Your grade: 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 70%. We keep your highest score.

Next item

1. What is the primary objective of machine learning? 1 / 1 point

- ○ Building intelligent robots
- ○ Teaching computers to think like humans
- ○ Transforming data into ordered collections of features
- ◉ Enabling computers to find hidden insights from data

Correct
Correct! Machine learning aims to empower computers to analyze and learn from data, uncovering patterns and insights that may not be easily discernible through conventional programming approaches.

2. How do data engineers transform categorical features into a suitable format for analysis or machine learning algorithms? 1 / 1 point

- ○ By scaling the features
- ◉ By employing methods like one-hot encoding or label encoding
- ○ By reducing the number of features
- ○ By identifying the most relevant features

Correct
Correct! One-hot encoding and label encoding are common techniques that data engineers use for converting categorical features into a suitable format for analysis.

3. Why is the model lifecycle considered iterative? 1 / 1 point

- ○ It requires constant evaluation of the model.
- ○ It involves repeating the processes in a specific order.
- ○ It involves defining the problem multiple times.
- ◉ It requires going back and forth between the processes.

Correct
Correct! The model lifecycle is considered iterative because it often requires going back and forth between the different processes. If problems are identified with the deployed models, it may be necessary to revisit earlier steps in the lifecycle.

4. Why does unsupervised learning have fewer models and evaluation methods compared to supervised learning? 1 / 1 point

- ◉ Unsupervised learning deals with unlabeled data, which limits the availability of models and evaluation methods.
- ○ Unsupervised learning algorithms are simpler, requiring fewer models and evaluation methods.
- ○ Unsupervised learning does not require evaluation methods as the outcomes are less controllable.
- ○ Unsupervised learning is a less popular approach, resulting in fewer models and evaluation methods.

Correct
Correct! Unsupervised learning deals with unlabeled data, which poses a challenge in terms of the availability and applicability of models and evaluation methods. Since there are no predefined labels, it becomes harder to evaluate the performance of unsupervised learning algorithms.

5. Which of the following regression techniques is used to handle multicollinearity and avoid overfitting in the presence of multiple independent variables? 1 / 1 point

- ○ Gradient boosting
- ◉ Ridge regression
- ○ Polynomial regression
- ○ Support vector regression

Correct
Correct! Ridge regression is a regularization technique that adds a penalty term to the regression model, which helps in handling multicollinearity and reducing overfitting.

6. Which algorithm predicts the probability of a class? 1 / 1 point

- ○ K-nearest neighbor (KNN)
- ○ Decision trees
- ○ Support vector machines
- ◉ Logistic regression

Correct
Correct! Logistic regression predicts the probability of a class in classification problems.

7. Which error measurement in regression calculates the average of squared differences between the prediction and the true output? 1 / 1 point

- ◉ Mean squared error (MSE)
- ○ Root mean squared error (RMSE)
- ○ R-squared
- ○ Mean absolute error (MAE)

Correct
Correct! Mean squared error (MSE) calculates the average of squared differences between the prediction and the true output in regression modeling.

8. What is the purpose of splitting data into a training set and a test set in machine learning? 1 / 1 point

- ○ To increase the model's complexity
- ○ To ensure all data points are used for training
- ○ To reduce the number of features in the dataset
- ◉ To evaluate how well the model performs on new data

Correct
Correct! Splitting data into a training set and a test set allows us to evaluate how well the model performs on new data.

9. Which clustering algorithm is based on the concept of density and identifies clusters as dense regions separated by areas of lower density? 1 / 1 point

- ○ K-means
- ◉ DBSCAN
- ○ Agglomerative clustering
- ○ Divisive clustering

Correct
Correct! DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a density-based clustering algorithm that identifies clusters based on the concept of density. It separates clusters as dense regions separated by areas of lower density.

10. Which AI-powered coding assistant developed by OpenAI helps developers write code efficiently by analyzing code and generating suggestions? **1 / 1 point**

- (●) GitHub Copilot
- ( ) Bard
- ( ) DeepDream
- ( ) AlphaFold

Correct
Correct! GitHub Copilot is an AI-powered coding assistant developed by OpenAI, and GitHub is designed to help developers write code efficiently by analyzing code and generating suggestions.