



TO PASS 80% or higher

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

✓ Correct

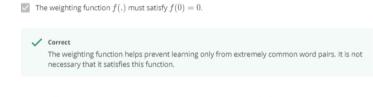
✓ Correct Yes! Keep Learning

GRADE 90%

Natural Language Processing & Word Embeddings

	LATEST SUBMISSION GRADE 90%	
1.	Suppose you learn a word embedding for a vocabulary of 10000 words. Then the emb should be 10000 dimensional, so as to capture the full range of variation and meanin words. True False	
	Correct The dimension of word vectors is usually smaller than the size of the vocabulary. No sizes for word vectors ranges between 50 and 400.	Aost common
2.	2. What is t-SNE? A linear transformation that allows us to solve analogies on word vectors A non-linear dimensionality reduction technique A supervised learning algorithm for learning word embeddings An open-source sequence modeling library	1/1 point
	✓ Correct Yes	
3.	 Suppose you download a pre-trained word embedding which has been trained on a function text. You then use this word embedding to train an RNN for a language task of recognomeone is happy from a short snippet of text, using a small training set. 	
	x (input text) y (happy?)	
	I'm feeling wonderful today!	
	I'm bummed my cat is ill.	
	Really enjoying this!	
	Then even if the word "ecstatic" does not appear in your small training set, your RNR reasonably be expected to recognize "I'm ecstatic" as deserving a label $y=1$.	N might
	Correct Yes, word vectors empower your model with an incredible ability to generalize. Th "ecstatic would contain a positive/happy connotation which will probably make yo classified the sentence as a "1".	
4.	Which of these equations do you think should hold for a good word embedding? (Che apply)	eck all that

	$igsquare$ $e_{boy} - e_{brother} pprox e_{sister} - e_{girl}$				
5.	Let E be an embedding matrix, and let o_{1234} be a one-hot vector corresponding to word 1234. Then to get the embedding of word 1234, why don't we call $E*o_{1234}$ in Python?	1/1 point			
	It is computationally wasteful.				
	\bigcirc The correct formula is $E^T * o_{1234}$.				
	This doesn't handle unknown words (<unk>).</unk>				
	None of the above: calling the Python snippet as described above is fine.				
	Correct 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			
	True				
	○ False				
	✓ Correct				
7.	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 point			
	$\bigcirc \ c$ is the one word that comes immediately before $t.$				
	$\textcircled{\scriptsize 0}$ c and t are chosen to be nearby words.				
	$\bigcirc \ c$ is the sequence of all the words in the sentence before $t.$				
	$\bigcirc \ c$ is a sequence of several words immediately before $t.$				
	✓ Correct				
8.	Suppose you have a 10000 word vocabulary, and are learning 500-dimensional word embeddings. The word2vec model uses the following softmax function:	1/1 point			
	$P(t \mid c) = rac{e^{ heta_t^{c} c_c}}{\sum_{10000} e^{ heta_t^{c} c_c}}$				
	Which of these statements are correct? Check all that apply.				
	$ec{arphi}_t$ and e_c are both 500 dimensional vectors.				
	✓ Correct				
	$\ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ $				
	$ ec{\psi} \; heta_t$ and e_c are both trained with an optimization algorithm such as Adam or gradient descent.				
	✓ Correct				
	$\hfill \Box$ After training, we should expect θ_t to be very close to e_c when t and c are the same word.				
9.	Suppose you have a 10000 word vocabulary, and are learning 500-dimensional word embeddings.The GloVe model minimizes this objective:	1/1 point			
	$\min \sum_{i=1}^{10,000} \sum_{j=1}^{10,000} f(X_{ij}) (heta_i^T e_j + b_i + b_j' - log X_{ij})^2$				
	Which of these statements are correct? Check all that apply.				
	$\hfill \qquad \theta_i$ and e_j should be initialized to 0 at the beginning of training.				
	extstyle ext				
	✓ Correct				
	V lathe weeks of the sured language is the				
	$igspace{\ } X_{ij}$ is the number of times word i appears in the context of word j.				
	✓ Correct				



10. 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?

0 / 1 point

\circ	m_1	>>	m_2	
0	m_1	<<	m_2	

Incorred