

Machine Learning Test

❖ Read Carefully Multiple Options could be correct

1. In the context of machine learning, the ability to extract patterns from data relies on distinct learning paradigms. Instance-based learning and model-based learning represent two primary approaches. Which of the following statements accurately describe how these approaches contribute to pattern extraction from data?

A. Instance-based learning memorizes the training instances and uses a similarity measure to make predictions for new data, without explicitly identifying underlying patterns.
B. Model-based learning involves the creation of a global model that abstracts patterns from the training data by generalizing based on statistical or mathematical principles.
C. Instance-based learning is more computationally efficient during prediction time compared to model-based learning, as it does not need to store all training instances for future use.
D. Model-based learning optimizes a loss function during training, enabling the model to generalize well to unseen data, while instance-based learning struggles with high-dimensional data due to overfitting.

2. In linear regression, there are multiple techniques to compute coefficients and the intercept. Which of the following methods are used, and which one is applied by Scikit-learn's LinearRegression model?

A. Ordinary Least Squares (OLS) minimizes the sum of squared errors between predicted and actual values.
B. Gradient Descent iteratively adjusts coefficients to minimize the cost function, often used for large datasets.
C. Maximum Likelihood Estimation (MLE) estimates parameters by maximizing the likelihood function.
D. Scikit-learn's LinearRegression uses Stochastic Gradient Descent internally, optimizing coefficients for small datasets.

3. Which of the following is NOT an assumption of linear regression?

A. Linearity of the relationship between independent and dependent variables.
B. Homoscedasticity (constant variance of errors).
C. Independence of errors.
D. Multicollinearity among independent variables should be high.

4. Which of the following does NOT satisfy the condition for the difference between parametric and non-parametric tests?

A. Parametric tests assume a specific distribution of the data, usually normal.
B. Non-parametric tests do not assume any specific distribution and are more flexible.
C. Parametric tests are typically used for small sample sizes, while non-parametric tests are better for large samples.
D. Non-parametric tests are more robust in the presence of outliers compared to parametric tests.

5. Which of the following statements accurately describes how to select the appropriate metric and the consequences of choosing the wrong one?

A. The metric depends on the problem type (e.g., classification or regression) and the specific goal (e.g., accuracy, precision, recall).
B. Mischoosing the metric can lead to overfitting the model to training data.
C. Accuracy is always the best metric, even for imbalanced classification problems.
D. Using the wrong metric can result in poor real-world performance, as it may not align with the business goal.

6. In machine learning, gradient descent comes in different variants: stochastic, mini-batch, and batch gradient descent. Which of the following statements accurately describes their differences, advantages, and disadvantages?

A. Stochastic Gradient Descent (SGD) updates weights after every training example, which can introduce noise but helps escape local minima.
B. Batch Gradient Descent uses the entire dataset to compute gradients, leading to more stable updates but higher computational cost per iteration.
C. Mini-batch Gradient Descent strikes a balance between SGD and batch gradient descent by using small batches,

combining the benefits of both faster convergence and reduced computational cost.

D. Stochastic Gradient Descent is computationally faster than mini-batch gradient descent, but it converges more smoothly because it avoids oscillations.

7. Which of the following correctly describes the differences between R^2 and Adjusted R^2 in linear regression?

A. R^2 shows the variance explained by the model but increases even with irrelevant features.

B. Adjusted R^2 accounts for the number of predictors, penalizing unnecessary features.

C. R^2 is always better than Adjusted R^2 , especially with many predictors.

D. Adjusted R^2 is better for models with many predictors as it reduces the risk of overfitting.

8. Which of the following correctly explains the difference between feature selection and feature engineering in machine learning?

A. Feature selection removes irrelevant or redundant features to improve model performance.

B. Feature engineering creates new features or transforms existing ones to capture more information.

C. Feature selection focuses on enhancing feature representation, while feature engineering removes irrelevant features.

D. Both feature selection and feature engineering aim to reduce dimensionality, but feature engineering creates new information.

9. Is multicollinearity always bad in machine learning, and how can it be detected?

A. Multicollinearity is not always bad; it mainly affects interpretability rather than model performance in predictive tasks.

B. High multicollinearity inflates variance of coefficients, making them unreliable for interpretation.

C. It can be detected using Variance Inflation Factor (VIF), where a VIF value above 10 often indicates serious multicollinearity.

D. The p-values of predictors can also reveal multicollinearity if they are low but the R^2 value is high.

10. Which of the following correctly identifies the disadvantages of the filter and wrapper methods in feature selection?

A. The filter method selects features based only on statistical metrics, ignoring interactions between features, which can lead to suboptimal feature subsets.

B. The wrapper method is computationally expensive, especially with large datasets, as it evaluates multiple models for different feature subsets.

C. The filter method is slower compared to the wrapper method because it evaluates each subset using cross-validation.

D. The wrapper method risks overfitting because it tightly optimizes features for a specific model.

11. Which of the following techniques are used to prevent overfitting in machine learning, and what are the key differences between L1 and L2 regularization?

A. L1 regularization (Lasso) adds an absolute value penalty, shrinking some coefficients to zero, leading to sparse models.

B. L2 regularization (Ridge) penalizes large coefficients without setting them to zero, helping reduce multicollinearity.

C. Dropout randomly removes neurons during training to prevent overfitting in neural networks.

D. L2 regularization tends to produce sparser models than L1 regularization, making it more suitable for feature selection.

12. Which of the following statements correctly explains the bias-variance tradeoff in model evaluation and its impact on machine learning model performance?

A. Bias refers to the error due to overly simplistic assumptions in the learning algorithm, leading to underfitting.

B. Variance refers to the model's sensitivity to fluctuations in the training data, which can lead to overfitting.

C. A model with high bias performs well on training data but poorly on unseen data due to overfitting.

D. The tradeoff means that reducing bias typically increases variance, and vice versa, affecting overall model performance.

13. Which of the following statements correctly describes feature extraction in machine learning?

- A.** Feature extraction transforms original features into a new feature space to capture essential information.
- B.** Feature extraction always retains all original features without modification.
- C.** Techniques like Principal Component Analysis (PCA) are commonly used for feature extraction.
- D.** Feature extraction can lead to dimensionality reduction by creating fewer, more informative features.

14. How can data leakage occur during hyperparameter tuning, and what are its consequences?

- A.** Data leakage can occur if hyperparameters are tuned on the entire dataset before splitting into training and validation sets.
- B.** Data leakage results in overly optimistic model performance during tuning but poor performance on unseen data.
- C.** Using cross-validation during hyperparameter tuning prevents data leakage by ensuring that validation data remains unseen during training.
- D.** Data leakage does not affect hyperparameter tuning as it only impacts the training phase of the model.

15. What is the main purpose of the Durbin-Watson test in regression analysis?

- A.** To check for autocorrelation in the residuals of the model.
- B.** To verify if the model follows the linearity assumption.
- C.** To detect multicollinearity among the predictors.
- D.** To check if the residuals follow a normal distribution.

16. Why might Mean Absolute Error (MAE) not be the best choice as a loss function for some machine learning models?

- A.** MAE is not differentiable at all points, which makes it harder for optimization algorithms like gradient descent to work efficiently.
- B.** MAE gives equal weight to all errors, which may not be ideal when larger errors need to be penalized more heavily.
- C.** MAE can lead to multiple optimal solutions in models with more complex feature spaces, making it unstable.
- D.** MAE performs poorly when the data contains significant outliers, as it heavily penalizes them.

17. How do homoscedasticity and heteroscedasticity impact the performance of regression models in machine learning?

- A.** Homoscedasticity refers to constant variance in the residuals, which is an assumption of linear regression models for reliable predictions.
- B.** Heteroscedasticity refers to non-constant variance in the residuals, which can lead to biased and inefficient parameter estimates in regression models.
- C.** In cases of heteroscedasticity, models like Ordinary Least Squares (OLS) regression can still perform well without any adjustments.
- D.** Homoscedasticity violations indicate that the model may not have captured the true relationship between the variables.

18. What is hyperparameter tuning in machine learning, and what are some common strategies for effectively searching for optimal hyperparameters?

- A.** Hyperparameter tuning involves finding the best set of hyperparameters that control the model training process but are not learned from the data.
- B.** Grid search explores a predefined set of hyperparameter values by evaluating all possible combinations.
- C.** Random search randomly samples hyperparameter combinations and can often find good solutions faster than grid search.
- D.** Bayesian optimization selects hyperparameters based on prior knowledge and performance of previous trials to improve search efficiency.

19. Which of the following correctly describes the loss function in logistic regression?

- A.** Logistic regression uses log loss (binary cross-entropy) to measure error in predicted probabilities.
- B.** Log loss penalizes wrong predictions more when probabilities are far from actual values.

- C.** It optimizes for maximum likelihood instead of using mean squared error like in linear regression.
- D. Logistic regression uses hinge loss for optimization.

20. How does logistic regression manage multiclass classification?

- A.** One-vs-Rest (OvR): Trains a binary classifier for each class against all others.
- B. One-vs-One (OvO): Trains classifiers for every pair of classes.
- C.** Softmax Regression: Directly predicts probabilities for all classes in one model.
- D. Logistic regression can only handle binary classification.

21. Which of the following statements are not assumptions of logistic regression?

- A. The dependent variable must be binary (0 or 1).
- B. The relationship between the independent variables and the log-odds of the dependent variable is linear.
- C. There should be no multicollinearity among the independent variables.
- D.** The residuals should be normally distributed.

22. What is the time complexity of the K-Nearest Neighbors (KNN) algorithm?

- A.** Training Phase: $O(n)$ for storing training samples.
- B.** Prediction Phase: $O(n \cdot d)$ for calculating distances to all training samples.
- C. Brute-force approach for distance calculation: $O(n \log n)$.
- D. Using KD-trees or Ball Trees can reduce complexity to $O(\log n)$.

23. What is the importance of the kernel function in Support Vector Machines (SVM)?

- A.** The kernel function transforms input data into a higher-dimensional space.
- B. It reduces the computational cost of training the SVM model.
- C.** Different kernels (linear, polynomial, RBF) help adapt to various data shapes.
- D. The kernel function is used for data normalization before training.

24. Why are decision trees often considered overfitted models?

- A.** Decision trees can create very complex structures that fit the training data perfectly.
- B.** They are sensitive to noise in the training data, which can lead to capturing irrelevant patterns.
- C. They require large amounts of data to generalize well, which is often not available.
- D. Decision trees typically have high bias, making them less flexible.

25. What are ensemble methods, and how do they improve model performance? Discuss the differences between bagging and boosting.

- A.** Ensemble methods combine multiple models to improve overall performance and reduce overfitting.
- B.** Bagging builds independent models in parallel and combines their predictions by averaging or voting.
- C.** Boosting builds models sequentially, where each model corrects the errors of the previous ones.
- D.** Bagging usually reduces variance, while boosting reduces bias and can increase variance.

26. In a decision tree for classification, we use Gini impurity. What should we use for regression tasks?

- A.** Mean Squared Error (MSE)
- B. Gini Impurity
- C.** Mean Absolute Error (MAE)
- D. Entropy

27. What is the role of Out-of-Bag (OOB) error in bagging?

- A.** OOB error estimates model performance in bagging (e.g., Random Forest).
- B.** OOB samples are unused data points that act as a test set for each tree.
- C.** OOB error gives an unbiased estimate without needing a validation set.
- D. OOB error is only used in boosting methods.

28. Why is feature scaling applied in machine learning?

- ☒ A. Ensures all features contribute equally.
- ☒ B. Helps gradient descent converge faster.
- ☒ C. Prevents larger features from dominating.
- ☒ D. Not needed for tree-based models.

29. Which of the following methods won't effectively address an imbalanced dataset?

- A. SMOTE increases the minority class samples by synthesizing new instances.
- B. Class weighting adjusts the model to penalize misclassifying minority class examples.
- C. Undersampling reduces the number of majority class samples to balance the dataset.
- ☒ D. Relying on the accuracy metric for evaluation is an effective approach.

30. What is the primary goal of unsupervised machine learning techniques?

- A. To predict outcomes based on labeled data.
- ☒ B. To find hidden patterns or groupings in data without prior labels.
- C. To minimize error in classification tasks.
- D. To classify data into predefined categories.

31. Which of the following is an example of using unsupervised learning?

- ☒ A. Grouping news articles based on their topics without knowing their categories beforehand.
- B. Predicting the price of a house based on size and location.
- C. Classifying emails as spam or not spam using labeled examples.
- D. Diagnosing diseases by learning from past medical records with outcomes.

32. Which of the following is an example of supervised learning?

- ☒ A. Predicting the price of a house using data about its size and location.
- B. Grouping images of animals without knowing their categories in advance.
- C. Finding hidden patterns in customer purchase behavior without labels.
- D. Clustering social media posts based on content similarity.

33. Do regression and classification tasks come under supervised or unsupervised learning?

- ☒ A. Supervised learning
- B. Unsupervised learning
- C. Both
- D. Neither

34. Which of the following describes reducible and irreducible errors in machine learning?

- ☒ A. Bias and variance are part of the reducible error that can be minimized with better models.
- ☒ B. Irreducible error comes from noise in the data and cannot be reduced by the model.
- C. Reducible error is caused only by variance in the model.
- D. Irreducible error can be minimized with better data preprocessing.

35. Which of the following is a misconception about how PCA reduces dimensionality?

- A. PCA finds linear combinations of the original features to form new axes, maximizing variance along each axis.
- B. The first few principal components capture the most variance, allowing dimensionality reduction by retaining only those.
- ☒ C. PCA reduces dimensions by identifying and removing features that are not correlated with the target variable.
- D. PCA generates uncorrelated components (orthogonal) that capture different aspects of the data's variance.

36. Which of the following is a common misconception about kurtosis in statistics?

- A. High kurtosis means the data has more extreme outliers than a normal distribution.
- B. Kurtosis measures the tailedness of a distribution.
- C. Low kurtosis indicates that the data is light-tailed, with fewer extreme values than normal.
- ☒ D. Kurtosis directly measures the peak or flatness of the distribution.

Case Study: Medical Diagnosis for a Rare Disease

A healthcare organization is developing a machine learning model to diagnose a rare disease based on patient data. The disease affects only 1% of the population. The organization is particularly concerned about accurately identifying patients who have the disease to ensure they receive the necessary treatment while minimizing false positives.

37. In evaluating the model's performance, which metric would be the most valid for this scenario?

- A. Accuracy
- B. Precision
- ☒ C. Recall
- D. F1 Score

38. The model reports a precision of 0.85 and a recall of 0.70. What does this imply about the model's performance?

- ☒ A. The model is good at identifying true positive cases but may have many false negatives.
- B. The model is equally good at identifying true positives and true negatives.
- C. The model has a balanced performance with no trade-offs between precision and recall.
- D. The model has a low false positive rate and a high false negative rate.

39. If the organization wants to balance precision and recall, which metric should they consider using?

- A. Accuracy
- B. Specificity
- ☒ C. F1 Score
- D. ROC-AUC

40. Given the same model's performance, with precision at 0.85 and recall at 0.70, what is the F1 Score?

- A. 0.80
- B. 0.75
- ☒ C. 0.78
- D. 0.82