Google Colab Setting

1) Google Colab Connection

a. You can connect to Google Colab via your google account by going to https://colab.research.google.com/. Click on “New Notebook” and start working!

b. Once you are in the new notebook, you need to choose what Processor you want. So click on Edit-> Notebook Settings and select a GPU when you need one.

c. You will see a “Connect” button on top and tada...you have a free GPU now!

2) Connecting a Colab Notebook to your Google drive

a) Once you are connected to a notebook, there will be times when you would want to connect the storage of this notebook to your drive for various reasons, like, getting input data, saving or loading your models etc etc.

b) To do that, type the below command and authenticate as instructed.

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

If you want to navigate inside any folder to fetch data/save models, you can do “cd” or “oschdir” to it.

```
%cd gdrive/My\ Drive/handouts
```

3) Connecting to Kaggle in Colab

a) For most of the homeworks the data would be huge for you to download and then upload somewhere. The best option is to directly download the data from kaggle everytime you load the notebook.

This may sound troublesome but this saves a lot of time.

b) To begin with, you need to install Kaggle directly in the Colab VM.

```
!pip install kaggle
!mkdir .kaggle
```

c) You would need to create a new API token in Kaggle as mentioned in the article below.
https://towardsdatascience.com/setting-up-kaggle-in-google-colab-ebb281b61463

d) Once Kaggle is installed, you need to set up the kaggle keys to let it identify your account of Kaggle with the current notebook settings.

```
import json
token = {"username":"your_username","key":"your_key"}
```

```
with open('/content/.kaggle/kaggle.json', 'w') as file:
    json.dump(token, file)
```

e) To set the correct token permissions, type the following.

```
!chmod 600 /content/.kaggle/kaggle.json
!cp /content/.kaggle/kaggle.json /root/.kaggle/
!kaggle config path -p /content
```

f) In case you get errors after running the above command, try running (e) again. Else, try the below method.

```
import os, zipfile, tarfile, ipdb
os.environ['KAGGLE_USERNAME'] = "your_username"
os.environ['KAGGLE_KEY'] = "your_key"
```

4) Downloading data from Kaggle

Once you do the above steps, go to kaggle competition and check the name of the competition from the URL.
Run the command like below to download the data from Kaggle.

```
!kaggle competitions download -c 11-785-s20-hw2p2-classification
```

5) Saving Models and Loading them from Drive

An important part of the models you are building is that you save them from time to time and then reload them later to improvise. A simple way to do this is to save them after some epochs.

```
torch.save({
                    'model_state_dict': model.state_dict(),
                    'optimizer_state_dict': optimizer.state_dict(),
                    'scheduler_state_dict' : scheduler.state_dict(),
        }, "gdrive/My Drive/hw2p2/"+"Model_"+str(epoch))
```

An important concept to understand here is that what all are you saving in the model. You may only save the model if you want for inference but you need to save the entire states if you want to run it further. You might want to refer to the documentation for more understanding. https://pytorch.org/tutorials/beginner/saving_loading_models.html

To load for re-run,

```
temp = torch.load("gdrive/My
Drive/hw2p2_ResNet_v1/"+"Model_Resv1_27")
network.load_state_dict(temp['model_state_dict'])
```

```
optimizer.load_state_dict(temp['optimizer_state_dict'])
scheduler.load_state_dict(temp['scheduler_state_dict'])
```

Where “network” is the model name and you would have defined your optimizer and schedulers. Do not worry, you will get to know about these terms as the course starts.

6) Limitations and Tips for Google Colab

Colab is a savior but it comes at a price.

a) You can run only 2 notebooks per google ID.

b) There is a limitation per google ID in the GPU usage per 24 hours. So be wary about how much free GPU you are utilizing. If you have the notebook on but are not using it, change to a CPU or terminate the session if possible.

c) Once, you reach the usage limit, you will be banned for up to 24 hours from using any GPU.
Pro-tip - You may want to use your other google IDs if you have.

d) Also, if you want to run your models for a long time, do not travel with your laptop, network disconnectivity will be frustrating with Colab. So, try to be in one place.

e) In case, you would have to leave your model running for long hours unattended, make sure your laptop does not sleep and put the below command for restraining Colab from detecting zero activity.

```
function ClickConnect(){
console.log("Working");
document.querySelector("colab-toolbar-button").click()
}setInterval(ClickConnect,600000)
```