

Advanced AI and Machine Learning Fundamentals

Computer Science · Answer Key · 15 Questions

1. Which activation function is defined by the mathematical expression $f(x) = \max(0, x)$ and is widely used to mitigate the vanishing gradient problem in deep neural networks?

- A) Sigmoid
- B) Tanh

C) Rectified Linear Unit (ReLU)

- D) Softmax

2. In the context of supervised learning, what is the specific term for the phenomenon where a model performs exceptionally well on training data but fails to generalize to unseen data?

- A) Underfitting

B) Overfitting

- C) Bias-Variance Tradeoff
- D) Gradient Descent

3. Which mathematical theorem provides the basis for updating the probability of a hypothesis as more evidence or information becomes available?

A) Bayes' Theorem

- B) Central Limit Theorem
- C) Pythagorean Theorem
- D) Fundamental Theorem of Calculus

4. In a Convolutional Neural Network (CNN), what is the primary purpose of a 'pooling layer'?

- A) To increase the number of trainable parameters
- B) To perform non-linear activation

C) To reduce the spatial dimensions of the input volume

- D) To normalize the weights of the network

5. What does the 'k' represent in the K-Nearest Neighbors (KNN) algorithm?

- A) The number of iterations for training

B) The number of nearest data points to consider for classification

- C) The constant factor in the learning rate
- D) The total number of clusters in the dataset

6. Which optimization algorithm is an extension of Stochastic Gradient Descent that utilizes the moving average of gradients to accelerate convergence?

- A) Backpropagation
- B) Adam (Adaptive Moment Estimation)**
- C) Principal Component Analysis
- D) Linear Regression

7. What is the specific role of 'Dropout' in a deep neural network during the training process?

- A) To increase the learning rate
- B) To prevent overfitting by randomly disabling neurons**
- C) To convert input data into a vector format
- D) To reduce the time required for data pre-processing

8. Which term describes the process of transforming raw data into a format suitable for machine learning algorithms, often involving scaling and encoding?

- A) Feature Engineering**
- B) Backpropagation
- C) Supervised Learning
- D) Reinforcement Learning

9. In reinforcement learning, what does the 'exploration vs. exploitation' trade-off refer to?

- A) Choosing between supervised and unsupervised learning
- B) Deciding between trying new actions or choosing known high-reward actions**
- C) The balance between CPU and GPU utilization
- D) The trade-off between training speed and model accuracy

10. What is the primary function of the 'Softmax' function when used in the output layer of a multi-class classification neural network?

- A) To normalize the output into a probability distribution**
- B) To transform negative values into zero
- C) To increase the weight of hidden neurons
- D) To calculate the loss function

11. Which type of machine learning model is specifically designed to handle sequential data, such as time series or natural language, by maintaining a hidden state?

- A) Recurrent Neural Network (RNN)**
- B) Decision Tree
- C) Random Forest
- D) Support Vector Machine

12. In data science, what is the 'Curse of Dimensionality'?

- A) The inability of computers to process large datasets
- B) The observation that data becomes sparse as the number of features increases**
- C) The requirement for more memory to store high-dimensional arrays
- D) The difficulty of visualizing more than three dimensions

13. What is the purpose of the 'learning rate' hyperparameter in gradient-based optimization?

- A) To define the depth of the neural network
- B) To determine the step size at each iteration while moving toward a minimum**
- C) To select the number of epochs for training
- D) To initialize the weights of the model

14. Which clustering algorithm relies on the calculation of centroids and iteratively assigns data points to the nearest cluster center?

- A) K-Means**
- B) Logistic Regression
- C) Naive Bayes
- D) Linear Discriminant Analysis

15. In unsupervised learning, which technique is used to reduce the number of variables in a dataset while retaining as much variance as possible?

- A) Principal Component Analysis (PCA)**
- B) Gradient Boosting
- C) Cross-Validation
- D) Data Augmentation