

Spring 2019

Home

Announcements

Assignments

Discussions

Grades

People

Pages

Files

Syllabus

Conferences

Collaborations

Chat

Course Captures

View All Pages

Assignment 3 Description

Natural Language Processing

Welcome to the third homework of CS182, in this assignment, you will learn about processing and generating text. Specifically, you will build a neural network to generate news headlines, through the training of an LSTM-based language model. Then you will train a Transformer to summarize news articles.

Create an "assignment3" directory and download the assignment zip file from [here](#). Note the zip file unpacks directly into the parent directory, it does not create a subdirectory.

Installation

This assignment requires Python 3.5 or greater. If you plan to complete this assignment with a local machine, and not in the Google Colab, please install Python 3. You can check the python version with `python3 --version`). If not, install an updated version of python using the instructions here: <https://www.tecmint.com/install-python-in-ubuntu/>

Linux / MAC Os

Installing on Linux is very similar to the previous assignments. From the assignment directory, you can run the following:

```
virtualenv .env --python=python3
```

To create a virtual environment. Running:

```
source .env/bin/activate
```

will activate the environment. You now must install the required packages for the assignment into your virtualenv. If you are using a GPU-enabled machine, from the main assignment directory do:

```
pip3 install -r requirements_gpu.txt
```

Otherwise:

```
pip3 install -r requirements.txt
```

To use your GPU, You must have installed the CUDA9 toolkit, and you must have your library path pointing to CUDA-9 libraries. i.e.

```
echo $LD_LIBRARY_PATH
```

should include `/usr/local/cuda-9.0/lib64` on Linux. Instructions for installing CUDA are in the Tensorflow setup pages.

Make sure that you are running a version of python >= 3.5 in your Jupyter notebooks. Open one of the .ipynb assignment files through Jupyter in a browser and make sure you are using the right Python environment by going into the "Kernel" tab "change kernel" and selecting the correct kernel to run the code (should be Python 3). If you need to figure out the exact version of python you are running, create a new cell in your notebook and do

```
import sys
sys.version
```

Windows

Windows is not officially supported. There is a chance that you will be able to get this working with Anaconda on windows, however we strongly suggest using a virtual machine (virtual box) with an installation of linux, as it will likely be more stable, and the GSIs will be able to provide additional assistance.

Google Colab

We have adapted the notebook to make it simple to run in the Google Colab. If you want to use the Google Colab assignment, you will have to go to the following Google Drive folder: https://drive.google.com/open?id=1TNhUy9ldZ5mv_GLNNmCBFnLFT3DXwntF and "Add it to your Drive" (so that you can modify the files).

You must be logged in to your UC Berkeley Google account to see the folder on Google Drive.

You must still download the files, test them locally for validity and run use our submission script to produce your submission ZIP-file. You will be responsible for download your model files from Google Drive and making sure they work.

To open the Jupyter notebooks in Google drive, open the assignment Google Drive folder, double click on one of the 2 .ipynb assignments, and select "Open with Colaboratory".

Make sure you are using the correct runtime (kernel). Go to the Runtime tab > Change runtime type. You should use Python 3, and "hardware accelerator" should be "GPU".

Downloading the Data

To download the data, run the following command from the assignment root directory:

```
bash download_data.sh
```

If you get the error "bash: ./download_data.sh: Permission denied" run ``chmod +x download_data.sh`` and try again.

Preparing a Submission

To prepare a submission, run the following command from the assignment root directory:

```
bash prepare_submission.sh
```

This will create a "submission.zip" file which can be submitted for the homework. You may get a warning if the files required do not exist - please pay attention to this, as we will not be responsible if you forget to include these files in the zipped folder. Check that your submission contains:

- "1 Language Modeling.ipynb" and "2 Summarization.ipynb"
- "capita.py", "transformer.py", "transformer_attention.py", "transformer_layers.py"
- Your model files (.index, .cptk, .data) for each notebook. The names of the models you upload should match the names you put in the Jupyter notebooks. This should be a total of 6 files.

If you get the error "bash: ./prepare_submission.sh: Permission denied" run `chmod +x download_data.sh` and try again.

Submit your file to the submission page [here](#).

Questions

Ask your questions on Piazza.

If you have questions about the creation of the dataset or the assignment, contact Philippe Laban (phillab@berkeley.edu) or John Canny (canny@berkeley.edu).

If you are an instructor looking for the unreleased test-set and Solution notebooks, you can contact us as well.