

CS224d: Deep Learning for Natural Language Processing

Due Date: 4/19 (Tue) 11:59 PM PST. Hard deadline: 4/22 (Fri) 11:59 PM PST with 3 late days.

In this assignment we will familiarize you with basic concepts of neural networks, word vectors, and their application to sentiment analysis.

Setup

Note: Please be sure you have Python 2.7.x installed on your system. The following instructions should work on Mac or Linux. If you have any trouble getting set up, please come to office hours and the TAs will be happy to help.

Get the code: [Download the starter code here](#) and [the complementary written problems here](#).

Update: The written assignment has been updated to more closely match the notation in the slides. Here is the [old version for reference](#). If you have already completed the problems, we will accept either form.

[Optional] virtual environment: Once you have unzipped the starter code, you might want to create a [virtual environment](#) for the project. If you choose not to use a virtual environment, it is up to you to make sure that all dependencies for the code are installed on your machine. To set up a virtual environment, run the following:

```
cd assignment1
sudo pip install virtualenv      # This may already be installed
virtualenv .env                 # Create a virtual environment
source .env/bin/activate        # Activate the virtual environment
pip install -r requirements.txt # Install dependencies
# Work on the assignment for a while ...
deactivate                     # Exit the virtual environment
```

Install requirements (without a virtual environment): To install the required packages locally without setting up a virtual environment, run the following:

```
cd assignment1
pip install -r requirements.txt # Install dependencies
```

Download data: Once you have the starter code, you will need to download the Stanford Sentiment Treebank dataset. Run the following from the `assignment1` directory:

```
cd cs224d/datasets
./get_datasets.sh
```

Submitting your work

Once you are done working, run the `collectSubmission.sh` script; this will produce a file called `assignment1.zip`. Rename this file to `<your-sunet-id>.zip`, for instance if your stanford email is `jdoe@stanford.edu`, your file name should be

```
cd cs224d/datasets
jdoe.zip
```

For the written component, please upload a PDF file of your solutions to Gradescope. If you are enrolled in the class you should have been signed up automatically. If you added the class late or are not signed up, post privately to Piazza and we will add you to the roster. When asked to map question parts to your PDF, please map the parts accordingly as courtesy to your TAs. This is crucial so that we can provide accurate feedback. If a question has no written component (completely programatic), map it on the same page as the previous section or next section.

Please upload your programming submission below.

Submit Programming Assignment

Please upload the zip file containing your programming portion of the assignment. Please name your file as <your-sunet-id>.zip (e.g. jdoe.zip). If you do not have an sunet-id, then please identify yourself in the description below.

File
Upload:

Choose File

No file chosen

Description
(255
characters
maximum):

Your e-mail
(required):

Upload

This upload widget is externally owned. Uploading files here will not upload them to your Box account.

powered by **box**

Tasks

There will be four parts to this assignment. Each part has written and code components. The assignment is designed to be completed in order as later sections will leverage solutions to earlier parts. We recommend reading the assignment carefully and starting early as some parts may take significant time to run.

Q1: Softmax (10 points)

Q2: Neural Network Basics (30 points)

Q3: word2vec (40 points + 5 bonus)

Q4: Sentiment Analysis (20 points)