

# Train ANNs in Python

Siyu Qi

June 9, 2020

## 1 Introduction

There are two platforms to train the ANNs in Python: Google Colab and a local computer. In this tutorial, we will learn how to run scripts on Google Colab. Google Colab is an online environment, based on Python Jupyter notebooks, which provides free Tesla K80 GPUs for use. You can run your Python code in Colab environment, and all you need is a Google account.

Note that Google Colab automatically recycles a runtime (or session) if it lacks activity for 90 minutes or it has existed for 12 hours, which means you have to interact with the web page every 90 minutes, and finish all your work within 12 hours. Otherwise, your job will be killed and files will be deleted.<sup>1</sup>

## 2 Contents

1. Preparation
2. Create a notebook on Google Colab by uploading a script from local
3. Setup the environment and upload dataset
4. Train ANNs
5. Download trained model

## 3 Preparation

Files in folder "python\_ANN":

1. `ann_helper.py`: a helper script which users do not need to modify or run.

---

<sup>1</sup>One easiest way to keep your Colab job active is to open a separate browser window (instead of a tab in your current window) for it, then you don't need to go back to it every 90 minutes.

2. `train_single_output_ANN.ipynb` and `train_multi_output_ANN.ipynb` : Training scripts. Users can train single-output or multi-output ANNs by running them on Google Colab platform.
3. `test_single_output_ANN.py` and `test_multi_output_ANN.py` : Testing scripts. Users can test trained ANN models by running them on local computer.

Also, we need a dataset file `ANN_data.xlsx`. Before running our scripts on Google Colab platform, we need to upload the helper script and dataset to Google Drive. Go to <https://drive.google.com/>, click on `New`, then `Folder Upload` and select `python_ANN`. Then (important!) enter the uploaded `python_ANN` folder, select `File upload` and upload `ANN_data.xlsx`.

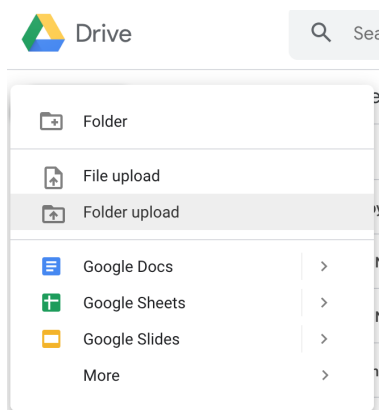
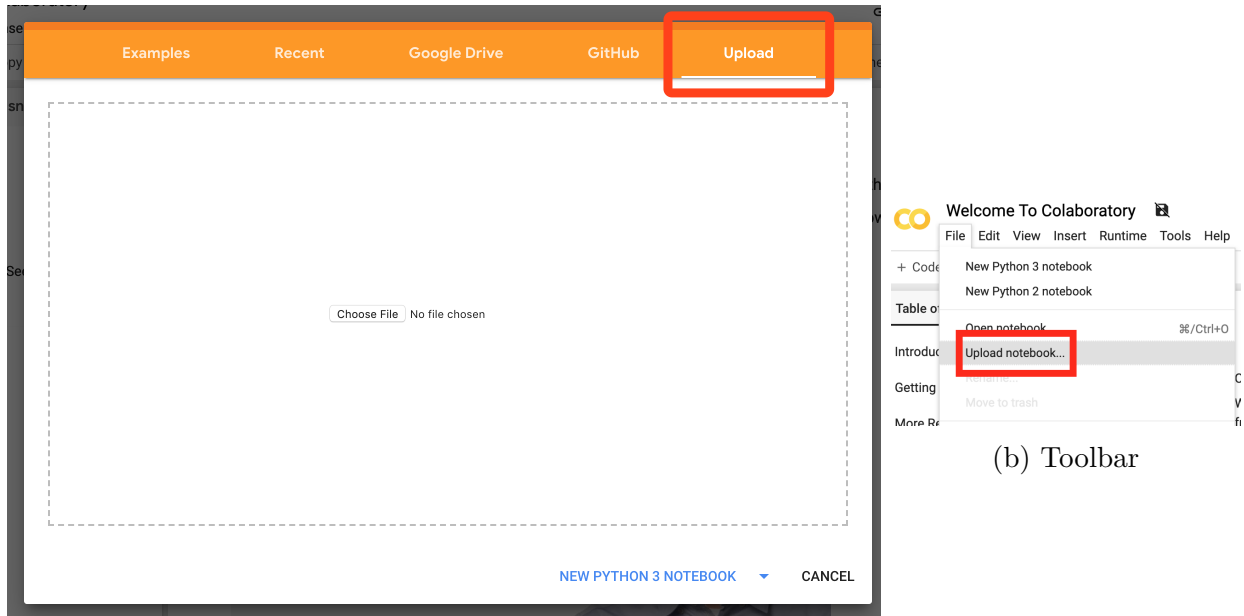


Figure 1: Upload the complete folder

## 4 Open Notebook

Go to Google Colab (<http://colab.research.google.com>) and click `Upload` as in Fig.2a. If this message box does not pop up, select `File` → `Upload files` as in Fig.2b. Then click `Choose File` and select the `.ipynb` file in `python_ANN` folder.

If you already have the notebook created, you can directly open it from `Recent` tab and skip to next section.



(a) Welcome page

(b) Toolbar

Figure 2: Upload notebook

If uploading is successful, you will be redirected to the newly created notebook on Google Colab just like Fig.3.

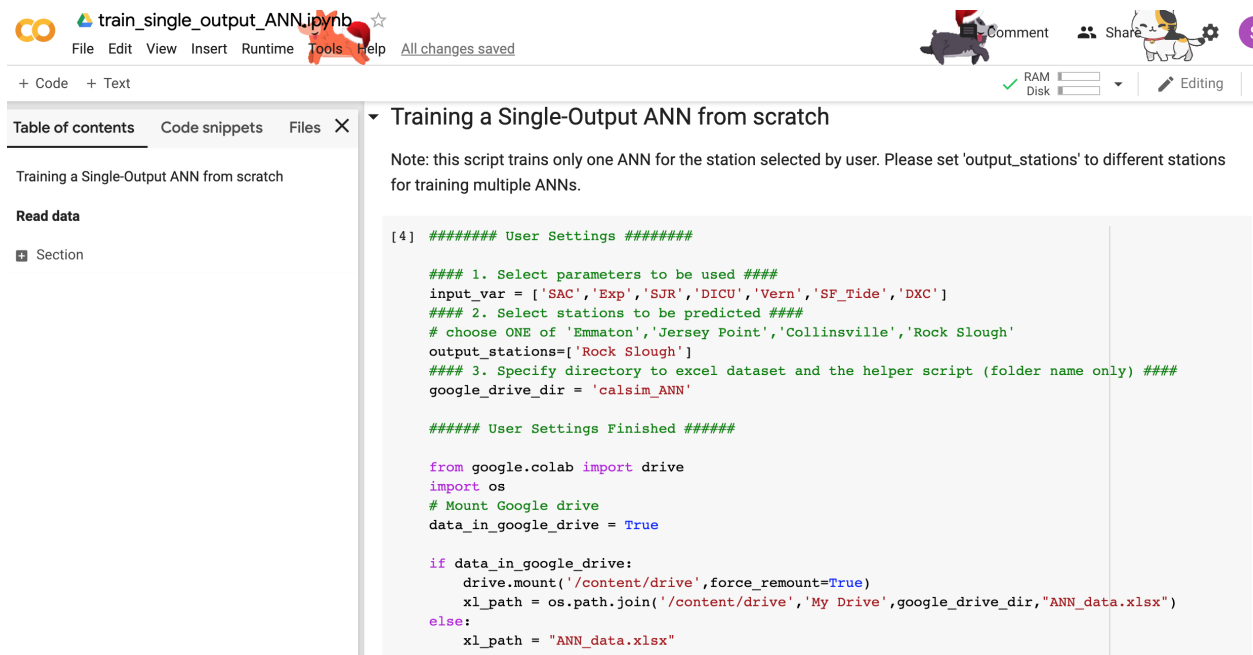
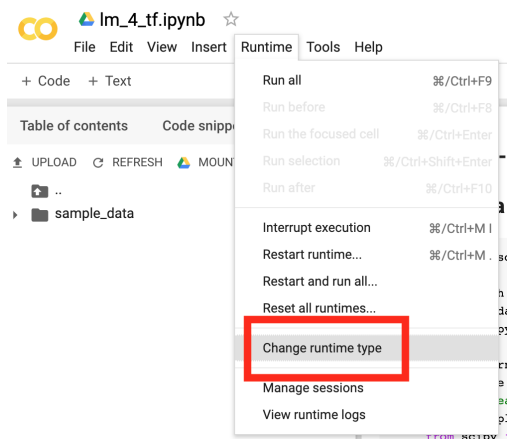


Figure 3: New notebook uploaded

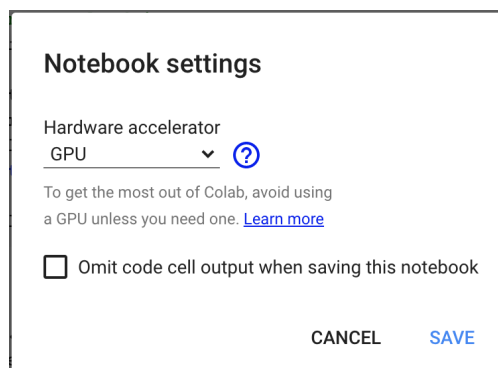
## 5 Setup Environment and Prepare Dataset

### 5.1 Environment Setup

Choose `Runtime` → `Change runtime type` in toolbar, set `Hardware accelerator` to `GPU` like Fig.4. The settings will be saved, so you can skip this step next time running this same notebook.



(a) Set runtime type



(b) Settings

Figure 4: Environment setup

### 5.2 Prepare Dataset

Now the environment is ready. Next we will mount our Google Drive to access the dataset.


Go to the first code cell in the notebook and set `google_drive_dir` to the Google Drive folder where `ann_helper.py` and `ANN_data.xlsx` are in, here it should be:

```
google_drive_dir='python_ANN'
```

Note: `google_drive_dir` should be set to a path, not a URL. To find the path, first locate `ann_helper.py` and `ANN_data.xlsx` in Google Drive, then the path (excluding "My Drive") is at top of the page. Fig. 5 is an example.

| My Drive > some_folder > python_ANN      |       |               |           |       |
|--|-------|---------------|-----------|-------|
| Name                                     | Owner | Last modified | File size |       |
| __pycache__                              | me    | Dec 2, 2019   | me        | —     |
| train_single_output_ANN_from_excel.ipynb | me    | 12:57 PM      | me        | 47 KB |
| train_multi_output_ANN.ipynb             | me    | Jun 5, 2020   | me        | 91 KB |
| train_multi_output_ANN_from_csv.ipynb    | me    | May 29, 2020  | me        | 28 KB |
| test_single_output_ANN.py                | me    | Dec 10, 2019  | me        | 5 KB  |
| test_multi_output_ANN.py                 | me    | Dec 10, 2019  | me        | 5 KB  |
| ann_helper.py                            | me    | May 29, 2020  | me        | 11 KB |
| ANN_data.xlsx                            | me    | Dec 7, 2019   | me        | 25 MB |

Figure 5: Example: google\_drive\_dir='some\_folder/python\_ANN'

Click the  button of first code cell, then click on the URL shown in the interactive output block to verify your account, copy authorization code, paste it in the text entry box pointed by the red arrow in Fig.6 and press enter.

```

##### User Settings #####

### 1. Select parameters to be used ###
input_var = ['SAC','Exp','SJR','DICU','Vern','SF_Tide','DXC']
### 2. Select stations to be predicted ###
# choose ONE of 'Emmaton','Jersey Point','Collinsville','Rock Slough'
output_stations=['Rock Slough']
### 3. Specify directory to excel dataset and the helper script (folder name only) ###
### if data is uploaded instead of in your google drive, set data_in_google_drive=False ####
data_in_google_drive = True
google_drive_dir = 'calsim_ANN'

##### User Settings Finished #####

from google.colab import drive
import os
# Mount Google drive
if data_in_google_drive:
    drive.mount('/content/drive',force_remount=True)
    xl_path = os.path.join('/content/drive','My drive',google_drive_dir,"ANN_data.xlsx")
else:
    xl_path = "ANN_data.xlsx"

... Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client\_id=947318989803-6bn
Enter your authorization code:

```

Figure 6: Variable settings and verification

## 6 Train your ANN

Before running the script, we can modify the variables in "User Settings" section in the first code block as needed.

1. `input_var`: a list of (one or more) input variables to be used for training.
2. `output_stations`: In single-output ANN script, set it to the single output station. In multi-output ANN script, set it to a list of multiple stations.
3. `google_drive_dir`: the folder where `ann_helper.py` and `ANN_data.xlsx` locate.

Click **Runtime** → **Run all**, and now your program is running!

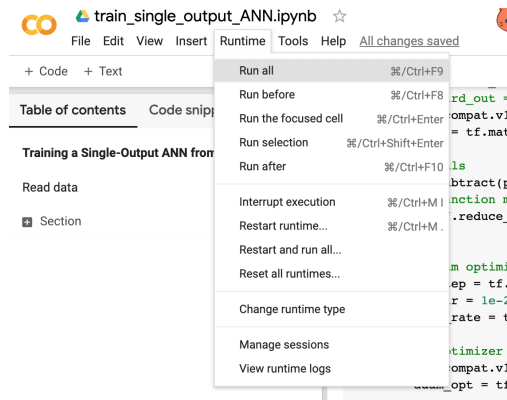


Figure 7: Run your training job

## 7 Find Trained Model

Once training is finished, in the folder where **ANN\_data.xlsx** locates, you can find a folder named **models**, the trained model and fortran file are automatically saved there.

My Drive > python\_ANN ▾

| Name ▾                                | Owner | Last modified   | File size |
|---------------------------------------|-------|-----------------|-----------|
| models                                | me    | 8:54 PM me      | —         |
| __pycache__                           | me    | Dec 2, 2019 me  | —         |
| train_single_output_ANN.ipynb         | me    | 8:57 PM me      | 41 KB     |
| train_multi_output_ANN.ipynb          | me    | 9:03 PM me      | 87 KB     |
| train_multi_output_ANN_from_csv.ipynb | me    | May 29, 2020 me | 28 KB     |
| test_single_output_ANN.py             | me    | Dec 10, 2019 me | 5 KB      |
| test_multi_output_ANN.py              | me    | Dec 10, 2019 me | 5 KB      |
| ann_helper.py                         | me    | 8:51 PM me      | 11 KB     |
| ANN_data.xlsx                         | me    | Dec 7, 2019 me  | 25 MB     |

Figure 8: Saved models in Google Drive