Supervised Machine Learning

1. What is Machine Learning?
   A. A technique that enables systems to learn patterns from data
   B. A hardware optimization tool
   C. A database software
   D. A manual programming method
   Answer: A
2. Machine Learning mainly improves performance through:
   A. Hardware upgrades
   B. Experience from data
   C. Manual rule writing
   D. Increasing storage
   Answer: B
3. The role of statistics in ML is to:
   A. Remove algorithms
   B. Design hardware
   C. Provide mathematical foundation for inference
   D. Increase speed
   Answer: C
4. Probability in ML is primarily used to:
   A. Normalize datasets
   B. Remove noise
   C. Delete features
   D. Measure uncertainty
   Answer: D
5. Structured data is typically stored in:
   A. Tables with rows and columns
   B. Audio signals
   C. Images
   D. Text documents
   Answer: A
6. Unstructured data includes:
   A. CSV files
   B. Images and text documents
   C. Excel tables
   D. SQL databases
   Answer: B
7. Categorical data represents:
   A. Measured quantities
   B. Continuous values
   C. Categories or labels
   D. Percentages
   Answer: C
8. Numerical data can be:
   A. Text only
   B. Binary only
   C. Categorical only
   D. Continuous or discrete
   Answer: D

9. The first step in a Machine Learning pipeline is:
   A. Data collection
   B. Model testing
   C. Deployment
   D. Model evaluation
   Answer: A
10. Data preprocessing is important because it:
   A. Reduces model complexity
   B. Improves model performance
   C. Increases bias
   D. Removes labels
   Answer: B
11. Data splitting is done to:
   A. Remove noise
   B. Increase dataset size
   C. Validate model generalization
   D. Reduce features
   Answer: C
12. The training set is used to:
   A. Deploy the model
   B. Evaluate performance
   C. Remove outliers
   D. Learn patterns from data
   Answer: D
13. The test set is used to:
   A. Check performance on unseen data
   B. Train the model
   C. Normalize features
   D. Increase variance
   Answer: A
14. High bias typically results in:
   A. Overfitting
   B. Underfitting
   C. Perfect accuracy
   D. High variance
   Answer: B
15. High variance leads to:
   A. Stable model
   B. Underfitting
   C. Overfitting
   D. Low complexity
   Answer: C
16. Supervised learning requires:
   A. Unlabeled data
   B. No dataset
   C. Random variables
   D. Labeled data
   Answer: D

17. Regression is used when the output is:
    A. Continuous
    B. Categorical
    C. Binary only
    D. Text-based
    Answer: A
18. Classification is used when the output is:
    A. Continuous
    B. Categorical
    C. Numeric only
    D. Ratio scale
    Answer: B
19. Simple Linear Regression assumes:
    A. Non-linear relation
    B. Exponential relation
    C. Linear relationship
    D. Categorical relation
    Answer: C
20. A common regression evaluation metric is:
    A. Accuracy
    B. Precision
    C. Recall
    D. Mean Squared Error
    Answer: D
21. R-squared measures:
    A. Model fit quality
    B. Data normalization
    C. Bias
    D. Error count
    Answer: A
22. Logistic regression is used for:
    A. Continuous prediction
    B. Binary classification
    C. Clustering
    D. Dimensionality reduction
    Answer: B
23. Logistic regression uses:
    A. Linear equation only
    B. Polynomial equation
    C. Sigmoid function
    D. Square root function
    Answer: C
24. Accuracy in classification measures:
    A. Only true positives
    B. Only false negatives
    C. Model complexity
    D. Total correct predictions
    Answer: D

25. Precision measures:
    A. Correct positive predictions out of predicted positives
    B. Total samples
    C. True negatives only
    D. Variance
    Answer: A
26. Recall measures:
    A. Model bias
    B. Correct positives out of actual positives
    C. Total dataset size
    D. Feature importance
    Answer: B
27. A confusion matrix is used for:
    A. Regression
    B. Clustering
    C. Classification evaluation
    D. Data cleaning
    Answer: C
28. Decision Trees are mainly used in:
    A. Visualization
    B. Data storage
    C. Dimensionality reduction
    D. Supervised learning tasks
    Answer: D
29. Decision Trees split data based on:
    A. Feature importance
    B. Random choice
    C. Hardware speed
    D. Dataset size
    Answer: A
30. Overfitting in Decision Trees can be reduced by:
    A. Increasing depth
    B. Pruning
    C. Adding noise
    D. Removing data
    Answer: B
31. The ML pipeline ends with:
    A. Feature selection
    B. Data cleaning
    C. Deployment
    D. Data splitting
    Answer: C
32. If a model performs well on training data but poorly on test data, it indicates:
    A. Underfitting
    B. Low variance
    C. High bias
    D. Overfitting
    Answer: D

33. In supervised ML, features are also called:
    A. Independent variables
    B. Labels
    C. Outputs
    D. Targets
    Answer: A
34. Model evaluation ensures that:
    A. Model memorizes data
    B. Model generalizes well
    C. Data increases
    D. Bias increases
    Answer: B
35. The target variable in regression is:
    A. Categorical
    B. Text-based
    C. Continuous
    D. Binary only
    Answer: C
36. Bias-variance tradeoff helps to:
    A. Remove data
    B. Normalize features
    C. Increase complexity
    D. Balance underfitting and overfitting
    Answer: D
37. Probability helps ML models to:
    A. Estimate likelihood of outcomes
    B. Delete features
    C. Normalize datasets
    D. Increase bias
    Answer: A
38. Model training involves:
    A. Deployment
    B. Learning parameters from data
    C. Removing labels
    D. Data cleaning
    Answer: B
39. Supervised learning algorithms include:
    A. Clustering
    B. Dimensionality reduction
    C. Regression and Classification
    D. Association rules
    Answer: C
40. The main objective of supervised learning is to:
    A. Discover hidden groups
    B. Reduce dimensions
    C. Visualize trends
    D. Predict outputs from labeled data
    Answer: D