

What is an outlier ?

- ✓ A. An outlier is a data point that differs significantly from other observations in a dataset.
- B. A data point that is the median value of a dataset.
- C. A data point that represents the most common value in a dataset
- D. A data point that is similar to other observations in a dataset.

Why is feature scaling important in machine learning?

- ✓ A. To prevent some features from dominating others during model training.
- B. It simplifies the dataset by removing unnecessary features.
- C. To improve the interpretability of the model
- D. To decrease the complexity of the model.

Outliers can occur due to

- ✓ A. Measurement or recording errors
- ✓ B. Genuine extreme values in population
- ✓ C. Inherent variability in the data

If there are outliers in the dataset

- ✓ A. Balanced feature scaling among features cannot be guaranteed if we used min-max scaling
- B. Balanced feature scaling among features can be guaranteed if we used standardization scaling
- C. Balanced feature scaling among features can be guaranteed if we used max-abs scaling
- D. Balanced feature scaling among features can be guaranteed always

Which of the following statements best describes the purpose of a validation set in machine learning?

- A. It is used to evaluate the model's performance on data that it has never seen before.
- ✓ B. It is used to fine-tune the model's hyperparameters and assess its generalization ability.
- C. It is used to train the model on a subset of data before testing it on the entire dataset.
- D. It is used to compare the model's predictions with the ground truth labels to calculate accuracy.

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

- A. To evaluate the model's performance.
- B. To fine-tune the model's hyperparameters.
- ✓ C. To learn the model's parameters using labeled data.
- D. To compare different machine learning algorithms.

What is the purpose of using a test set?

- A. To train the model's parameters.
- B. To fine-tune the model's hyperparameters.
- ✓ C. To evaluate the model's performance on unseen data.
- D. To provide feedback for adjusting the model during training.

Which of the following factors contributes to irreducible error in machine learning?

- A. Model complexity.
- B. Training time.
- C. Data preprocessing techniques.
- ✓ D. Inherent variability and noise in the data.

Which of the following statements best describes irreducible error in the context of machine learning?

- A. It represents the error that occurs due to the model being too simplistic.
- B. It refers to the error that occurs due to underfitting of the model.
- ☒ C. It is the error that remains even if a model perfectly captures the underlying patterns in the data.
- D. It is the error introduced by high variance in the model predictions.

Which of the following scenarios is a typical indication of overfitting in a machine learning model?

- ☒ A. The model performs well on the training data but poorly on the test data.
- B. The model consistently performs poorly on both the training and test data.
- C. The model has a high bias and low variance.
- D. The model is trained using a small dataset.

Which of the following best describes bias in the context of machine learning models?

- A. The error that occurs due to noise in the training data.
- ✓ B. The error that occurs when a model is too simplistic to capture the underlying relationships in the data.
- C. The error that occurs when a model fits too closely to the training data and fails to generalize to new data.
- D. The error that occurs due to variability in the training data.

In k-fold cross-validation, which of the following is a potential drawback when compared to a simple train-test split?

- ✓ A. It requires more computational resources.
- B. It may lead to overfitting due to multiple training iterations.
- C. It is more prone to bias in estimating model performance.
- D. It does not allow for hyperparameter tuning.

What kind of feedback does an agent receive in reinforcement learning to improve its actions?

- A. Labels for training data
- B. Rewards or penalties from the environment
- C. Clustering results
- D. Dimensionality reduction scores



What is the primary advantage of k-fold cross-validation compared to a simple train-test split?

- A. It guarantees that the model is tested on **all possible** combinations of training and test data.
- ✓ B. It provides a more accurate estimate of model performance by averaging over multiple train-test splits.
- C. It requires less computational resources and training time.
- D. It ensures that the model is trained on **all possible combinations** of training and validation data.

Which of the following statements best explains why variability in model performance occurs with different train/test splits?

- ✓ A. It is due to differences in the data used for testing the model.
- ✓ B. It is due to differences in the data used for training the model.
- ✓ C. Variations in the distribution of data between training and testing sets.
- D. It reflects changes in the model's hyperparameters.

In the context of machine learning, which of the following best describes the process of learning?

- A. The algorithm applies a fixed set of rules to predict outcomes.
- B. Memorizing the training data.
- ✓ C. The algorithm improves its performance on tasks based on experience with data.

Which of the following machine learning techniques is most suitable for finding hidden patterns or clusters in unlabeled data?

- A. Supervised learning
- B. Reinforcement learning.
- ✓ C. Unsupervised learning.
- D. Semi-supervised learning.

In machine learning, what does "generalization" refer to?

- A. The ability of the model to memorize the training data perfectly.
- ✓ B. The ability of the model to perform well on new, unseen data.
- C. The process of adjusting model parameters to fit the training data.
- D. The process of increasing model complexity to improve performance.

Which type of learning could use feedback from the environment to improve future performance?

- A. Supervised Learning
- B. Unsupervised Learning
- C. Semi-Supervised Learning
- ✓ D. Reinforcement Learning