CS224n: Natural Language Processing with Deep Learning

Due Date: 1/26 (Thurs) 11:59 PM PST.
Hard deadline: 1/29 (Sun) 11:59 PM PST with 3 late days. Assignments turned in past this date will be penalized (see the grading page).

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.

Written solutions: here. We do not plan to release code solutions online. Come to office hours if you have questions about your code.

[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 and pretrained GloVe vectors. Run the following from the `assignment1` directory:

```
sh get_datasets.sh
```

Submitting your work

See Piazza for submission instructions. Remember, you MUST follow the instructions and submit your code and writeup.

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. Read the assignment carefully and start early as some parts may take significant time to run.

Q1: Softmax (10 points)

Q2: Neural Network Basics (30 points)

Q3: word2vec (40 points + 2 bonus)

Q4: Sentiment Analysis (20 points)