Quiz, 10 questions

Congratulations! You passed! Next Item Suppose you learn a word embedding for a vocabulary of 10000 words. Then the embedding vectors should be 10000 dimensional, so as to capture the full range of variation and meaning in those words. 1/1 points True False Correct What is t-SNE? A linear transformation that allows us to solve analogies on word vectors 1/1 points A non-linear dimensionality reduction technique Correct A supervised learning algorithm for learning word embeddings An open-source sequence modeling library

Suppose you download a pre-trained word embedding which has been trained on a huge corpus of text. You then use this word embedding to train an RNN for a language task of recognizing if someone is happy from a short snippet of text, using a small training set.

1/1

x (input text)

Natural Language Processing & Word Embeddings

Quiz, 10 questions

ıa	ge Processing & Word Finbeddings	1	9/10 points (90%)
	I'm bummed my cat is ill.	0	
	Really enjoying this!	1	

y (happy?)

Then even if the word "ecstatic" does not appear in your small training set, your RNN might reasonably be expected to recognize "I'm ecstatic" as deserving a label y = 1.



True

Correct

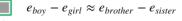


False



 $\begin{tabular}{ll} 4. & Which of these equations do you think should hold for a good word embedding? (Check all that apply) \\ \end{tabular}$

1/1 points



Correct



 $e_{boy} - e_{girl} \approx e_{sister} - e_{brother}$

Un-selected is correct



 $e_{boy} - e_{brother} \approx e_{girl} - e_{sister}$

Correct



 $e_{boy} - e_{brother} \approx e_{sister} - e_{girl}$

Un-selected is correct

Quiz, 10 questions Let E be an embedding matrix, and let e_{1234} be a one-hot vector corresponding to word 1234. Then to get the embedding of word 1234, why don't we call $E * e_{1234}$ in Python? 1/1 It is computationally wasteful. points Correct The correct formula is $E^T * e_{1234}$. This doesn't handle unknown words (<UNK>). None of the above: Calling the Python snippet as described above is fine. When learning word embeddings, we create an artificial task of estimating $P(target \mid context)$. It is okay if we do poorly on this artificial prediction task; the more important by-product of this task is that we learn a useful set of word embeddings. 1/1 points True Correct False In the word2vec algorithm, you estimate $P(t \mid c)$, where t is the target word and c is a context word. How are t and c chosen from the training set? Pick the best answer. 1/1 c is a sequence of several words immediately before t. points c and t are chosen to be nearby words. Correct

c is the sequence of all the words in the sentence before t.

Natural Language Processing & Word Embeddings

9/10 points (90%)

Quiz, 10 questions



Suppose you have a 10000 word vocabulary, and are learning 500-dimensional word embeddings. The word2vec model uses the following softmax function:

0/1 points

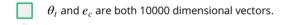
$$P(t \mid c) = \frac{e^{\theta_t^T e_c}}{\sum_{t'=1}^{10000} e^{\theta_t^T e_c}}$$

Which of these statements are correct? Check all that apply.

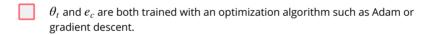


 θ_t and e_c are both 500 dimensional vectors.

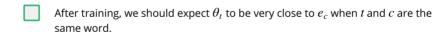
Correct



Un-selected is correct



This should be selected



Un-selected is correct



Suppose you have a 10000 word vocabulary, and are learning 500-dimensional word embeddings. The GloVe model minimizes this objective:

1/1 points

$$\min \sum_{i=1}^{10,000} \sum_{j=1}^{10,000} f(X_{ij}) (\theta_i^T e_j + b_i + b_j' - \log X_{ij})^2$$

Which of these statements are correct? Check all that apply.



Natural Language Processing & Word Embeddings

9/10 points (90%)

Quiz, 10 questions



Correct



Correct



Correct

The weighting function helps prevent learning only from extremely common word pairs. It is not necessary that it satisfies this function.



1 / 1 points You have trained word embeddings using a text dataset of m_1 words. You are considering using these word embeddings for a language task, for which you have a separate labeled dataset of m_2 words. Keeping in mind that using word embeddings is a form of transfer learning, under which of these circumstance would you expect the word embeddings to be helpful?



 $m_1 >> m_2$

Correct



 $m_1 << m_2$

9/10 points (90%)

Natural Language Processing & Word Embeddings

Quiz, 10 questions