Train ANNs in Python

Siyu Qi

April 1, 2021

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.[1]

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.

[1] 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 a local computer.

Also, we need a dataset file ANN_data.xlsx. Before running our scripts on Google Colab platform, please upload the helper script and dataset to Google Drive:

1. Go to `https://drive.google.com/`, click on New, then Folder Upload and select python_ANN.

2. (Important!) Enter the uploaded python_ANN folder, select File upload and upload ANN_data.xlsx.

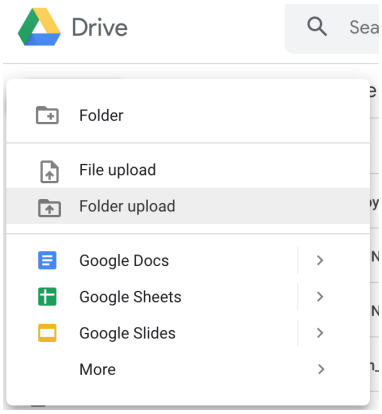
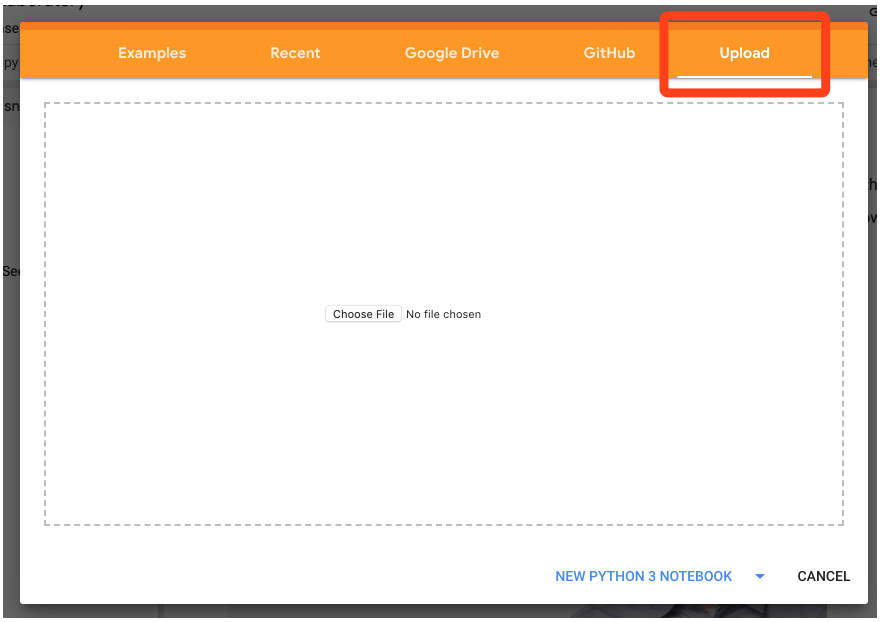


Figure 1: Upload the complete folder

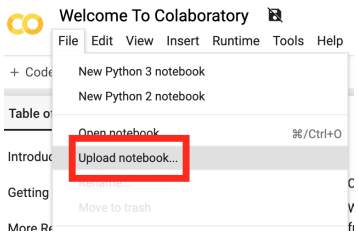
4 Open Notebook

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

If you have uploaded the notebook before, you can directly open it from Recent tab and skip to next section.



(a) Welcome page



(b) Toolbar

Figure 2: Upload notebook

Once uploading succeed, you will be redirected to the newly created notebook on Google Colab just like Fig.3.

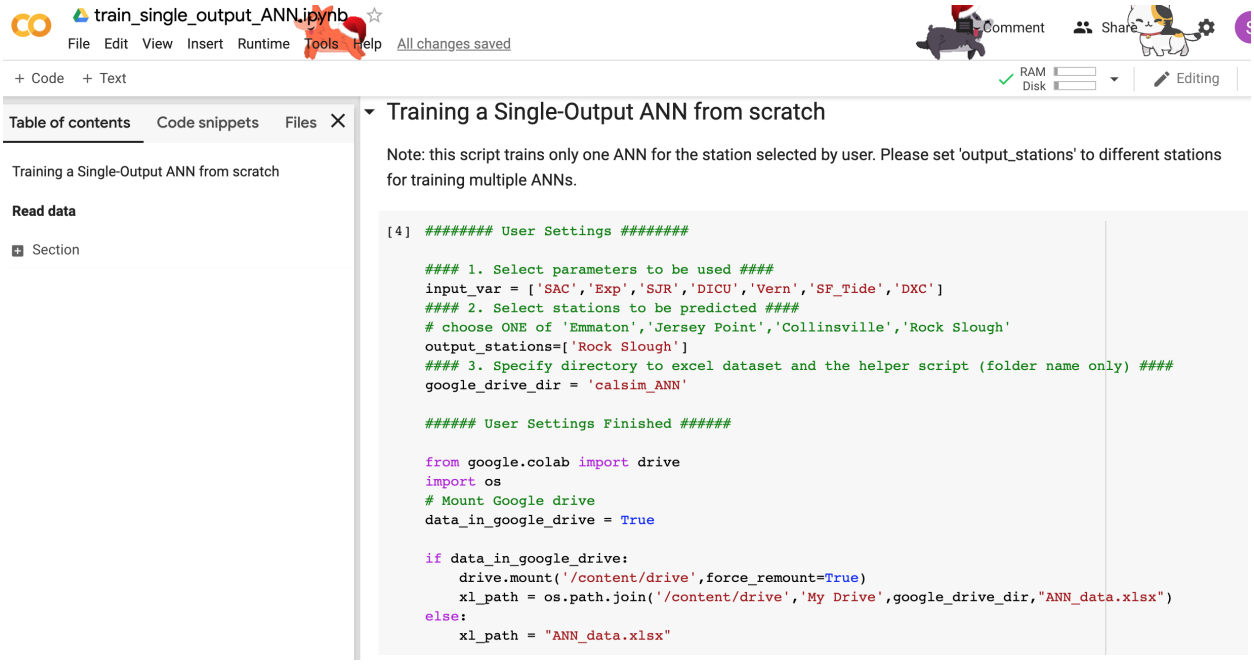
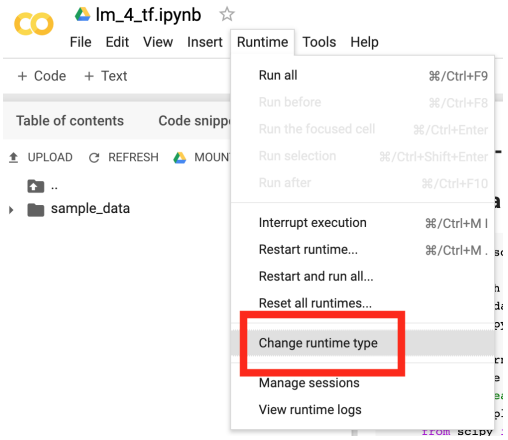


Figure 3: New notebook uploaded

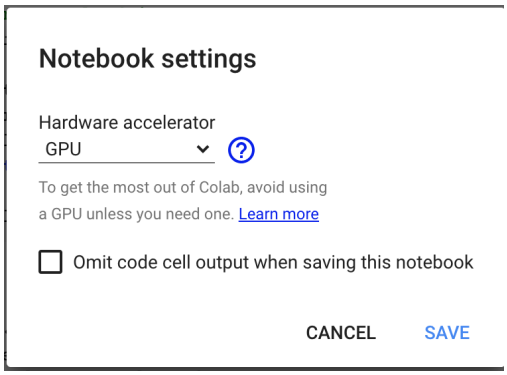
5 Setup Environment and Prepare Dataset

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, let's 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.xlxs` 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.xlxs` in Google Drive, then the path (excluding "My Drive") is at top of the page. Fig. 5 is an example.

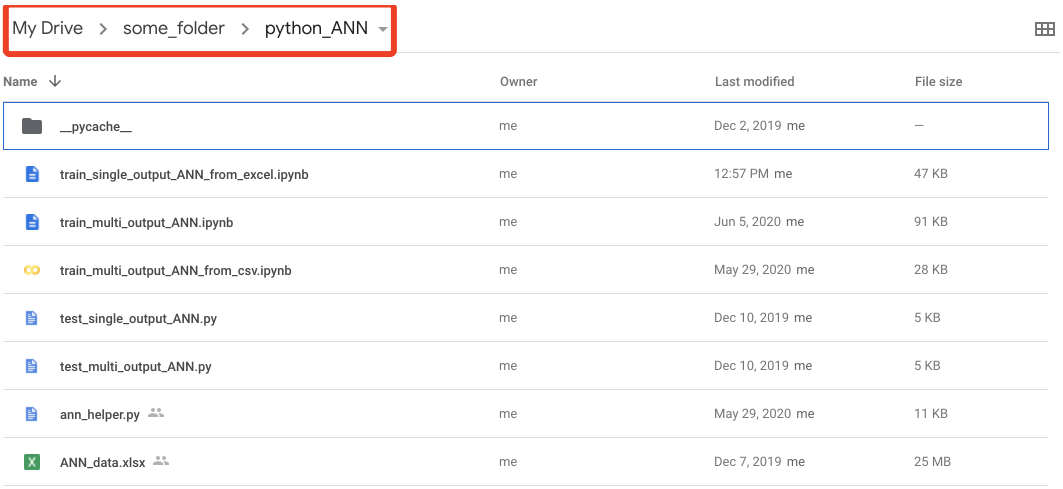


Figure 5: Example: google_drive_dir='some_folder/python_ANN'

6 Train your ANN

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

1. `input_var`: a list of (one or more) input variables for training ANNs.
2. `output_stations`: in single-output ANN script, set it to a 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.xlxs` locate.
4. `is_quick_test`: set it to 'yes' to quickly run through the code and check for bugs, or set it to 'no' to get the ANNs fully trained.
5. `running_on_colab`: a boolean variable, set it to True to run the script on Colab, or set it to False to run the code on a local computer.

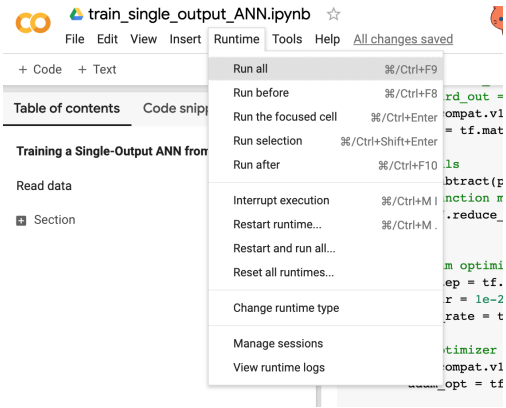


Figure 6: Run your training job

Click Runtime → Run all as in Fig. 6, and the second code cell will ask for an authorization code. Click on the URL shown in the interactive output block to verify your account, copy the authorization code, paste it in the text entry box pointed by the red arrow in Fig.7 and press enter.

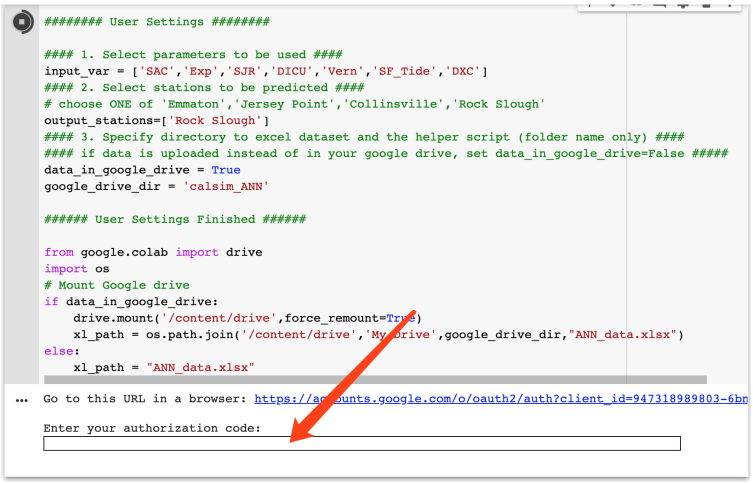


Figure 7: Variable settings and verification

7 Find Results

Once training is finished, in the Google Drive folder where ANN_data.xlsx locates, you can find a folder named models. The trained model, two text files storing ground truth and ANN estimations as well as a fortran file are automatically saved there.

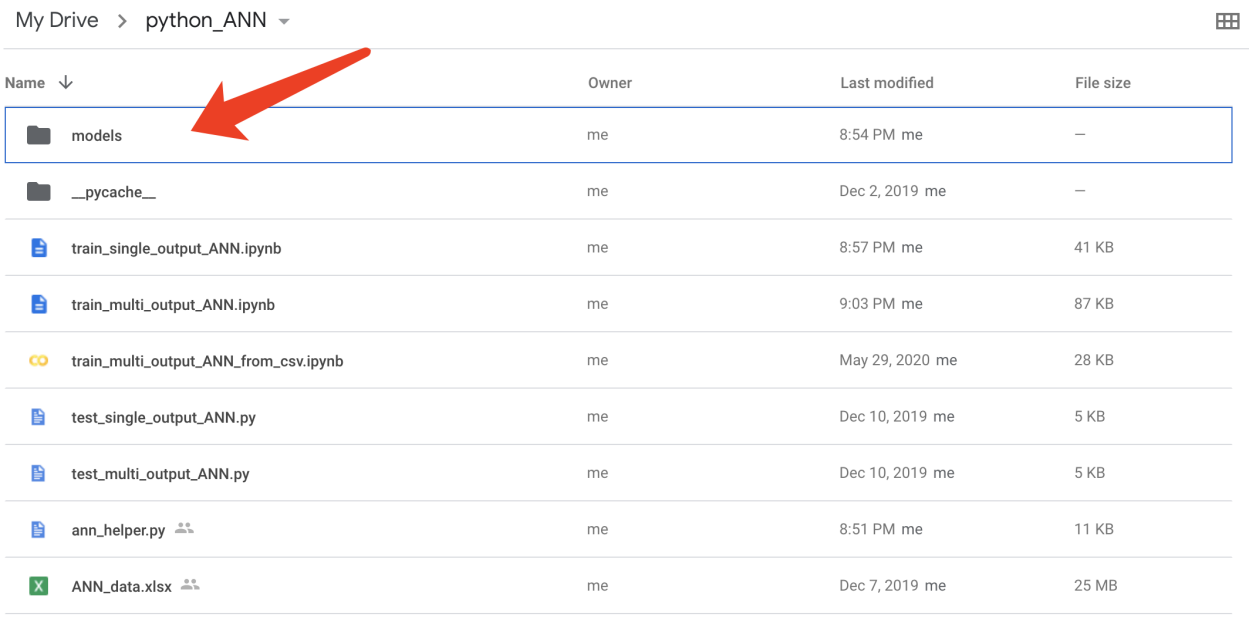


Figure 8: Saved models in Google Drive