

Introduction to Machine Learning

Assignment - 0

TYPE OF QUESTION: MCQ/MSQ

Number of questions: 15

Total mark: $2 * 15 = 30$

Q1.

What is the main goal of machine learning?

- a. To write algorithms that can perform specific tasks
- b. To develop intelligent machines that can think and reason
- c. To enable computers to learn from data and improve performance over time
- d. To create systems that can mimic human behavior

Correct Answer: c

Detailed Solution: The main goal of machine learning is to develop algorithms and models that can learn from data and improve their performance on specific tasks without being explicitly programmed.

Q2.

What is the purpose of the training set in machine learning?

- a. To test the performance of the model on new data
- b. To fine-tune the hyperparameters of the model
- c. To evaluate the performance of the model
- d. To teach the model to make predictions based on labeled examples

Correct Answer: d

Detailed Solution: The training set is used to teach the machine learning model by providing labeled examples. The model learns patterns and relationships in the data to make accurate predictions on new, and unseen data.

Q3.

You are given reviews of a few movies marked as positive, negative, or neutral. Classifying reviews of a new movie is an example of

- a. Reinforcement learning
- b. Semi-Supervised learning
- c. Unsupervised learning
- d. Supervised learning

Correct Answer: d

Detailed Solution: Supervised learning is defined as learning from labeled examples. Given labeled reviews, the model can learn from them and predict new examples.

Q4.

In machine learning, what is the definition of a feature?

- a. A measurement or characteristic that describes an input data point.
- b. The output or prediction generated by a machine-learning model.
- c. The process of transforming raw data into a usable format for machine learning algorithms.
- d. The evaluation metric used to assess the performance of a machine learning model.

Correct Answer: a

Detailed Solution: In machine learning, a feature refers to a measurable or observable characteristic of an input data point. Features are used to represent the data and provide information that helps the machine learning algorithm understand and make predictions or classifications. Features can be numerical, categorical, or binary and are typically represented as columns or variables in a dataset.

Q5.

What is the goal of feature selection in machine learning?

- a. To transform categorical variables into numerical representations
- b. To balance the class distribution in imbalanced datasets
- c. To reduce the dimensionality of the dataset
- d. To normalize the data for better model performance

Correct Answer: c

Detailed Solution: Feature selection aims to identify and select the most informative features from the dataset, thereby reducing the dimensionality. It helps in improving model performance, reducing computational complexity, and avoiding the curse of dimensionality.

Q6.

What is the purpose of feature scaling in machine learning?

- a. To increase the complexity of the model
- b. To reduce the dimensionality of the data
- c. To normalize the range of features to a similar scale
- d. To eliminate outliers from the dataset

Correct Answer: c

Detailed Solution: Feature scaling is used to ensure that all features or variables in the dataset are on a similar scale. This is important because some machine learning algorithms are sensitive to the scale of features. Scaling the features helps in achieving faster convergence, preventing dominant features from overshadowing others, and ensuring the algorithm treats all features equally.

Q7.

Which of the following is a categorical feature?

- a. Amount of rainfall in a day
- b. Price of petroleum
- c. Mother tongue of a person
- d. Height of a person

Correct Answer: c

Detailed Solution: Categorical variables represent types of data that may be divided into groups. All other features are continuous.

Q8.

What is the primary goal of supervised learning?

- a. To cluster similar data points together
- b. To uncover hidden patterns in the data
- c. To learn from labeled examples and make predictions or decisions on unseen data
- d. To learn from rewards and punishments to maximize performance

Correct Answer: c

Detailed Solution: The primary goal of supervised learning is to learn from labeled examples, where the input data is paired with corresponding output labels. The model is trained to make predictions or decisions on unseen data based on the patterns learned from the labeled examples.

Q9.

Which of the following task(s) can be solved using unsupervised learning?

- a. Image classification
- b. Customer segmentation
- c. Anomaly detection
- d. Spam email detection

Correct Answers: b, c

Detailed Solution: Unsupervised learning is used for tasks like customer segmentation, where patterns or clusters are identified in the data, and anomaly detection, where abnormal data points are identified.

Q10.

Which of the following is a classification task?

- a. Predict the number of copies of a book that will be sold this month
- b. Predict the price of a house based on floor area, number of rooms, etc.
- c. Predict the age of a person
- d. Predict whether there will be abnormally heavy rainfall next year

Correct Answer: d

Detailed Solution: From the definition of the classification task.

Q11.

Which of the following statement(s) is/are true about unsupervised learning in machine learning?

- a. Unsupervised learning algorithms require labeled training data
- b. Unsupervised learning algorithms discover patterns and structures in unlabeled data
- c. Clustering and dimensionality reduction are examples of unsupervised learning techniques
- d. Unsupervised learning is used for classification tasks
- e. Anomaly detection is a common application of unsupervised learning

Correct Answers: b, c, e

Detailed Solution: Unsupervised learning is a type of machine learning where the algorithm learns from unlabeled data without any specific target variable or labels. The goal is to discover patterns, structures, or relationships within the data.

Clustering is a technique in unsupervised learning that groups similar data points together, while dimensionality reduction aims to reduce the number of features while retaining important information.

Anomaly detection involves identifying abnormal or unusual patterns in the data. Unsupervised learning does not require labeled training data or perform classification tasks.

Q12.

Which of the following statement(s) is/are true about supervised learning in machine learning?

- a. Supervised learning requires labeled training data
- b. The goal of supervised learning is to discover hidden patterns in unlabeled data
- c. Classification and regression are examples of supervised learning tasks
- d. Supervised learning algorithms can make predictions on new, unseen data
- e. K-means and Hierarchical Clustering are supervised learning algorithms

Correct Answers: a, c, d

Detailed Solution: Supervised learning is a type of machine learning where the algorithm learns from labeled training data, where each data instance has corresponding input features and output labels.

The goal of supervised learning is to learn a mapping between the input features and the output labels in order to make predictions on new, unseen data.

Classification is a supervised learning task where the goal is to predict the class or category of a given input, while regression is a supervised learning task where the goal is to predict a continuous value.

K-means and Hierarchical Clustering are examples of unsupervised learning algorithms, not supervised learning.

Q13.

Which of the following technique(s) is/are used for handling missing data in machine learning?

- a. Removing the rows with missing data
- b. Imputing the missing values with mean, median, or mode
- c. Imputing missing values with zero
- d. Using regression models to predict missing values
- e. Ignoring the missing values and training the model

Correct Answers: a, b, d

Detailed Solution: Missing data is a common issue in real-world datasets, and various techniques can be used to handle it. Removing the rows with missing data is one approach, but it may result in a loss of information if the missing data is substantial.

Imputing the missing values with mean, median, or mode is a common technique where the missing values are replaced with the central tendency of the corresponding feature.

Imputing missing values with zero is not a suitable approach unless zero has a specific meaning in the context of the data.

Using regression models to predict missing values is another approach where the model is trained using the available data and used to predict the missing values.

Ignoring the missing data during training is generally not recommended as it can lead to biased or incomplete models.

Q14.

Which of the following is NOT a type of machine learning algorithm?

- a. Supervised learning
- b. Unsupervised learning
- c. Reinforcement learning
- d. Data preprocessing

Correct Answer: d

Detailed Solution: Data preprocessing is not a type of machine learning algorithm. It refers to the steps taken to prepare and clean the data before applying machine learning techniques.

Q15.

Which of the following is NOT a common evaluation metric used in machine learning?

- a. Accuracy
- b. Precision
- c. Recall
- d. Profit margin

Correct Answer: d

Detailed Solution: Accuracy, precision, and recall are commonly used evaluation metrics in machine learning. Accuracy measures the overall correctness of the model's predictions. Precision measures the proportion of true positive predictions out of all positive predictions, while recall measures the proportion of true positive predictions out of all actual positive instances. These metrics are used to assess the performance of classification models.

On the other hand, profit margin is not a direct evaluation metric used in machine learning. Profit margin is a business metric that relates to the financial performance of a company and is not specifically tied to model evaluation. While machine learning models can be used to optimize business decisions and improve profitability, the evaluation metrics for the models are typically focused on their predictive performance rather than financial metrics like profit margin.
