1/1 point

**Grade received 100%** Latest Submission Grade 100% To pass 80% or higher

- 1. True/False: Suppose you learn a word embedding for a vocabulary of 60000 words. Then the embedding vectors could be 60000 dimensional, so as to capture the full range of variation and meaning in those words.
  - False
  - True

	Correct  No, the dimension of word vectors is usually smaller than the size of the vocabulary. Most common sizes for word vectors range between 50 and 1000.	
2.	True/False: t-SNE is a non-linear dimensionality reduction technique.	1 / 1 point
	○ False	
	True	
	∠ <sup>¬</sup> Expand	
	Correct t-SNE is a non-linear dimensionality reduction technique.	

3.	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 point

x (input text)	y (happy?)	
I'm feeling wonderful today!	1	
I'm bummed my cat is ill.	0	
Really enjoying this!	1	

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
- False



Correct
Yes, word vectors empower your model with an incredible ability to generalize. The vector for "ecstatic" would contain a positive/happy connotation which will probably make your model classify the sentence as a "1".

✓ Correct Yes!

✓ Correct Yes!

 $e_{boy} - e_{girl} pprox e_{sister} - e_{brother}$ 

 $ightharpoonup e_{boy} - e_{qirl} pprox e_{brother} - e_{sister}$ 

 $e_{boy} - e_{brother} pprox e_{sister} - e_{girl}$ 



## Correct

Great, you got all the right answers.

1/1 point







The correct formula is  $E^T * o_{1234}$ 

This doesn't handle unknown words (<UNK>).

It is computationally wasteful.

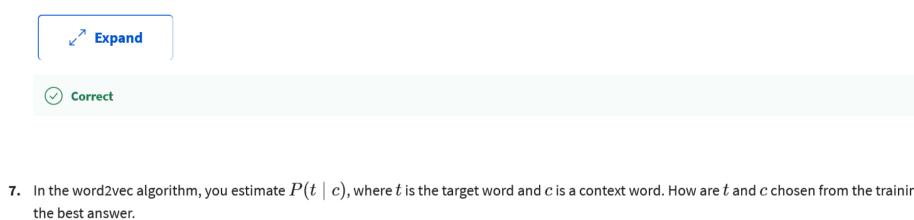
Yes, the element-wise multiplication will be extremely inefficient.

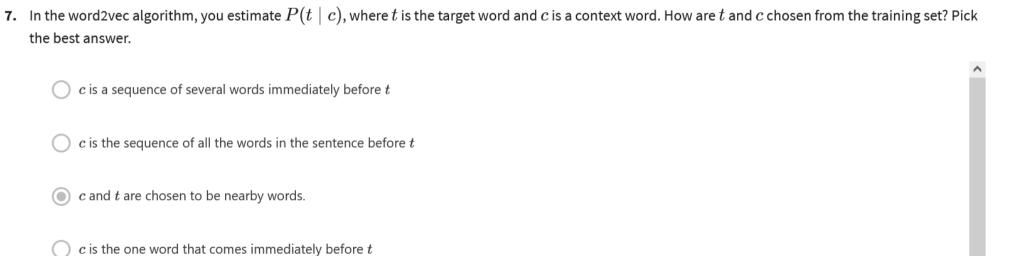
**6.** When learning word embeddings, we create an artificial task of estimating  $P(target \mid context)$ . It is okay if we do poorly on this artificial prediction task; the more important by-product of this task is that we learn a useful set of word embeddings.

1/1 point

False

True





### **⊘** Correct

8. Suppose you have a 10000 word vocabulary, and are learning 100-dimensional word embeddings. The word2vec model uses the following softmax function:

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

True/False: After training, we should expect  $heta_t$  to be very close to  $e_c$  when t and c are the same word.

- False
- True

To review this concept watch the lecture.

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

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

True/False:  $X_{ij}$  is the number of times word j appears in the context of word i.

- True
- False



 $X_{ij}$  is the number of times word j appears in the context of word i.



**⊘** Correct

10.	. You have trained word embeddings using a text dataset of $t_1$ words. You are considering using these word embeddings for a language task, for which you	1/1 point
	have a separate labeled dataset of $t_2$ words. Keeping in mind that using word embeddings is a form of transfer learning, under which of these circumstances would you expect the word embeddings to be helpful?	
	$igcup$ When $t_1$ is equal to $t_2$	
	$igcup$ When $t_1$ is smaller than $t_2$	
	$lacksquare$ When $t_1$ is larger than $t_2$	
	∠ <sup>¬</sup> Expand	

**⊘** Correct

Transfer embeddings to new tasks with smaller training sets.