Your grade: 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 60%. We keep your highest score.

Next item

1. Which statement is true regarding one-hot encoding? 1 / 1 point

- (●) It converts categorical data into feature vectors.
- ( ) It sums up all the vectors from bag-of-words and multiplies the result with the embedding matrix.
- ( ) It is a representation of an entire document as the aggregate or average of vectors.
- ( ) It combines embedding weights to form a matrix.

Correct
One-hot encoding refers to converting categorical data into feature vectors. In this method, each category is assigned a unique index. A vector is created where all elements are set to zero except for the one corresponding to the category, which is set to one.

2. To determine the class of an article, logits are input from the output layer into the argmax function. What does the Argmax function do? 1 / 1 point

- ( ) It decides whether a neuron should be activated or not.
- ( ) It processes the text through the pipeline.
- (●) It identifies the index of the highest logit value.
- ( ) It corresponds to the vocabulary size.

Correct
Argmax is an operation used to find the argument that gives the maximum value from a target function. The argmax function identifies the index of the highest logit value, corresponding to the most likely class.

3. Which of the following expressions represents the gradient descent equation? 1 / 1 point

- ( ) $\mu_f = \frac{1}{N}\sum_{n=1}^{N} f(z_n)$
- ( ) $\Sigma_y P(y) \cdot (\ln(P(y)) - \ln(P(y \mid x, \theta)))$
- ( ) $P(y = i \mid x) = \frac{e^{(z^{[2]})i}}{\sum_{j=1}^{K} e^{(z^{[2]})j}}$
- (●) $\theta_{k+1} = \theta_k - \eta \nabla_\theta L(\theta_k)$

Correct
The gradient descent equation is key to reducing loss in the model. Here's the breakdown:

θ_{k+1} = Next parameter vector

θ_k = Current parameter vector

η = Learning rate (sets your step size)

$\nabla_\theta L(\theta_k)$ = Gradient of the loss function (points where the loss changes most)

By moving in the reverse direction of this gradient, you can fine-tune your parameters, thereby decreasing the loss with each step.

4. What happens during neural network training? 1 / 1 point

- ( ) Learnable parameters are encoded to generate text.
- ( ) Learnable parameters are optimized for text-to-text transfer.
- (●) Learnable parameters are fine-tuned to enhance model performance.
- ( ) Learnable parameters are decoded to generate text.

Correct
In neural network training, learnable parameters are fine-tuned to enhance model performance. This process is steered by the loss function, which serves as a measure of accuracy.

5. Match the following and select the correct option. 1 / 1 point

1. Training data	a. Evaluating real-world performance
2. Validation data	b. Hyperparameter tuning
3. Test data	c. Learning

- ○ 1, a; 2, b; 3, c
- ○ 1, b; 2, c; 3, a
- ○ 1, a; 2, c; 3, b
- ◉ 1, c; 2, b; 3, a

Correct
Typically, the data set is partitioned into three subsets: training data for learning, validation data for hyperparameter tuning, and test data for evaluating real-world performance.