

# Setup

August 1, 2023

## 1 Assignment 4

**1.1 Note: DON'T RUN THIS UNTIL YOU HAVE MADE SURE YOUR CODE WORKS LOCALLY; THIS NOTEBOOK TAKES ~1 HOUR TO RUN.**

**1.2 See <https://edstem.org/us/courses/33056/discussion/2523780> For a guide to use Colab GPUs.**

Please make a copy of this Colab notebook in your own Google Drive to edit it.

This serves as an Azure alternative in case we have a dreadful case of GPU shortage. Feel free to copy this setup for other assignments.

### 1.3 1. Upload your files

#### 1.3.1 Option 1: Upload your files to your Google Drive, and mount the drive

You can upload your files to your own Google drive. This is the "safest" way to set up the Colab, since your files are stored in your Google Drive instead of the Colab Runtime, which may terminate if you leave it on for too long (~8 hours, or if you put your laptop to sleep mode) and lose your files.

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

#### 1.3.2 Option 2: Zip your folder, upload, then unzip

Zip your code folder. Upload the folder of your code to the runtime (drag and drop) so that the structure of Files would look like the following: - .. - sample\_data - your\_folder.zip

Then run the following code block to unzip the file.

NOTE THAT IF YOU LEAVE YOUR COLAB NOTEBOOK ALONE FOR TOO LONG, THERE IS A RISK THAT YOU WILL LOSE YOUR FILES.

```
[ ]: ! unzip your_folder.zip # replace this with the name of your zip file
```

##2. Install Dependencies

Run the following cell to install dependencies for this assignment.

```
[ ]: !pip install docopt sentencepiece sacrebleu
```

### 1.4 3. Relocate to your code directory

```
[ ]: %cd directory_name  
! ls # verify that you are in the right directory
```

### 1.5 4. Run assignment-specific commands

To run on GPU, please make sure your runtime type is using the GPU accelerator! To do so, click on:

on the upper-right hand corner, then click “Change runtime type” in the bottom-left of the popped-up panel, then select “GPU” as your hardware accelerator.

Now you can directly run the commands specified by the assignment files in the following code blocks.

```
[ ]: ! sh run.sh train
```

```
[ ]: ! sh run.sh test
```

```
[ ]: # After you have verified that everything works, you can collect your submission  
! sh collect_submission.sh
```

### 1.6 5. Download your files

If you used Option 2, where your files are not directly in your Google drive, you would need to download the files you have generated in the process. We primarily need your test\_outputs.txt file in your outputs/ folder to compute the BLEU score. You can also zip the entire folder and download the whole folder in the following cell block:

```
[ ]: !zip folder_name downloading.zip  
# if the above does not work, run the following  
# !zip -r folder_name downloading.zip  
from google.colab import files  
files.download('downloading.zip')
```