1. Which one of the following word representations is most likely to correspond to a word embedding representation in a general-purpose vocabulary? In other words, which one is most likely to capture meaning and important information about the words?

1 / 1 point

- ○ car -> 2

  caravan -> 3

- ◉ car -> (0.1 1)

  caravan -> (-0.1 0.9)

- ○ car -> (0 1 0 0)

  caravan -> (0 0 1 0)

- ○ car -> (1 0.1)

  caravan -> (-1 -0.9)

Correct

Correct. This is a vector representation where similar terms are closer together. (In a vehicle-specific vocabulary where cars and caravans are seen as dissimilar, d could have been a possible representation.)

2. Which one of the following statements is correct? **1 / 1 point**

- ○ To learn word embeddings you only need a vocabulary and an embedding method.
- ○ Learning word embeddings using a machine learning model is unsupervised learning as the input data set is not labelled.
- ○ The objective of a machine learning model that learns word embeddings is to predict word embeddings.
- ◉ The meaning of the words, as carried by the word embeddings, depends on the embedding approach.

Correct
Correct: the specifics of the task are what will ultimately define the meaning of the individual words, e.g. assuming that words that are surrounded by the same kinds of words have similar meaning.

3. Which one of the following statements is false? **1 / 1 point**

- ELMo may have different word embeddings for the word "stable" depending on the context.
- You need to train a deep neural network to learn word embeddings.
- word2vec-based models cannot create word embeddings for words they did not see in the corpus they were trained on.
- You can use a pre-trained BERT model to learn word embeddings on a previously unseen corpus.

Correct
Correct.

4. Consider the corpus "A robot may not injure a human being or, through inaction, allow a human being to come to harm." and assume you are preparing data to train a CBOW model. Ignoring punctuation, for a context half-size of 3, what are the context words of the center word "inaction"?

1 / 1 point

- ○ “being or through inaction allow a human”
- ○ “through inaction allow”
- ○ “being inaction human”
- ◉ “being or through allow a human”

Correct
Correct, the context words are 3 words to both the left and the right of the center word.

5. Which one of the following statements is false? **1 / 1 point**

- ○ Given the corpus "I think therefore I am", the word "think" could be represented by the one-hot vector (1 0 0 0).
- ○ Consider the corpus "A robot may not injure a human being or, through inaction, allow a human being to come to harm." and assume you are preparing data to train a CBOW model. Ignoring punctuation, for a context size of 3, the context words of the center word "inaction" are: "a", "allow", "being", "human", "or", and "through"
- ◉ The continuous bag-of-words model learns to predict context words given a center word.
- ○ Given the corpus "I think therefore I am", the word "you" cannot be represented.

Correct

Correct. It's the reverse: the continuous bag-of-words model learns to predict a center word given context words. The continuous skip-gram model, presented in an earlier video, learns to predict context words given a center word.

6. You are designing a neural network for a CBOW model that will be trained on a corpus with a vocabulary of 8000 words. If you want it to learn 400-dimensional word embedding vectors, what should be the sizes of the input, hidden, and output layers?

1 / 1 point

- ○ 8000 (input layer), 8000 (hidden layer), 400 (output layer)
- ○ 400 (input layer), 400 (hidden layer), 8000 (output layer)
- ◉ 8000 (input layer), 400 (hidden layer), 8000 (output layer)
- ○ 8000 (input layer), 400 (hidden layer), 400 (output layer)

Correct
Correct.

7. If you are designing a neural network for a CBOW model that will be trained on a corpus of 8000 words, and if you want it to learn 400-dimensional word embedding vectors, what should be the size of W1, the weighting matrix between the input layer and hidden layer, if it is fed training examples in batches of 16 examples represented by a 8000 row by 16 column matrix?

1 / 1 point

Hint: if X is the input matrix, H the matrix for the hidden layer, and B1 the bias matrix, then H = ReLU(W1X + B1).

- ○ 400 rows by 16 columns
- ○ 16 rows by 8000 columns
- ◉ 400 rows by 8000 columns
- ○ 8000 rows by 16 columns

Correct

Correct. The size of W1 does not depend on the batch size.

8. Given the input vector x below, a trained continuous bag-of-words model outputs the vector ŷ below. What is the word predicted by the model? **1 / 1 point**

x		ŷ
0.25	am	0.267
0.5	I	0.005
0	therefore	0.726
0.25	think	0.002

- ○ I
- ◉ Therefore
- ○ am
- ○ Think

Correct

Correct. Values in ŷ are interpreted as probabilities of each word being the center word. As "therefore" corresponds to the element in ŷ with the highest value, it is the model's prediction.

9. The following weighting matrix W_1 has been learned after training a CBOW model. You are also given word-to-row mapping for the input column vectors. **1 / 1 point**

$$W_1 = \begin{pmatrix} 1.76 & -0.21 & 4.56 & -2.39 & 2.11 & -2.98 \\ 1.64 & -2.5 & -2.94 & -1.3 & 0.09 & 1.81 \\ -1.9 & -1.18 & 2.61 & -1.7 & -4.36 & -4.54 \\ -3.65 & -0.3 & -1.16 & 1.75 & -4.05 & -3.38 \end{pmatrix} \qquad x = \begin{pmatrix} \\ \\ \\ \\ \\ \end{pmatrix} \begin{matrix} \text{all} \\ \text{one} \\ \text{ring} \\ \text{rule} \\ \text{them} \\ \text{to} \end{matrix}$$

What is the word embedding vector for "ring"?

- ○ [-1.9; -1.18; 2.61; -1.7; -4.36; -4.54]
- ○ [-2.39; -1.3; -1.7; 1.75]
- ◉ [4.56; -2.94; 2.61; -1.16]
- ○ [0; 0; 1; 0; 0; 0]

✓ **Correct**
Correct.

10. Select all that are correct. **1 / 1 point**

- [x] Extrinsic evaluation evaluates actual usefulness of embeddings, is time consuming and is more difficult to trouble shoot.

  Correct
  Correct.

- [x] You can perform intrinsic evaluation by using a clustering algorithm to group similar word embedding vectors, and determining if the clusters capture related words.

  Correct
  Correct

- [ ] To evaluate word embeddings with intrinsic evaluation, you use the word embeddings to perform an external task, which is typically the real-world task that you initially needed the word embeddings for. Then, use the performance metric of this task as a proxy for the quality of the word embeddings.

- [x] To evaluate word embeddings with extrinsic evaluation, you use the word embeddings to perform an external task, which is typically the real-world task that you initially needed the word embeddings for. Then, use the performance metric of this task as a proxy for the quality of the word embeddings.

  Correct
  Correct.