\leftarrow Back

Congratulations! You passed!

Grade received 100% To pass 80% or higher

Natural Language Processing & Word Embeddings

Latest Submission Grade 100%

1.	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 point

False

O True

○ Correct
The dimension of word vectors is usually smaller than the size of the vocabulary. Most common sizes for word vectors range between 50 and 400.

2	What	is	t-SNF2

1 / 1 point

- A non-linear dimensionality reduction technique
- A linear transformation that allows us to solve analogies on word vectors
- An open-source sequence modeling library
- A supervised learning algorithm for learning word embeddings



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.

O False

True

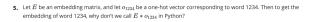
Correct
Yes, word vectors empower your model with an incredible ability to generalize. The vector for "ecstatic" would contain a positive/happy connotation which will probably make your model classify the sentence as a ""."

4. Which of these equations do you think should hold for a good word embedding? (Check all that apply)

1/1 point

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

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



1/1 point

- $\bigcirc \ \ \, \text{The correct formula is } E^T*o_{1234}.$
- None of the above: calling the Python snippet as described above is fine.
- This doesn't handle unknown words (<UNK>).
- It is computationally wasteful.

 Correct Yes, the element-wise multiplication will be extremely inefficient. 	
6. When learning word embeddings, we create an artificial task of estimating P(target 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	
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
lacktriangledown 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 the one word that comes immediately before $t.$	
\bigcirc c is a sequence of several words immediately before $t.$	
⊘ Correct	
Suppose you have a 10000 word vocabulary, and are learning 500-dimensional word embeddings. The word2vec model uses the following softmax function: T. T. T.	1/1 point
$P(t\mid c) = rac{\sigma_{t}^{p_{t}}}{\sum_{l'=1}^{p_{t}}\sigma_{t'}^{p_{t'}}}$	
Which of these statements are correct? Check all that apply.	
After training, we should expect θ_t to be very close to e_c when t and c are the same word.	
$ extstyle extstyle extstyle heta_t$ and e_c are both 500 dimensional vectors.	
○ Correct	
igstyleigstyle hinspace hinspa	
⊙ Correct	
$\ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ $	
9. Suppose you have a 10000 word vocabulary, and are learning 500-dimensional word embeddings. The GloVe model minimizes this objective: $\min \sum_{j=1}^{10,000} \sum_{j=1}^{10,000} f(X_{ij}) (\theta_i^T e_j + b_i + b_j' - log X_{ij})^2$	1 / 1 point
Which of these statements are correct? Check all that apply.	
$\hfill \hfill $	
$\ensuremath{ arnothing } heta_i$ and e_j should be initialized randomly at the beginning of training.	
⊘ Correct	
$\ensuremath{ igsigma} X_{ij}$ is the number of times word j appears in the context of word i.	
⊘ Correct	
lacksquare The weighting function $f(.)$ must satisfy $f(0)=0.$	
 Correct The weighting function helps prevent learning only from extremely common word pairs. It is not necessary that it satisfies this function. 	
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 circumstances would you expect the word embeddings to be helpful? (a) $m_1 > m_2$	1/1 point
$\bigcap_{m_1 < m_2} m_1 < m_2$	
⊙ Correct	