REPORT

Assignment 1 - u1471339

1. The best learning rate observed was 0.01. The lowest development set loss is 6.017232191163811. Best Dev performance of 0.0020336863756240287 (F1 score) was at epoch 7

   For the embeddings, I used embeddings = list(model.parameters())[0]

2. a)

WORD PAIR - 1	VALUE	WORD PAIR -2	VALUE	RESULT
i) (Cat,Tiger)	0.1026	(plane,human)	0.1953	(plane,human) are more similar
ii) (my,mine)	0.1293	(happy,human)	0.0611	(my,mine) are more similar
iii) (happy,cat)	0.1185	(king,princess)	0.2633	(king,princess) are more similar
iv) (ball,racket)	0.1051	(good,ugly)	-0.0413	(ball,racket) are more similar
v) (cat,racket)	0.1656	(good,bad)	0.2889	(good,bad) are similar

2. b)

Analogy pair 1	Analogy Pair 2	Predicted Result	Expected Result
i) (King:Queen)	(man:?)	friend	Woman
ii) (King:Queen)	(prince:?)	mother	Princess
iii) (King:Man)	(queen:?)	friend	Woman
iv) (woman:man)	(princess:?)	lover	Prince
v) (prince:princess)	(man:?)	tyrant	Woman

3. A. Word Similarity Test :

Word Pair 1	Value	Word Pair 2	Value	Result
(child,kids)	0.2547	(avoid,omit)	0.1072	(child,kids) are more similar
(beautiful,beauty)	0.3087	(disliked,hatred)	0.1856	(beautiful,beauty) are more similar
(before,after)	0.4135	(politics,government)	0.2335	(before,after) are more similar

B. Analogy Test:

Analogy Pair 1	Analogy Pair 2	Predicted Result	Expected Result
(man,woman)	(grandfather:?)	wheeled	grandmother
(king,queen)	(groom:?)	renounce	bride
(woman,man)	(nurse:?)	highlanders	doctor

4. Word Similarity Test Pearson Correlation: 0.21060226272100077
   Accuracy on Analogy Test: 0.03389830508474576


```
Word Similarity Test Pearson Correlation: 0.21060226272100077
Accuracy on Analogy Test: 0.03389830508474576
```

5. The paper **"Bag of Tricks for Efficient Text Classification"** by Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov introduces various techniques and improvements for text classification tasks.

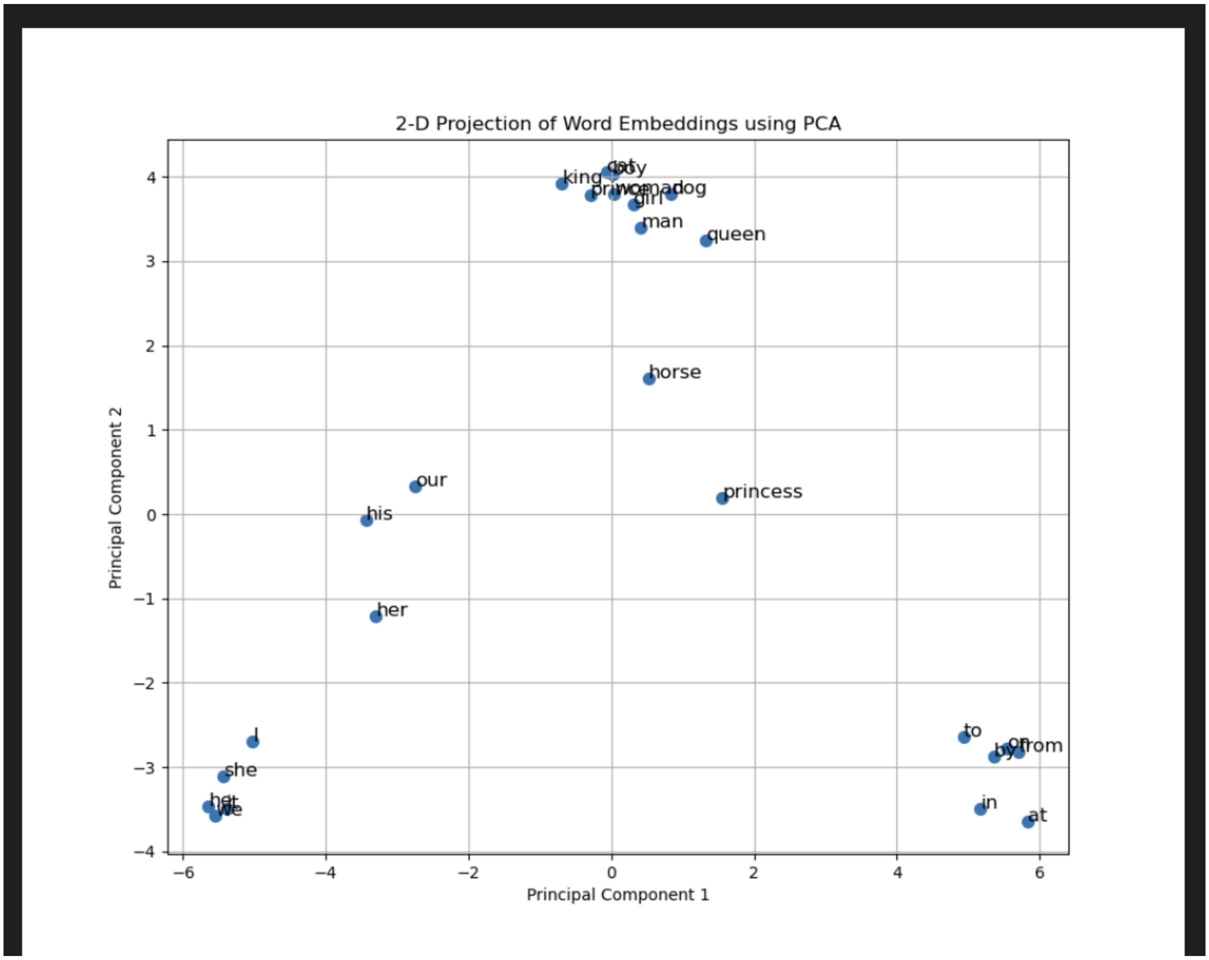
CBOW (Continuous Bag of Words):

- It is a word embedding technique used to predict a target word based on its surrounding context words. CBOW considers the surrounding words as context.The aim is to maximize the likelihood of the target word given its context words.
- The main use of CBOW is the creation of semantically rich word embeddings. It frequently appears in NLP exercises like language modeling, word analogies, and word similarity.
- Word embeddings are learned by CBOW using a neural network architecture, and the system seeks to represent words as continuous vectors. Classification is not specifically addressed.

Bag of Tricks:

- The paper addresses text classification tasks where the context is an entire document or a piece of text. The paper seeks to effectively classify text documents into predetermined categories. Instead of word embeddings, the paper focuses on text categorization challenges. It suggests a number of methods for enhancing fastText and other text classification algorithm's efficacy and efficiency. It does not directly include word prediction, unlike CBOW.
- The paper is concerned with text classification, where the goal is to assign a label or category to a text document based on its content. It covers a variety of text classification-related topics, including feature engineering, model effectiveness, and training methods.
- The article offers a number of useful text categorization approaches, including hierarchical softmax, subword embeddings, and methods for lowering training-related memory and processing demands. The purpose of these methods is to enhance the functionality of text categorization models like fastText.

6. **2-D Projection of Word Embeddings using PCA**



Observation: Similar words tend to cluster together in the embedding space. Words with similar meanings or semantic relationships are likely to be close to each other. Closer points have higher cosine similarity, indicating that they are more similar in meaning. The words “to,by,on,from” tend to be closer to each other forming a cluster. The personal pronouns “like he,she,we,I, it” form a cluster which means they are very similar to each other.. Words like her,his and our are close to each other, capturing the semantic meaning. Words like man,woman,girl,boy are close to each other that describe gender qualities. We can also see that words like king and prince are close together due to the common character of royalty that they share. Words like prince and princess must be projected closer to each other but are not due to the low accuracy of the model.

Why 2-D projection may be misleading?

1. Information Loss: There is information loss when high-dimensional word embeddings are projected into a 2-dimensional space. Subtle semantic links may be captured in the original high-dimensional space but they may be lost in 2D.

2. Distances between data points might act differently in higher-dimensional spaces than they do in two dimensions which is a misrepresentation of distance. This implies that the 2D projection's depiction of the relative distances and similarities between points may not be a realistic reflection of the relationships present in the original data. The "curse of dimensionality" might provide false conclusions.

3. Ambiguity: Words with multiple meanings or polysemous words can be challenging to represent accurately in a 2D space.

4. The word embeddings like CBOW capture word co-occurrence patterns in the context. They don't fully represent the context in which words are employed. Words with comparable embeddings might not always signify the same thing in a 2D projection because context is simplified.

5. Misleading Clustering : In 2D, data points may appear to form clusters or patterns that do not exist in higher dimensions. These apparent clusters can lead to erroneous conclusions about the nature of the data.

6. Dimensionality Reduction: PCA is a linear dimensionality reduction technique. It can be unable to grasp complicated semantic structures or nonlinear interactions that exist in the high-dimensional space. Although nonlinear dimensionality reduction methods like t-SNE may offer more useful insights, they can also have drawbacks.