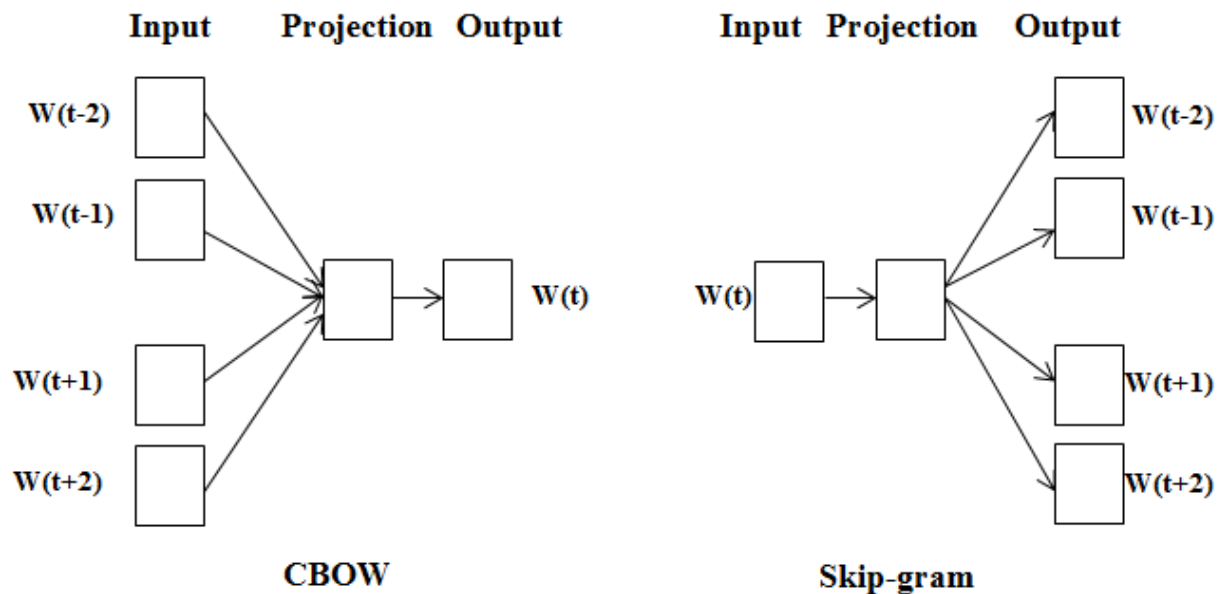


MCA Assignment 3

Text Representation and Retrieval

1. Word2Vec

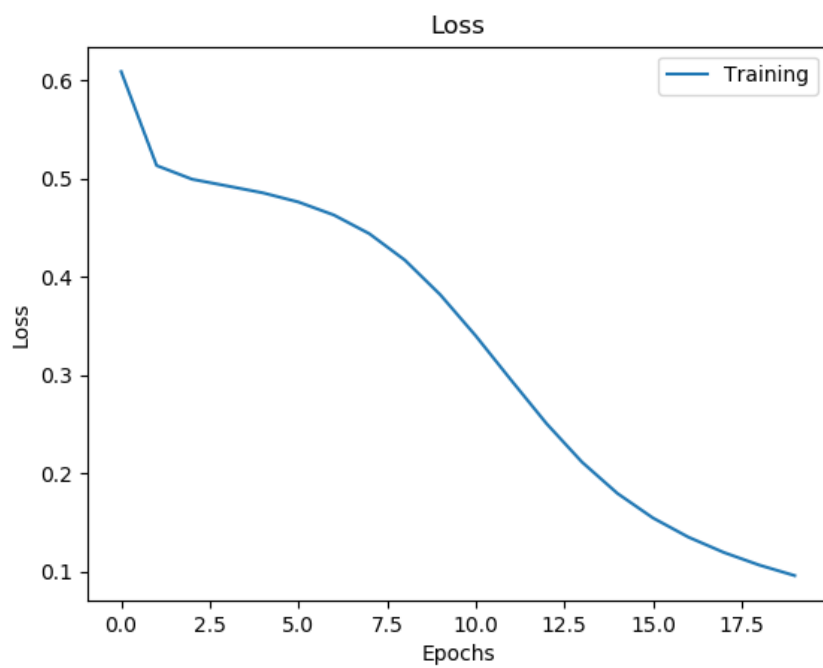
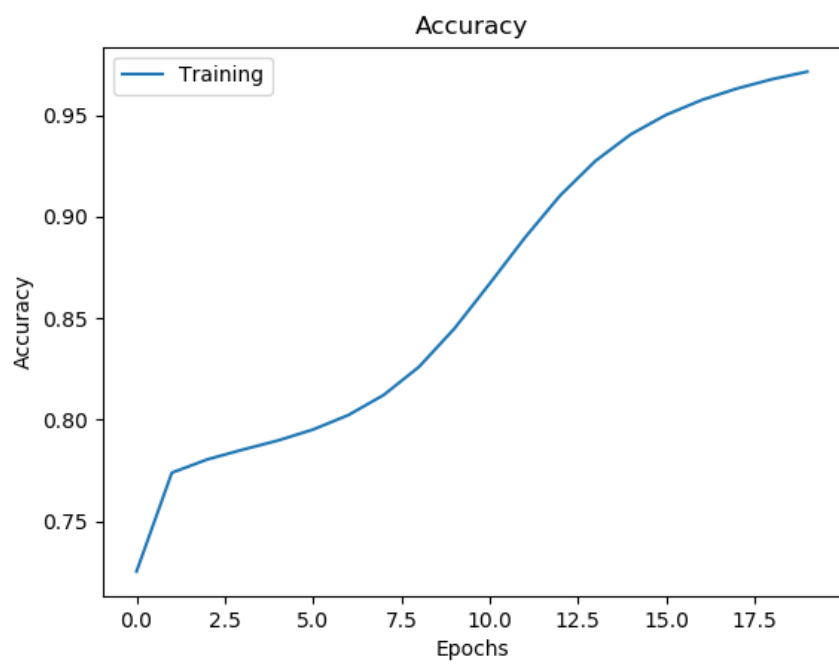
Word2Vec can be Implemented using two methods namely CBOW(Continuous Bag Of Words) and Skip-grams.



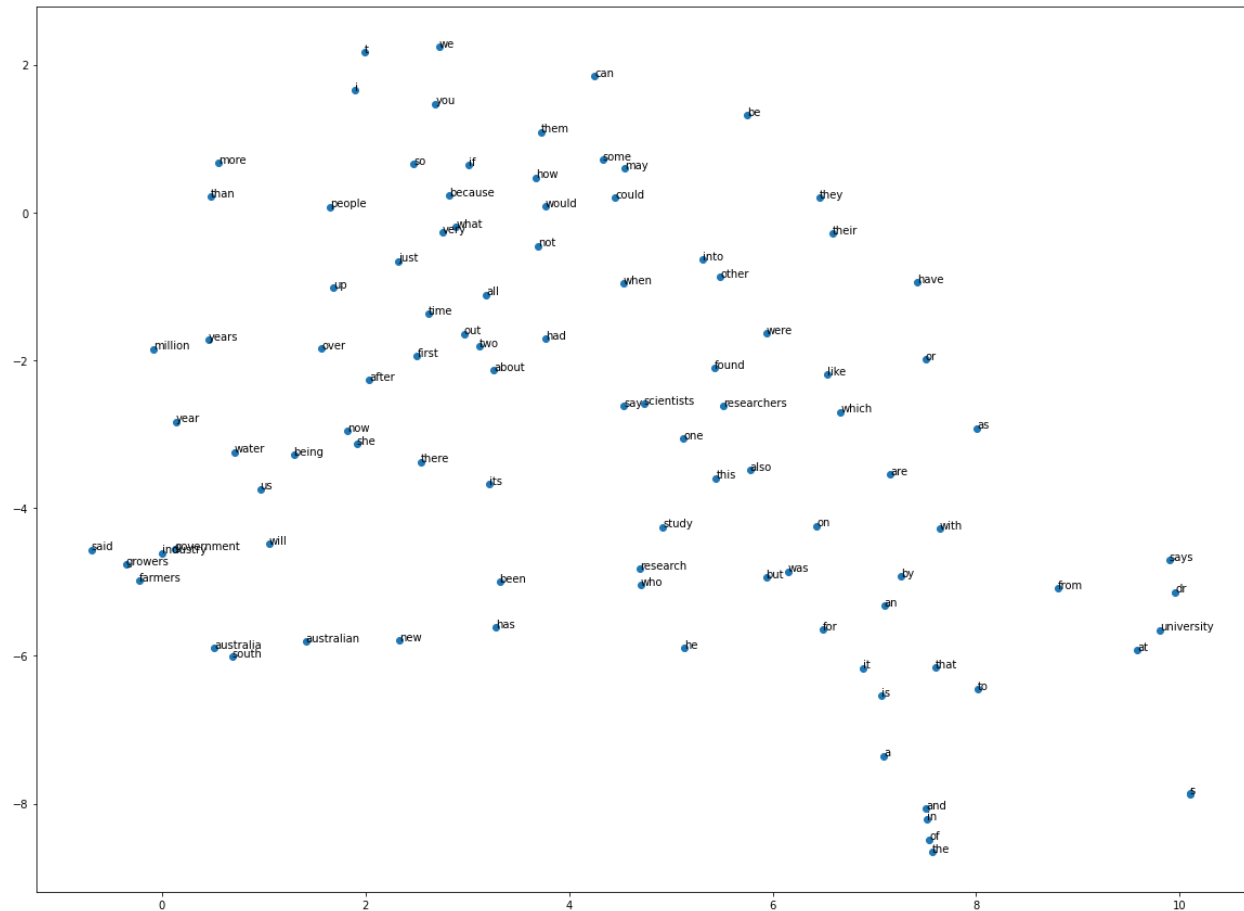
In this Implementation, I utilized the Skipgram Method. Skip-gram is one of the unsupervised learning techniques used to find the most related words for a given word.

Skip-gram is used to predict the context word for a given target word. It's the reverse of the CBOW algorithm. For the purpose of this word2vec implementation, I modified this sort of a regression problem where one predicts a context word from an input word and changed it into a classification problem where inputs are both input words and true and false context words in pairs with targets as 1 and 0 respectively for the positive context pair and negative context pair. These inputs feed into an Embedding Layer that then maps to dense layers and the target. The word Embedding is extracted from the output of the Embedding Layer of the trained model.

Model Performance



Example WordEmbedding Plot via TNSE dimensional reduction after 20 Epochs



- During Training while visualizing similar plots for every epoch I observed that words that are similar in context started coming together.

2. Relevance Feedback and Query Expansion

Retrieval Performance on 3 Epochs

```
Baseline Retrieval
```

```
MAP: 0.49176786896815833
```

```
Retrieval with Relevance Feedback
```

```
MAP: 0.5735731003211352
```

```
Retrieval with Relevance Feedback and query expansion
```

```
MAP: 0.5787353638123828
```

- The Changes in the Retrieval Performance seems in accordance with the logic that it should increase as the quality of the vectors representing the queries increase.