Your grade: 100%
Your latest: **100%** • Your highest: **100%** • To pass you need at least 80%. We keep your highest score.

Next item →

1. If the vector representations of the tokens 'I', 'love', and 'dogs' are as follows, select the bag-of-words representation of 'I love dogs'.   **1 / 1 point**

$$
I = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
\quad
love = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
\quad
dogs = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}
$$

○ I love dogs

$$
\begin{bmatrix}
0 & 1 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
1 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 1 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{bmatrix}
$$

◉ I love dogs

$$
\begin{bmatrix} 1 \end{bmatrix}
$$

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

○ I love dogs
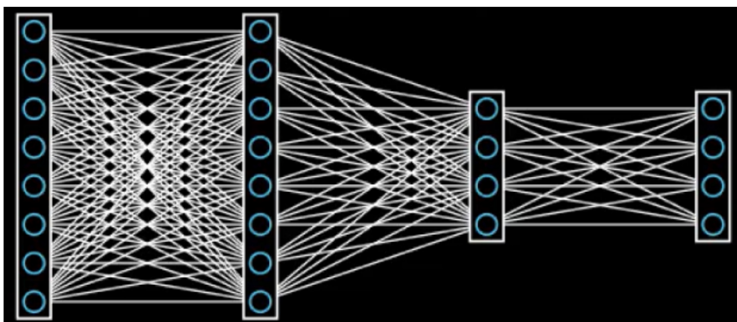
$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$$

○ I love dogs

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

✓ **Correct**
The bag-of-words representation portrays a document as the aggregate or average of one-hot encoded vectors. For 'I love dogs,' you add the one-hot vectors for 'I,' 'love,' and 'dogs'.

2. Consider the following neural network. Identify the number of neurons in the first and second embedding layers.   1 / 1 point

○ First layer: 8; Second layer: 8

⦿ First layer: 8; Second layer: 4

○ First layer: 4; Second layer: 4

○ First layer: 4; Second layer: 8

3. Which of the following expressions is used to find cross-entropy loss?   1 / 1 point

◉ $-\frac{1}{N}\sum_{n=1}^{N} log\ P(y_n|x_n, \theta)$

○ $\sum_{n=1}^{N} log\ P(y_n|x_n, \theta)$

○ $-\sum_{y} P(y_n|x_n, \theta)$

○ $-\frac{1}{N}\sum_{y}^{\hat{y}} tan\ P(y_n|x_n, \theta)$

4. Consider the following code, which is used while splitting a data set. What is the significance of specifying the batch size and shuffling the data?   1 / 1 point

```
BATCH_SIZE = 64

train_dataloader = DataLoader(split_train, batch_size=BATCH_SIZE,
shuffle=True, collate_fn=collate_batch)

valid_dataloader = DataLoader(split_valid, batch_size=BATCH_SIZE,
shuffle=True, collate_fn=collate_batch)

test_dataloader = DataLoader(test_dataset, batch_size=BATCH_SIZE,
shuffle=True, collate_fn=collate_batch)
```

○ Batch size promotes better optimization, and shuffling the data specifies the sample count for gradient approximation.

◉ Batch size specifies the sample count for gradient approximation, and shuffling the data promotes better optimization.

○ Batch size specifies the sample count for divergence, and shuffling the data reduces cross-entropy loss.

○ Batch size specifies the sample count for activation, and shuffling the data increases context size.

5. With reference to neural networks for n-gram language models, which of the following statements is **TRUE** regarding the context vector?   1 / 1 point

○ It is computed directly.

◉ It is defined as the product of your context size and the size of your vocabulary.

○ It should not be constructed by concatenating the embedding vectors.

○ It ignores the dependence on context size as it does not have a built-in mechanism to capture the order or position of words in a sentence.

6. Which of the following should be prioritized as your key performance indicator (KPI) while training the model?   1 / 1 point

○ Prediction

◉ Loss

○ Accuracy

○ Context

In training the model, prioritize the loss over accuracy as your key performance indicator. The following code appends the total loss for the epoch to your MY_LOSS list.

```python
MY_LOSS.append(total_loss/len(dataloader))
```