Assignment 1
Deadline: Thursday, 9th May (23:59)

April 30, 2019

Question 1 (5pt):

Build word embeddings with a Keras implementation where the embedding vector is of length 50, 150 and 300. Use the Alice in Wonderland text book for training. Use a window size of 2 to train the embeddings (*window_size* in the jupyter notebook).

1. Using the CBOW model.
2. Using the Skipgram model.
3. Analyze the different word embeddings:
   - Implement your own function to perform the analogy task with[1]. Do not use existing libraries for this task such as Gensim. Your function should be able to answer whether an analogy as in example 1 is true.

$$\begin{aligned} &\text{A king is to a queen as a man is to a woman} \\ &v_{\text{king}} - v_{\text{queen}} + v_{\text{woman}} = v_{\text{man}} \end{aligned} \tag{1}$$

   - Compare the performance on the analogy task between the word embeddings that you have trained.
   - Visualize your results and interpret the results.
4. Discuss:
   - What are the main advantages of CBOW and Skipgram?
   - What are the main drawbacks of CBOW and Skipgram?

Question 2 (0pt):

Peer Review paragraph (0 points) Finally, each group member must write a single paragraph outlining their opinion on the work distribution within the group. Did every group member contribute equally? Did you split up tasks in a fair manner, or jointly worked through the exercises. Do you think that some members of your group deserve a different grade from others?

[1] Mikolov et al., (2013), Distributed Representation of Words and Phrases and their Compositionality. Check the introduction (`https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf`)