Class- CS 6957
Mini Project -1
Name - Trupti Mohanty

A. Submitted the word embedding file. File name: embedding_output_final_submission.txt

1. Best learning rate: 0.001
   Lowest Development set loss (Average): 5.86
   Train loss: 5.44

2.

a. Cosine similarity

Sl no	Pair	Cosine Similarity	Pair	Cosine Similarity	Remark
1.	[cat, tiger]	0.93	[plane, human]	0.71	Cat: tiger is closer
2.	[my, mine]	0.38	[happy, human]	0.69	happy: human is closer
3.	[happy, cat]	0.78	[king, princess]	0.64	happy: cat is closer
4.	[ball, racket]	0.91	[good, ugly]	0.53	ball: racket is closer
5.	[cat, racket]	0.93	[good, bad]	0.70	cat: racket is closer

b. Analogy Test

Sl no	Pair	Analogy Pair	Remark (First 5 top suggested words based on cosine similarity)
1.	king: queen	man: woman	[woman, queen, creature, soul, fellow]
2.	king: queen	prince: queen	[queen, bonds, bride, loneliness, grave]
3.	king: man	queen: woman	[woman, queen, creature, soul, fellow]
4.	woman: man	princess: princess	[princess, government, shogun, creator, doctrine]
5.	prince: princess	man: princess	[princess, woman, person, girl, dog] woman is in the second wrt similarity score.

3. **a. Word similarity test**

Pair	Cosine Similarity	Pair	Cos Similarity	Remark
pen, pencil	0.91	she, prince	0.33	pen: pencil is closer
apple, orange	0.97	man, her	0.20	apple: orange is closer
phone, battery	0.97	table, man	0.36	phone: battery is closer
table, chair	0.81	woman, he	-0.17	table: chair is closer women and he has negative score
time, year	0.61	time, pen	0.40	time: year is closer

b. Word Analogy test

Sl no	Pair	Analogy Pair
1.	dog: pet	cat: pet
2.	man: he	woman: she
3.	happy: sad	cheerful: sadness
4.	we: us	i: me
5.	you: your	he: his

4. **Evaluation of the embedding on test files**
   Word Similarity Test Pearson Correlation: 0.0558
   Accuracy on Analogy Test: 0.048

5. **FastText**

FastText provides a benchmark for text classifiers, which is more accurate and faster than the complex deep neural network models. In this paper, fastText has been used for tag prediction and sentiment analysis, which shows improved accuracy compared to the other deep learning classifier models.

CBOW is used for learning the word embeddings in an unsupervised way. However, fastText uses a simple linear layer with a lower rank where the sentence is represented as a bag of words. The first layer is the same as word embedding; furthermore, the word embeddings are averaged into a text representation, and the final layer is the linear classifier.

CBOW does not take care of the word order. However, in fastText, an additional feature can be added to represent the word order. It uses n-gram to represent the local word order partially. In this paper, bigram has been used, which shows improved accuracy as compared to base fastText.

CBOW uses the softmax function at the output layer to predict the labels, which is computationally challenging when there are many classes as the computation complexity linearly varies with the number of classes. FastText uses hierarchical softmax, which reduces the computation of the order log of the number of classes. Hierarchical softmax also has improved performance fast during test time.

Extra

PCA is used for representing the 100-dimensional word embedding in lower dimension (2D). The below figure shows the 2D projection of the given words using Principal Component Analysis. This figure shows different clusters for the prepositions, women-girl, pronouns (he, his, our she, her), and king-prince-boy. However, the queen is placed closer to the boy and prince than the woman and girl. PCA uses a linear combination of the features to determine the principal components. The first principal component axis is the one where there is maximum variability in the dataset; subsequently, the second principal component is orthogonal. The major disadvantage is that it takes care of only linear projection of the data, where the non-linear relationship in the features is not captured, and the axes are also not interpretable. While representing in PCA, we lose information, and the distance between the words also differs from the true distance in the high dimensional space.