**Combining Movie Reviews into full\_train.txt and full\_test.txt**

Link to get the IMDb Large Movie Review Dataset: <https://ai.stanford.edu/~amaas/data/sentiment/>

**Code to Combine Movie Reviews**

Shell commands have been used to unzip the downloaded dataset and merge the reviews that are present in different text files into one; each for train and test sets.

1. Move the tar file to the directory where the data has been stored.
2. Open terminal window and cd to the directory that has the downloaded dataset with ‘.tar.gz’.
3. gunzip -c aclImdb\_v1.tar.gz | tar xopf –
4. cd aclImdb && mkdir movie\_data
5. for split in train test; do for sentiment in pos neg; do for file in $split/$sentiment/\*; do cat

cat $file >> movie\_data/full\_${split}.txt;echo >> movie\_data/full\_${split}.txt; done; done;

done;

After the following commands are run, a folder named ‘movie\_data’ will be created that consists

of two files: full\_train.txt and full\_test.txt. These two files contain all the positive and negative

reviews which are used in the main code.