebook, but computations will be interrupted

- Maybe useful for quick testing/experimenting/sharing (but maybe tedious as you need to reinstall packages each time)
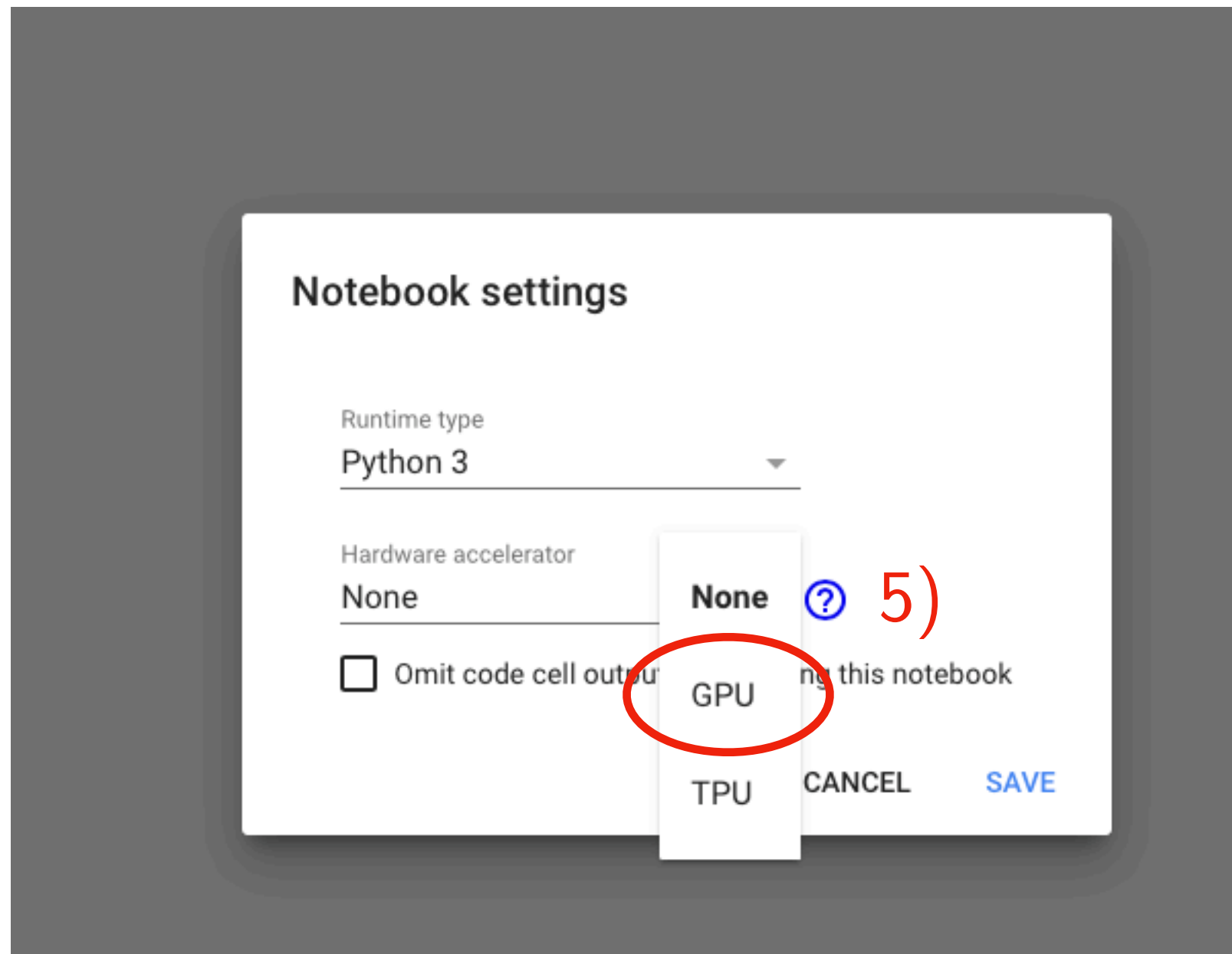
Option 1: Google Colab
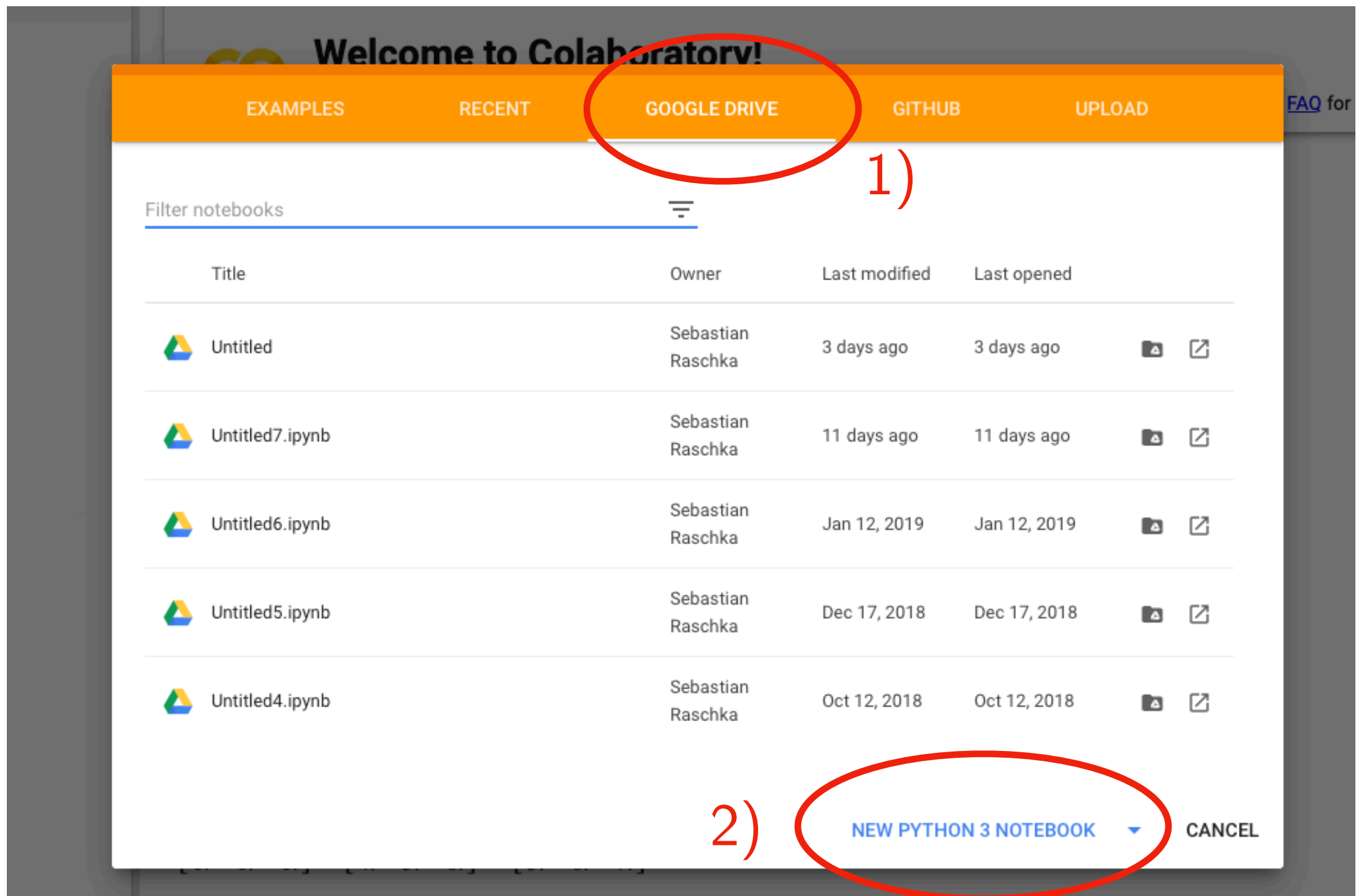
Menu appears if you visit https://colab.research.google.com
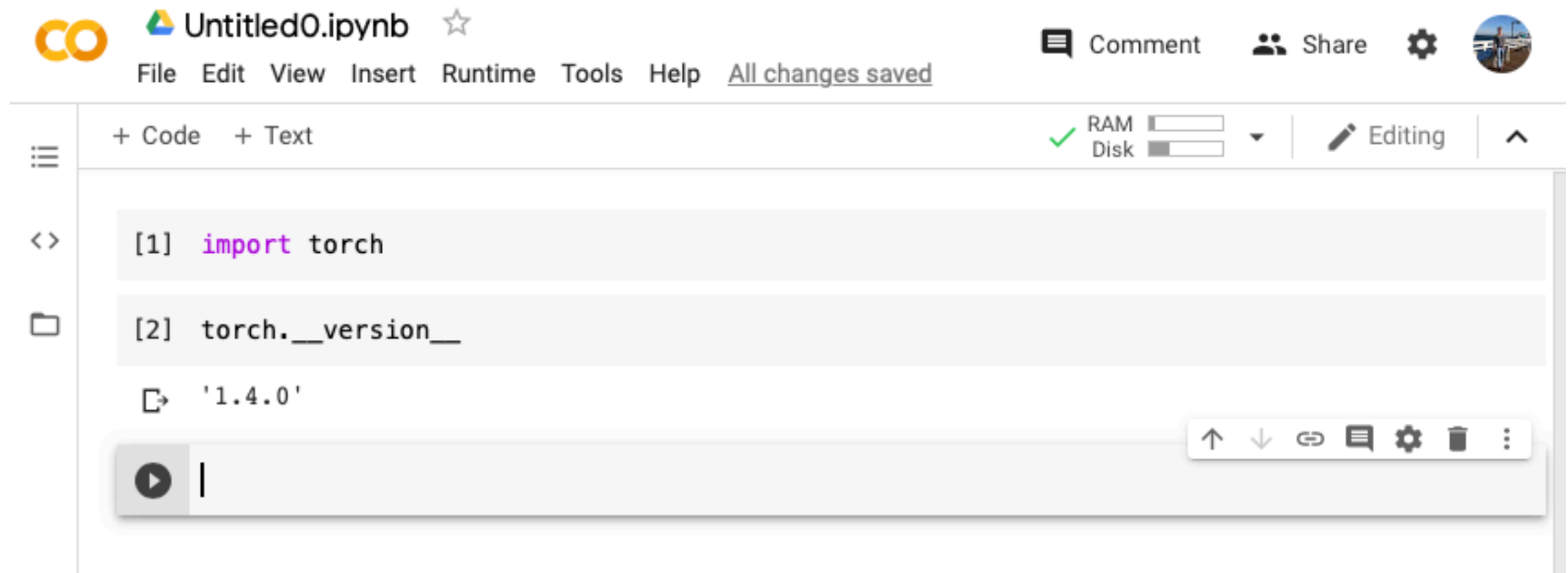
Option 1: Google Colab

Follow these steps for running code on GPU later (default is CPU)

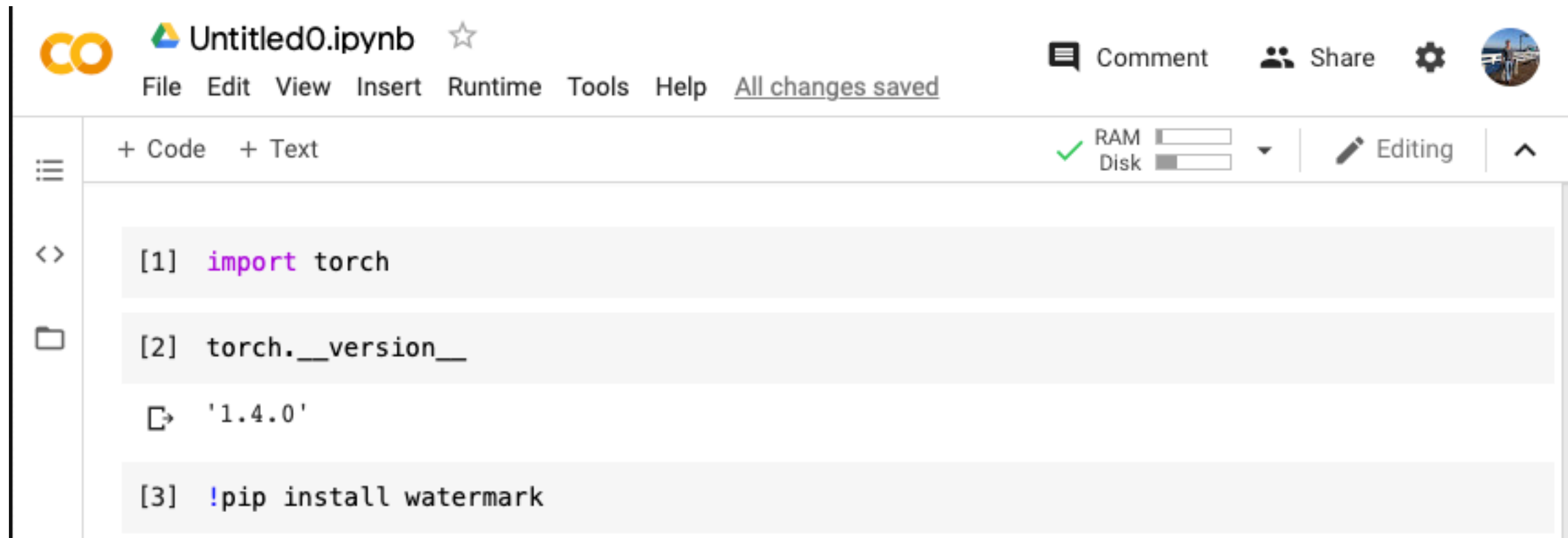# Option 1: Google Colab

Follow these steps for running code on GPU later (default is CPU)

Option 1: Google Colab

• This is nice! It appears that PyTorch is already pre-installed now (it wasn't always the case)

Option 1: Google Colab

- In any case, if you'd like/need to install packages, you can do it as shown in the example above

- Note that in Jupyter Notebooks, the "!" indicates that what follows on that line is a "shell command" (you can think of a "shell" as the Linux & macOS command-line terminal, e.g,. a Bash Shell)

Option 1: Google Colab
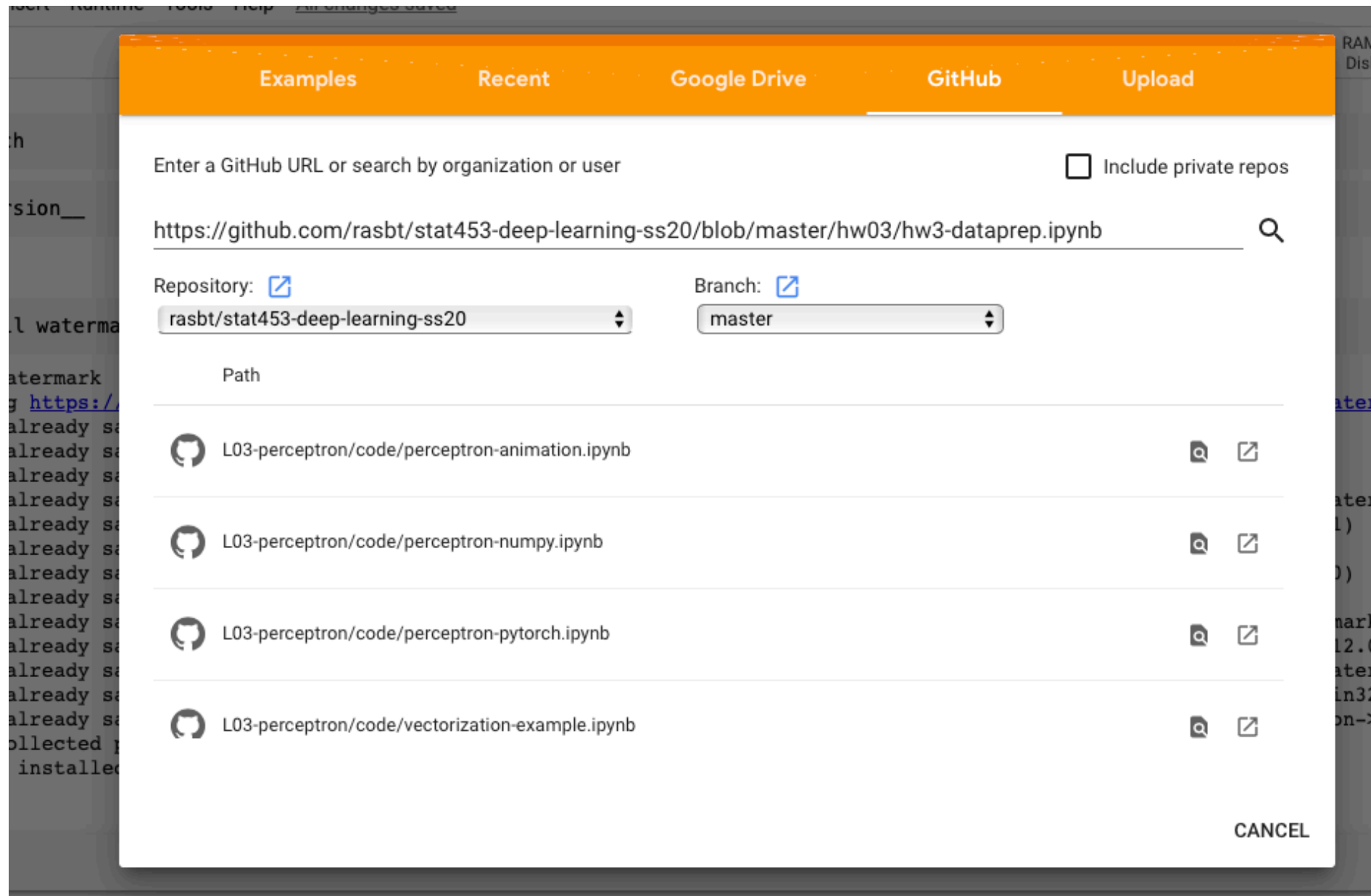
- You can also upload Notebooks or directly paste GitHub links to notebooks

Option 1: Google Colab

When you import a Notebook from a GitHub link, make sure to save it
in your Google Drive if you plan to make edits, otherwise it will be gone later

Option 1: Google Colab

```
[1]        return images, labels


    X_train, y_train = load_mnist('', kind='train')
    print('Rows: %d, columns: %d' % (X_train.shape[0], X_train.shape[1]))

    X_test, y_test = load_mnist('', kind='t10k')
    print('Rows: %d, columns: %d' % (X_test.shape[0], X_test.shape[1]))
```

```
-------------------------------------------------------------------
FileNotFoundError                         Traceback (most recent call last)
<ipython-input-1-da8e9179fc02> in <module>()
     25
     26
---> 27 X_train, y_train = load_mnist('', kind='train')
     28 print('Rows: %d, columns: %d' % (X_train.shape[0], X_train.shape[1]))
     29


<ipython-input-1-da8e9179fc02> in load_mnist(path, kind)
     10                              '%s-images-idx3-ubyte' % kind)
     11
---> 12     with open(labels_path, 'rb') as lbpath:
     13         magic, n = struct.unpack('>II',
     14                                  lbpath.read(8))

FileNotFoundError: [Errno 2] No such file or directory: 'train-labels-idx1-ubyte'
```

If you'd run the HW3 notebook, you'd likely encounter this error.
This is because it can't find the dataset via the specified, relative link ...

Option 1: Google Colab

... you'd need to get the datafile into the same location as the notebook (or provide an absolute file path).
First, locate the position of the saved Notebook in your Google Drive:
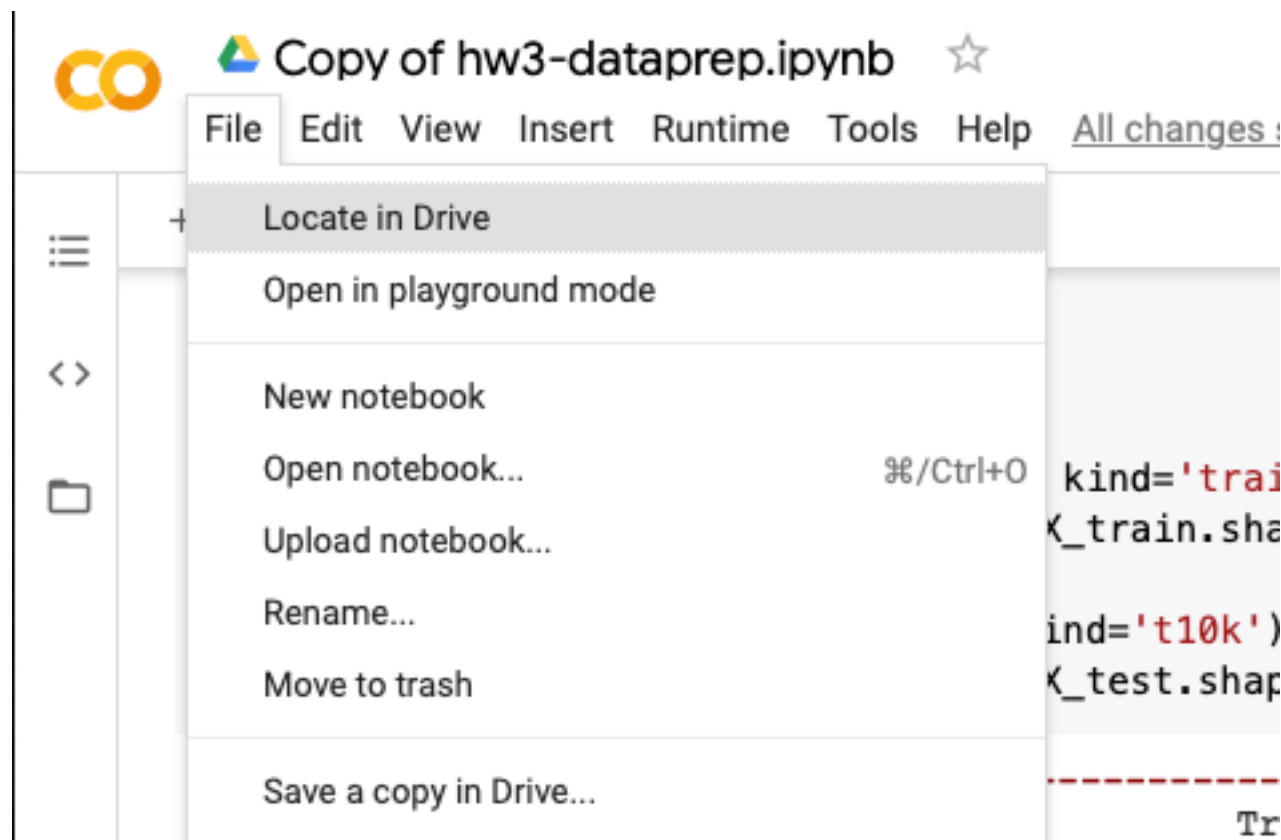
Option 1: Google Colab

Notebooks are usually in a directory called "My Drive/Colab Notebooks"
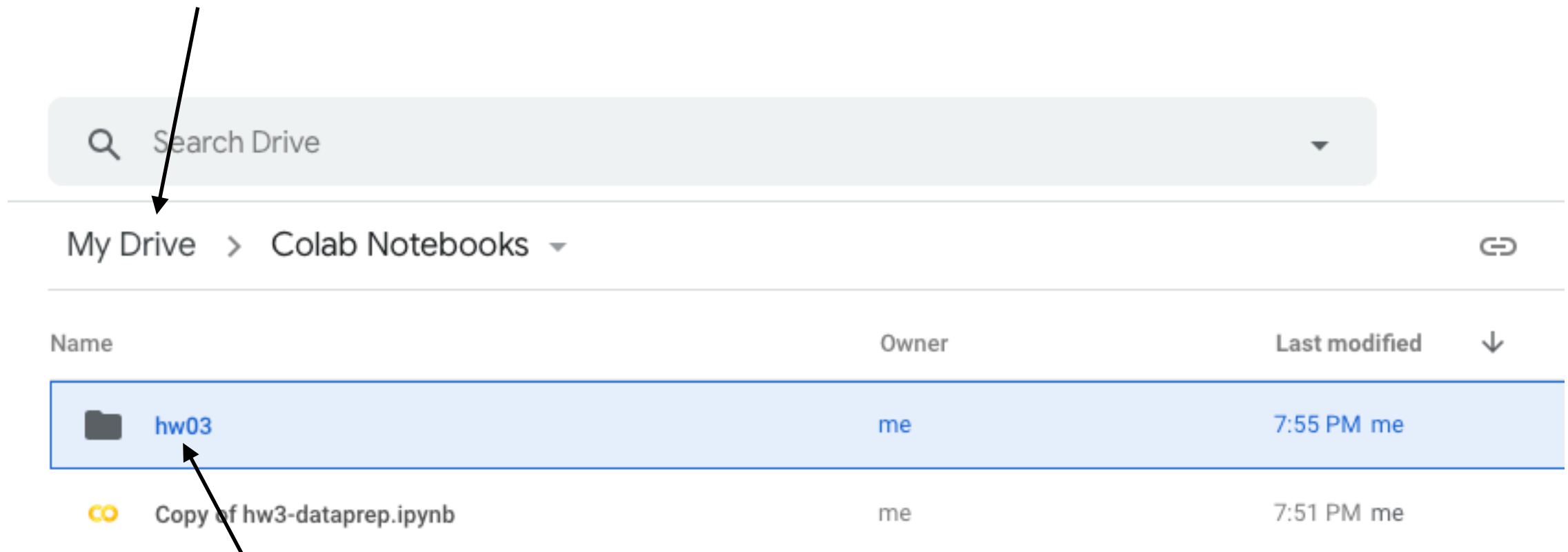
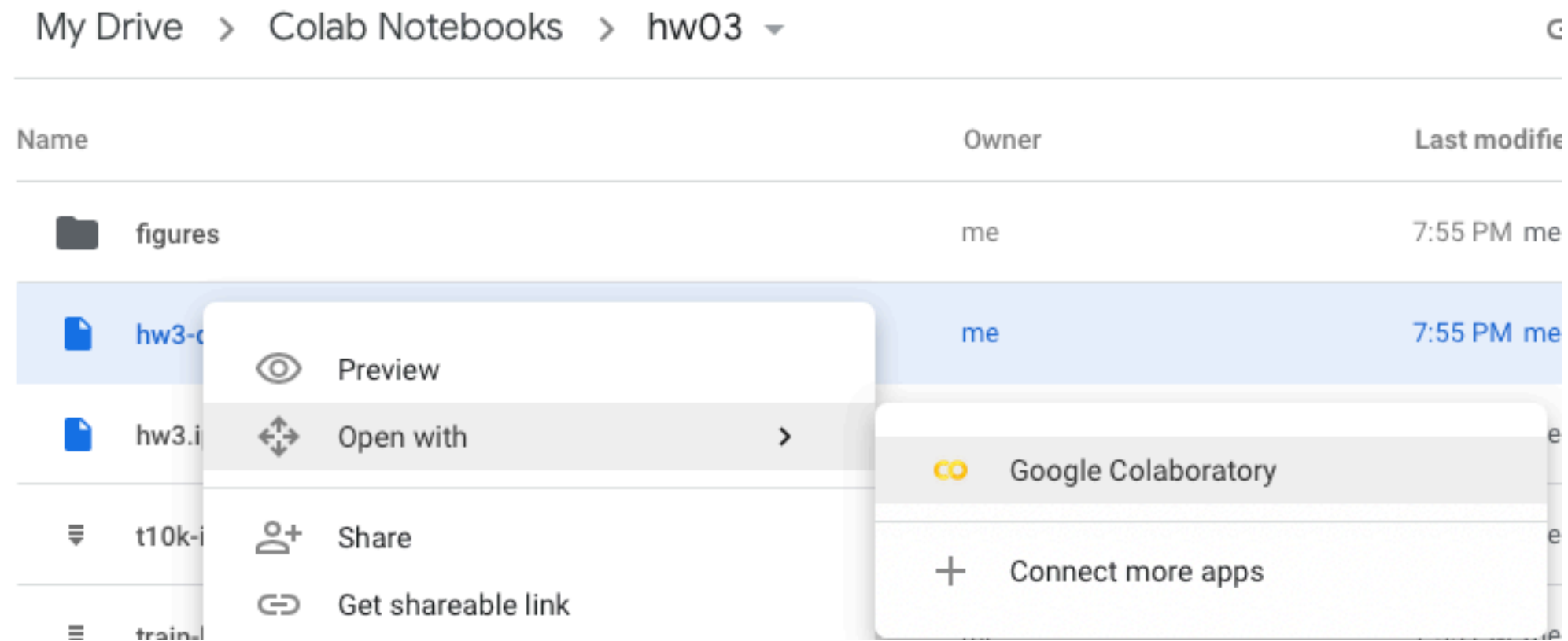I recommend uploading the whole folder though (simply drag&drop it from your computer into this window)

Option 1: Google Colab

Then, simply open the notebook in Colaboratory.

Option 1: Google Colab

Unfortunately, there's some extra step required: mounting your Google Drive to the computer that now runs the Notebook. You need to execute the following code:

1)

```
from google.colab import drive
drive.mount('/content/drive')
```

... Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client_id=9473189

Enter your authorization code:
[   ]

Then, click on the link and enter it in the field above

2)

```
from google.colab import drive
drive.mount('/content/drive')
```

... Go to this URL in a browser: https://accounts.google.co

Enter your authorization code:
....

3)

```
from google.colab import drive
drive.mount('/content/drive')
```

Go to this URL in a browser: https://ac

Enter your authorization code:
....
Mounted at /content/drive

Your Google Drive should now be finally mounted:

Option 1: Google Colab

Now, you simply need to provide the correct address to the dataset inside the Notebook and it should work:

```python
[16] # this code cell unzips the .gz files

     import sys
     import gzip
     import shutil
     import os

     writemode = 'wb'
     zipped_mnist = [f for f in os.listdir('/content/drive/My Drive/Colab Notebooks/hw03') if f.endswith('ubyte.gz')]
     for z in zipped_mnist:
         path = os.path.join('/content/drive/My Drive/Colab Notebooks/hw03', z)
         with gzip.GzipFile(path, mode='rb') as decompressed, open(path[:-3], writemode) as outfile:
             outfile.write(decompressed.read())
```

```python
     with open(labels_path, 'rb') as lbpath:
         magic, n = struct.unpack('>II',
                                  lbpath.read(8))
         labels = np.fromfile(lbpath,
                              dtype=np.uint8)

     with open(images_path, 'rb') as imgpath:
         magic, num, rows, cols = struct.unpack(">IIII",
                                                imgpath.read(16))
         images = np.fromfile(imgpath,
                              dtype=np.uint8).reshape(len(labels), 784)

     return images, labels


 X_train, y_train = load_mnist('/content/drive/My Drive/Colab Notebooks/hw03', kind='train')
 print('Rows: %d, columns: %d' % (X_train.shape[0], X_train.shape[1]))


 X_test, y_test = load_mnist('/content/drive/My Drive/Colab Notebooks/hw03', kind='t10k')
 print('Rows: %d, columns: %d' % (X_test.shape[0], X_test.shape[1]))
```

```
 Rows: 60000, columns: 784
 Rows: 10000, columns: 784
```