

CS 6320.002: Natural Language Processing

Fall 2020

Homework 4 – 55 points

Issued 01 Nov. 2021

Due 11:59pm CDT 15 Nov. 2021

Deliverables: Your completed Google Colab notebook `hw4.ipynb` and saved model file `hw4.pickle`, uploaded to Gradescope.

Warning! This assignment trains a neural network on a large dataset. The training process will take five to six hours at the absolute minimum; you will need to train for close to 24 hours to achieve good performance. Start early so you don't run out of time!

0 Getting Started

Make sure you have downloaded the skeleton notebook `hw4.ipynb`.

In this programming assignment, as well as the next one, we will be training neural networks. This is a very slow process that requires a significant amount of compute power. We will be using Google Colaboratory, a free, cloud-based coding environment for AI research that provides GPUs, as well as more powerful CPUs than you are likely to have access to on your own.

To get started, go to your Google Drive and click the “New” button. Under the “More” submenu, you will see “Google Colaboratory.” Click to launch a Colab session with a blank, untitled notebook.

Next, upload the skeleton notebook. Click on the “File” menu and select “Upload notebook”; this will open a pop-up dialog that will allow you to upload the provided file `hw4.ipynb`. Once this loads, you are ready to go. The remaining instructions for this programming assignment are in the notebook itself.