A2 Report

Shashwat Shivam
2016CS10328

Model Overview:-

- I used a basic CBOW model with two different weight matrices of dimensions (vocab_size x hidden_dim) and (hidden_dim x vocab_size).
- The context word embeddings generated using the first matrix were aggregated using sum operation and then passed to the second matrix.
- A softmax layer over the output predicted scores.
- Training was done using Adam optimizer and Cross Entropy Loss.
- Improvement was found in the model the 2 weight matrices were allowed to learn parameters separately rather than using the same weight matrice.

Text Preprocessing :-

- Tokenization was done using nltk word_tokenize library.
- Other preprocessing was kept to a minimum.
- A hyperparameter window-size was set to 1 to use words upto distance 1 on each side.

Using pre trained embeddings :-

- Pre trained google embeddings for tokens found in training data was used to initialize both the weight matrices mentioned in model overview bullet 1.
- I also tried to initialize embeddings for words occurring in test data but this decreased the performance.

Apart from this model i had also trained the skip gram model. This was not suited well for the task as hand as it used a single word to predict contexts. Similarity using contexts was more suitable and hence CBOW model performed much better.