

1. What is machine learning?

Machine learning is a subset of artificial intelligence that involves training algorithms on data to enable them to make predictions or decisions without being explicitly programmed for specific tasks.

2. List out any 3 applications of machine learning.

- Image Recognition: Identifying objects, faces, or features in images.
- Natural Language Processing (NLP): Language translation, sentiment analysis, and chatbots.
- Recommendation Systems: Suggesting products, movies, or content based on user preferences.

3. Describe overfitting and underfitting.

- Overfitting: A model learns the training data too well, capturing noise and details, resulting in poor generalization to new data.
- Underfitting: A model is too simple to capture the underlying pattern in the data, leading to poor performance on both training and new data.

4. Compare bias and variance.

- Bias: Error due to overly simplistic assumptions in the learning algorithm, leading to systematic inaccuracies.
- Variance: Error due to the model's sensitivity to small fluctuations in the training set, leading to overfitting and poor generalization.

5. Define SVM.

Support Vector Machine (SVM) is a supervised learning algorithm used for classification and regression tasks. It finds the optimal hyperplane that best separates the data into classes.

6. State the primary purpose of linear discriminant analysis.

Linear Discriminant Analysis (LDA) aims to find a linear combination of features that best separates two or more classes, maximizing the ratio of between-class variance to within-class variance.

7. List out ensemble methods.

- Bagging (Bootstrap Aggregating)
- Boosting (e.g., AdaBoost, Gradient Boosting)
- Random Forest

8. Define an unbalanced data set.

An unbalanced dataset has a disproportionate ratio of classes, leading to biased model predictions favoring the majority class over the minority class.

9. Explain Mean Squared Error.

Mean Squared Error (MSE) is a loss function that measures the average squared difference between the actual and predicted values, indicating the quality of a model's predictions.

10. Define AUC.

The Area Under the Curve (AUC) measures the ability of a classifier to distinguish between classes, representing the area under the Receiver Operating Characteristic (ROC) curve.

11. List the types of Regression.

- Linear Regression
- Logistic Regression
- Polynomial Regression

12. Name the applications of SVM.

- Text and Hypertext Categorization
- Image Classification
- Bioinformatics (e.g., gene classification)

13. Describe ensemble learning.

Ensemble learning combines multiple models to improve prediction accuracy and robustness compared to individual models. Techniques include bagging, boosting, and stacking.