

CS224d: Deep Learning for Natural Language Processing

Due Date: 4/16 (Thu) 11:59 PM PST.

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](#).

[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
```

Start IPython: After you have the Stanford Sentiment data, you should start the IPython notebook server from the `assignment1` directory. If you are unfamiliar with IPython, you should read this [IPython tutorial](#).

Submitting your work

Once you are done working, put the written part in the same directory as your IPython notebook file, and 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
```

Upload this file to [the Box for this assignment](#).

For the written component, please upload a PDF file of your solutions to [Scoryst](#). Please [sign up](#) with your stanford email and SUNet ID (letter ID) if applicable. When asked to map question parts to your PDF, please map the parts accordingly as courtesy to your TAs. The last part of each problem is a placeholder for the programming component, you could just map it to the page of the last part in your written assignment.

Tasks

There will be four parts to this assignment, the first three comprise of a written component and a programming component in the IPython notebook. The fourth part is purely programming-based, and we also give you an opportunity to earn extra credits by doing a programming-based optional part. For all of the tasks, you will be using the IPython notebook `wordvec_sentiment.ipynb`.

Q1: Softmax (10 points)

Q2: Neural Network Basics (30 points)

Q3: word2vec (40 points)

Q4: Sentiment Analysis (20 points)

For these four parts, please try to finish the written component before writing code. We designed the written component to help you think through the details in your code implementation. For each part, the written component is worth 40% the points of that part, and programming is 60%.

Extra Credit (optional): Improve Your Sentiment Analysis Model (+10 points)

For this optional part, please follow the instructions in the IPython notebook to finish your implementation and report results. Extra credit will be awarded based on relative progress.