START OF QUIZ
Student ID:
74832403,Kumar,Rakesh

Question 1

Topic: Lecture 5
Source: Lecture 5

What is PMI measuring? That is, what does it mean for two words to have high PMI? (1)

Question 2

Topic: Lecture 6
Source: Lecture 6

What are two significant shortcoming of the Word2Vec model? (1)

Question 3

Topic: Lecture 8
Source: Lecture 8

What is an anaphor? (1)

Question 4

Topic: Lecture 5
Source: Lecture 5

In class, we talked about how a "typical" dimensionality for embeddings is in the range of 100-500. What might be some consequences if we estimated too low or too high? (2)

Question 5

Topic: Lecture 7
Source: Lecture 7

Why do you think that pronouns must be high salience items from previous sentences? (1)

Question 6

Topic: Lecture 8
Source: Lecture 8

Why is it easy to create negative examples for lexical coherence tests? (1)

Question 7

Topic: Lecture 7
Source: Lecture 7

The TextTiling algorithm we looked at just looked at raw word overlap (possibly with stop-word removal). Describe a way that we could improve the algorithm to maximize coverage. (2)

Question 8

Topic: Lecture 6
Source: Lecture 6

Think back to week 1 of this block when we were doing word sense disambiguation. Do you think there would be disadvantages to disambiguating all words before running word2vec? Explain. (2)

Question 9

Topic: Coding
Source: Coding

Write a short function that tries to find a good value for k in truncated SVD. You'll essentially be writing your own version of the evaluate word analogies function. For each of our the analogies, you'll need to do the vector math we were doing in the capital city determination, and return the 1-closest vector - if it's what we're looking for, it's correct. If not, it's wrong. (3)

END OF QUIZ