

1. What is the role of an optimizer in deep learning?

- A. To compute the gradients
- B. To adjust the model's weights and biases during training
- C. To select the right learning rate
- D. To initialize the weights

Answer: B. To adjust the model's weights and biases during training

2. In the Perceptron Learning Algorithm, if the prediction is wrong, what is done?

- A. Gradients are computed
- B. The input vector is scaled and subtracted from the weight vector
- C. The input vector is added or subtracted from the weight vector
- D. The model is reset

Answer: C. The input vector is added or subtracted from the weight vector

3. What is the main purpose of backpropagation?

- A. To compute the loss
- B. To update the input data
- C. To compute gradients and adjust weights
- D. To initialize the weights

Answer: C. To compute gradients and adjust weights

4. What is the first step in the backpropagation algorithm?

- A. Compute error
- B. Update weights
- C. Perform forward propagation
- D. Perform backward propagation

Answer: C. Perform forward propagation

5. What happens during error computation in backpropagation?

- A. The predicted output is compared with the actual output
- B. The weights are updated
- C. The gradient is computed
- D. The loss function is defined

Answer: A. The predicted output is compared with the actual output

6. Which of the following is the optimization algorithm that updates weights using gradients to minimize the error in the model?

- A. Backpropagation
- B. Stochastic Gradient Descent

C. Perceptron Learning

D. Activation Function

Answer: B. Stochastic Gradient Descent

7. What does the gradient descent algorithm aim to do in deep learning?

A. Minimize the cost function

B. Maximize the cost function

C. Compute the loss function

D. Choose the best learning rate

Answer: A. Minimize the cost function

8. Which of the following is NOT a disadvantage of gradient descent?

A. High computational cost

B. Can get stuck in local minima

C. Sensitive to the initialization

D. Frequent updates of model parameters

Answer: D. Frequent updates of model parameters

9. What is the advantage of Stochastic Gradient Descent (SGD)?

A. It requires more memory than Batch Gradient Descent

B. It updates model parameters after each training example

C. It is slower to converge compared to Batch Gradient Descent

D. It uses the entire dataset to compute each update

Answer: B. It updates model parameters after each training example

10. What is a disadvantage of Stochastic Gradient Descent (SGD)?

A. It may overshoot the optimal solution

B. It requires more memory

C. It converges too slowly

D. It does not update model parameters

Answer: A. C.

11. Mini-batch Gradient Descent updates model parameters after:

A. Each data point

B. A fixed number of data points

C. The entire dataset

D. The first batch only

Answer: B. A fixed number of data points

12. Which of the following is a benefit of Mini-batch Gradient Descent?

- A. Requires more memory
- B. Slower convergence
- C. Updates are more stable than in SGD
- D. Can be stuck at local minima for long

Answer: C. Updates are more stable than in SGD

13. What is the purpose of the momentum term in momentum-based gradient descent?

- A. To reduce the learning rate
- B. To increase the learning rate
- C. To accelerate the convergence by adding past gradient information
- D. To adjust the number of iterations

Answer: C. To accelerate the convergence by adding past gradient information

14. Which of the following is true about Nesterov Accelerated Gradient (NAG)?

- A. It calculates gradients only at the current position
- B. It uses momentum without adjusting the direction
- C. It peeks ahead before estimating gradients
- D. It never overshoots the global minimum

Answer: C. It peeks ahead before estimating gradients

15. In the Adam optimizer, what does it combine from other optimizers?

- A. Momentum and SGD
- B. RMSProp and momentum
- C. AdaGrad and momentum
- D. Gradient descent and backpropagation

Answer: B. RMSProp and momentum
