

CSE 5243 HW2

Name: Neng Shi

I use the frequency of words in the sentence body to construct the feature vector for each sentence in the data set. Specifically, I first replace all the punctuations within a sentence with whitespaces and then segment the sentence into a collection of words. Finally, I count the frequency and construct an $M \times N$ matrix D , where M is the number of sentences, N is the number of words in the dataset, and $D_{i,j}$ means the count of word j in sentence i .