

1. Which of the following is an example of an unsupervised learning algorithm?

- A) Support Vector Machine
- B) Principal Component Analysis
- C) Logistic Regression
- D) Decision Tree

Answer: B

2. In ensemble methods, what is 'bagging' typically used for?

- A) To decrease variance
- B) To decrease bias
- C) To optimize the loss function
- D) To increase interpretability

Answer: A

3. Which of the following properties best distinguishes a reinforcement learning problem?

- A) The presence of labeled training data
- B) Learning from rewards and penalties in an environment
- C) The assumption that data points are independent of each other
- D) The use of unsupervised clustering techniques

Answer: B

4. Which of the following indicates that a model is underfitting the data?

- A) High training error and low validation error
- B) Low training error and high validation error
- C) High training error and high validation error
- D) Low training error and low validation error

Answer: C

5. In ensemble learning, what is the key assumption behind boosting?

- A) Weak learners can be combined sequentially to create a strong learner.
- B) All models in the ensemble must be independent of each other.
- C) The data needs to be divided into disjoint subsets.
- D) Boosting is effective only for unsupervised learning.

Answer: A

6. Which optimizer is an extension of stochastic gradient descent and adjusts the learning rate based on past gradients?

- A) Adam
- B) RMSProp
- C) Adagrad
- D) SGD

Answer: A

7. What does a 'vanishing gradient' problem typically affect?

- A) The performance of shallow neural networks
- B) The training of deep neural networks
- C) The dimensionality of the input data
- D) The size of the training dataset

Answer: B

8. Which of the following best describes the term 'early stopping' in deep learning?

- A) A technique to speed up training by skipping unnecessary layers.
- B) A regularization method to stop training once the model's performance on the validation set stops improving.
- C) A method to reduce the learning rate during training.
- D) A strategy to train models with fewer epochs for faster results.

Answer: B

9. Why are residual connections critical in very deep neural networks?

- A) They prevent the gradient from becoming too large.
- B) They mitigate the vanishing gradient problem by enabling gradient flow across layers.
- C) They reduce the overall number of parameters in the network.
- D) They ensure that the network output is always bounded.

Answer: B

10. What does the term "learning rate schedule" refer to in training deep learning models?

- A) Adjusting the batch size dynamically during training
- B) Decreasing the learning rate based on a predefined rule or performance metric
- C) Gradually increasing the model's capacity during training
- D) Restarting the training process at fixed intervals

Answer: B

11. What is the primary function of pooling layers in a CNN?

- A) To add non-linearity to the model
- B) To reduce the spatial dimensions of the feature maps
- C) To learn features from the data
- D) To normalize the feature maps

Answer: B

12. In object detection, which algorithm combines region proposal and classification in one step?

- A) R-CNN
- B) Fast R-CNN
- C) YOLO (You Only Look Once)
- D) SVM

Answer: C

13. In Faster R-CNN, what is the role of the Region Proposal Network (RPN)?

- A) To generate anchor boxes for object detection
- B) To classify regions as object or background
- C) To suggest candidate object regions in the feature map
- D) To refine bounding box coordinates for final output

Answer: C

14. What is the primary advantage of depth-wise separable convolutions in CNNs?

- A) They improve spatial resolution in feature maps.
- B) They reduce the number of parameters and computational cost.
- C) They enhance the ability to handle overfitting.
- D) They improve the interpretability of convolutional layers.

Answer: B

15. Which of the following techniques is commonly used for instance segmentation in computer vision?

- A) U-Net
- B) Mask R-CNN
- C) YOLO
- D) ResNet

Answer: B

16. What is the primary role of the attention mechanism in transformer models?

- A) To compress text into fixed-length vectors
- B) To allow the model to focus on relevant parts of the input sequence
- C) To preprocess text before embedding
- D) To reduce overfitting in large datasets

Answer: B

17. Which pre-trained model is commonly used for transfer learning in NLP?

- A) ResNet
- B) BERT
- C) AlexNet
- D) LeNet

Answer: B

18. What does 'beam search' optimize in the context of NLP models?

- A) The choice of hyperparameters
- B) The decoding process by exploring multiple output sequences
- C) The training speed of models
- D) The initialization of weights

Answer: B

19. What is the purpose of the "CLS" token in BERT?

- A) To serve as a separator between input sentences
- B) To represent the pooled representation of the input for classification tasks
- C) To signal the beginning of each sentence in a sequence
- D) To improve attention mechanism during training

Answer: B

20. In GPT models, what is the primary purpose of causal masking?

- A) To prevent the model from attending to future tokens in the sequence
- B) To ensure that all words in the sequence are treated equally
- C) To improve the efficiency of training by reducing the sequence length
- D) To prioritize rare words during training

Answer: A