Machine Learning

100 Answers with Explanations

Based on First 11 Pages

Nova IMS
Generated: January 16, 2026

LEVEL 1: BASIC ML CONCEPTS

Q1. What is machine learning?

ANSWER: B

Explanation: Machine learning enables computers to learn patterns and make decisions from data without being explicitly programmed for every scenario. This distinguishes it from traditional programming where developers must write specific rules for each situation.

Q2. What are 'features' in machine learning?

ANSWER: B

Explanation: Supervised learning is the only learning paradigm that uses labeled training data, where each input has a corresponding known output. This allows the algorithm to learn the mapping between inputs and outputs.

Q3. What are 'labels' in supervised learning?

ANSWER: B

Explanation: Features are the input variables or attributes that the model uses to make predictions. They represent the characteristics of the data that help the algorithm learn patterns.

Q4. What is a 'model' in machine learning?

ANSWER: C

Explanation: Unsupervised learning discovers hidden patterns and structures in data without labeled outputs. The algorithm identifies relationships, groupings, or representations on its own.

Q5. What is 'inference' in machine learning?

ANSWER: B

Explanation: Spam detection is a classification task because it assigns emails to discrete categories (spam or not spam). The other options predict continuous numerical values, making them regression tasks.

Q6. Which type of learning uses labeled training data with input-output pairs?

ANSWER: C

Explanation: Regression predicts continuous numerical values (e.g., prices, temperatures), while classification predicts discrete categories or classes (e.g., spam/not spam, species types).

Q7. What is the primary goal of supervised learning?

ANSWER: C

Explanation: Reinforcement learning agents learn by receiving rewards for good actions and penalties for bad ones, aiming to maximize cumulative rewards over time through trial and error.

Q8. Which of the following is a classification task?

ANSWER: B

Explanation: A label is the target output or correct answer associated with each training example in supervised learning. It's what the model learns to predict.

Q9. Which of the following is a regression task?

ANSWER: C

Explanation: Clustering is an unsupervised learning technique that groups similar data points together based on their characteristics, without predefined categories.

Q10. What distinguishes regression from classification?

ANSWER: A

Explanation: Data preprocessing (cleaning, transformation, handling missing values, etc.) typically consumes 60-80% of ML workflow time because real-world data is messy and requires extensive preparation.

Q11. What are common classification algorithms? (Select the correct statement)

ANSWER: B

Explanation: This principle emphasizes that model quality is fundamentally limited by data quality. Poor quality input data will inevitably produce poor quality results, regardless of algorithm sophistication.

Q12. What are common regression algorithms? (Select the correct statement)

ANSWER: B

Explanation: The test set provides an unbiased evaluation of how the model will perform on new, unseen data. This estimates real-world performance after deployment.

Q13. What is the primary goal of unsupervised learning?

ANSWER: C

Explanation: An 80/20 split (80% training, 20% testing) is the most common ratio, providing sufficient data for both training and reliable evaluation.

Q14. Which of the following is an unsupervised learning task?

ANSWER: B

Explanation: MCAR (Missing Completely at Random) means the missingness has no relationship with any variables - it's purely random, like data lost due to a computer crash.

Q15. What is clustering in unsupervised learning?

ANSWER: B

Explanation: Deletion (listwise deletion) simply removes all rows containing missing values. While wasteful, it's the simplest approach requiring no imputation logic.

LEVEL 2: ML TRIBES AND MATHEMATICAL FOUNDATIONS

Q16. What are common clustering algorithms?

ANSWER: B

Explanation: Standardization transforms each feature to have a mean of 0 and standard deviation of 1 using the formula: (x - μ) / σ.

Q17. What is dimensionality reduction?

ANSWER: B

Explanation: Mean imputation replaces missing values with the average (mean) of the available values for that feature.

Q18. What are dimensionality reduction techniques mentioned in the curriculum?

ANSWER: B

Explanation: The IQR (Interquartile Range) method is specifically designed for outlier detection by identifying values that fall beyond the typical range of the data.

Q19. In reinforcement learning, what is an 'agent'?

ANSWER: B

Explanation: The standard IQR rule defines outliers as values below Q1 - 1.5×IQR or above Q3 + 1.5×IQR, where Q1 and Q3 are the 25th and 75th percentiles.

Q20. In reinforcement learning, what is an 'environment'?

ANSWER: B

Explanation: Min-Max scaling transforms features to a [0, 1] range using the formula: (x - min) / (max - min).

Q21. What does a reinforcement learning agent learn to maximize?

ANSWER: C

Explanation: One-hot encoding creates a binary column for each category without implying any order, making it ideal for nominal (non-ordered) categorical variables with few categories.

Q22. What is a 'reward' in reinforcement learning?

ANSWER: B

Explanation: With high-cardinality features (many unique categories), one-hot encoding creates numerous new columns, dramatically increasing dimensionality and potentially causing the 'curse of dimensionality'.

Q23. What is a 'policy' in reinforcement learning?

ANSWER: B

Explanation: Stratified sampling maintains the class distribution in both training and test sets, which is essential for imbalanced datasets to ensure the minority class is adequately represented in both sets.

Q24. What are applications of reinforcement learning?

ANSWER: B

Explanation: K represents the number of folds (equal-sized subsets) that the data is divided into. The algorithm trains on K-1 folds and validates on the remaining fold, repeating K times.

Q25. What are common reinforcement learning algorithms mentioned?

ANSWER: B

Explanation: K=5 or K=10 are most common, balancing computational cost with reliable performance estimates. Larger K values give more reliable estimates but increase computation.

Q26. According to Pedro Domingos, what are the five "Machine Learning Tribes"?

ANSWER: B

Explanation: Symbolists use symbolic logic, rules, and inverse deduction. They believe intelligence can be achieved through formal logic and reasoning systems, exemplified by decision trees and rule-based expert systems.

Q27. Which ML tribe uses inverse deduction and logic rules?

ANSWER: C

Explanation: Symbolists focus on logic-based learning, using symbolic representations and rules derived from formal logic systems.

Q28. What algorithm family do Symbolists primarily use?

ANSWER: B

Explanation: Symbolists primarily use algorithms based on logic and rules, such as decision trees, rule induction systems, and knowledge-based systems.

Q29. Which ML tribe attempts to reverse-engineer the brain?

ANSWER: B

Explanation: Connectionists believe intelligence emerges from networks of simple units (like neurons in the brain), leading to the development of neural networks and deep learning.

Q30. What algorithms do Connectionists use?

ANSWER: B

Explanation: Connectionists use neural network architectures including deep learning, CNNs for images, RNNs for sequences, and transformers for language processing.

Q31. What is a strength of Connectionist (neural) approaches?

ANSWER: B

Explanation: Connectionists' neural networks excel at handling unstructured data (images, text, audio) and can learn complex hierarchical representations automatically through multiple layers.

Q32. Which ML tribe simulates evolution through selection and mutation?

ANSWER: C

Explanation: Evolutionaries apply principles of biological evolution - selection, mutation, crossover - to evolve solutions through successive generations.

Q33. What algorithms do Evolutionaries use?

ANSWER: B

Explanation: Evolutionaries use genetic algorithms, genetic programming, and evolutionary strategies that simulate natural selection to find optimal solutions.

Q34. Which ML tribe uses probabilistic inference and Bayes' theorem?

ANSWER: D

Explanation: Bayesians use probabilistic inference based on Bayes' theorem, updating beliefs as new evidence becomes available.

Q35. What algorithms do Bayesians use?

ANSWER: B

Explanation: Bayesian algorithms include Naive Bayes classifiers, Bayesian networks (graphical models), and Gaussian processes for regression.

Q36. What is a strength of Bayesian approaches?

ANSWER: B

Explanation: Bayesian approaches naturally quantify uncertainty through probability distributions and can incorporate prior knowledge into the learning process.

Q37. Which ML tribe learns by recognizing similarities?

ANSWER: D

Explanation: Analogizers learn by recognizing similarities between examples, using the principle that similar inputs should produce similar outputs.

Q38. What algorithms do Analogizers use?

ANSWER: B

Explanation: Analogizers use similarity-based methods including K-nearest neighbors, support vector machines with kernel functions, and case-based reasoning.

Q39. Which mathematical field is essential for understanding vectors and matrices in ML?

ANSWER: B

Explanation: Linear algebra provides the mathematical foundation for representing and manipulating vectors (data points) and matrices (datasets, weights) that are fundamental to ML algorithms.

Q40. Which mathematical field provides derivatives and gradients for ML optimization?

ANSWER: B

Explanation: Calculus provides derivatives and gradients essential for optimization algorithms like gradient descent that train ML models by minimizing loss functions.

Q41. Which mathematical field provides concepts like distributions and Bayes' theorem?

ANSWER: C

Explanation: Probability theory provides concepts like probability distributions, conditional probability, and Bayes' theorem that underpin many ML algorithms.

Q42. Which statistical concepts are important for ML?

ANSWER: B

Explanation: Statistical concepts including hypothesis testing, confidence intervals, mean, variance, and standard deviation are essential for understanding model performance and making inferences.

Q43. What is a common misconception about machine learning?

ANSWER: B

Explanation: A common misconception is that ML is a magical solution that doesn't require understanding or domain knowledge. In reality, ML requires careful problem formulation, data preparation, and interpretation.

Q44. Is the statement "more complex models always perform better" correct?

ANSWER: B

Explanation: Simpler models often outperform complex ones due to better generalization, especially with limited data. Complex models risk overfitting - memorizing training data rather than learning general patterns.

Q45. Can machine learning replace domain expertise entirely?

ANSWER: B

Explanation: Domain expertise remains essential for feature engineering, interpreting results, identifying data issues, and ensuring models align with real-world constraints and business objectives.

LEVEL 3: DATA PREPROCESSING FUNDAMENTALS

Q46. What percentage of ML workflow time is typically spent on data preprocessing?

ANSWER: C

Explanation: Studies consistently show that 60-80% of ML project time is spent on data preprocessing because real-world data contains numerous quality issues requiring extensive cleaning and transformation.

Q47. What does the principle "garbage in, garbage out" emphasize?

ANSWER: B

Explanation: This fundamental principle emphasizes that no amount of sophisticated modeling can compensate for poor quality data. Quality outputs require quality inputs.

Q48. What is MCAR in missing data terminology?

ANSWER: B

Explanation: MCAR (Missing Completely at Random) means missingness has no relationship with any observed or unobserved variables - it's purely random.

Q49. What does MCAR (Missing Completely at Random) mean?

ANSWER: B

Explanation: In MCAR, whether a value is missing is completely independent of all variables. For example, data lost due to a random equipment failure.

Q50. What is MAR in missing data terminology?

ANSWER: A

Explanation: MAR (Missing at Random) means missingness is related to observed data but not to the missing values themselves.

Q51. What does MAR (Missing at Random) mean?

ANSWER: B

Explanation: In MAR, missingness depends on observed variables. For example, men might be less likely to report their weight, but among those who report it, values are random.

Q52. What is MNAR in missing data terminology?

ANSWER: A

Explanation: MNAR (Missing Not at Random) means the probability of missingness is related to the actual missing value itself.

Q53. What does MNAR (Missing Not at Random) mean?

ANSWER: C

Explanation: In MNAR, missingness is related to the unobserved value. For example, people with very high income might be less likely to report it.

Q54. What is the simplest method to handle missing data?

ANSWER: C

Explanation: Listwise deletion removes entire rows containing any missing values. It's the simplest approach but can waste data if missingness is scattered across features.

Q55. What is listwise deletion?

ANSWER: B

Explanation: Listwise deletion means removing any row that contains at least one missing value from the dataset.

Q56. What is a disadvantage of listwise deletion?

ANSWER: B

Explanation: Listwise deletion is wasteful because it discards entire rows even when only one feature is missing, potentially removing large amounts of useful information.

Q57. What is mean imputation?

ANSWER: B

Explanation: Mean imputation replaces missing values with the arithmetic mean of the observed values for that feature.

Q58. What is median imputation typically used for?

ANSWER: B

Explanation: Median imputation is preferred when the distribution has outliers because the median is robust to extreme values, unlike the mean.

Q59. What is mode imputation used for?

ANSWER: B

Explanation: Mode imputation uses the most frequently occurring value, which is appropriate for categorical variables where mean/median don't make sense.

Q60. What is forward fill imputation?

ANSWER: B

Explanation: Forward fill uses the most recent previous value to fill missing values, which is logical for time series where values tend to persist.

Q61. What is KNN imputation?

ANSWER: B

Explanation: KNN imputation finds the K most similar instances (nearest neighbors) and uses a weighted average of their values to fill the missing value.

Q62. What is multiple imputation?

ANSWER: B

Explanation: Multiple imputation creates several different plausible imputed datasets, analyzes each separately, and combines results to account for imputation uncertainty.

Q63. What is the Z-score method used for?

ANSWER: B

Explanation: The Z-score method identifies outliers by calculating how many standard deviations a value is from the mean, with values beyond ±3σ typically considered outliers.

Q64. In the Z-score method, what typically indicates an outlier?

ANSWER: C

Explanation: The conventional threshold is ±3 standard deviations (±3σ), which captures approximately 99.7% of data in a normal distribution, making values beyond this likely outliers.

Q65. What does IQR stand for?

ANSWER: B

Explanation: IQR stands for Interquartile Range, which is the difference between the 75th percentile (Q3) and 25th percentile (Q1).

Q66. In the IQR method, outliers are typically defined as:

ANSWER: C

Explanation: The IQR method defines outliers as values below Q1 - 1.5×IQR or above Q3 + 1.5×IQR, a robust rule that works well regardless of distribution shape.

Q67. What is Isolation Forest?

ANSWER: B

Explanation: Isolation Forest is an ensemble method that isolates outliers by randomly partitioning the data. Outliers require fewer splits to isolate, making them easier to detect.

Q68. What is LOF (Local Outlier Factor)?

ANSWER: B

Explanation: LOF (Local Outlier Factor) compares the local density of a point to the density of its neighbors. Points with significantly lower density are considered outliers.

Q69. What is DBSCAN used for?

ANSWER: B

Explanation: DBSCAN (Density-Based Spatial Clustering) identifies clusters of high density and marks points in low-density regions as outliers.

Q70. What are common strategies for treating outliers?

ANSWER: B

Explanation: The appropriate strategy depends on context: remove if measurement errors, transform (log) to reduce impact, cap (winsorize) to limit influence, or keep if genuine extreme cases.

LEVEL 4: FEATURE SCALING

Q71. Why is feature scaling important?

ANSWER: B

Explanation: Many algorithms use distance-based calculations or gradient descent optimization that are sensitive to feature scales. Without scaling, features with larger ranges dominate.

Q72. Which algorithms require feature scaling?

ANSWER: B

Explanation: Distance-based algorithms (KNN, SVM, K-means), gradient descent (neural networks, logistic regression), and PCA all require feature scaling because they use distance metrics or are sensitive to feature magnitudes.

Q73. Which algorithms do NOT require feature scaling?

ANSWER: C

Explanation: Tree-based models split data based on feature values, not distances. Since splits are based on thresholds, they're invariant to feature scaling.

Q74. What is standardization (Z-score normalization)?

ANSWER: B

Explanation: Standardization (Z-score normalization) transforms each feature to have zero mean and unit variance using the formula $(x - \mu) / \sigma$.

Q75. What is the formula for standardization?

ANSWER: B

Explanation: The standardization formula is $(x - \mu) / \sigma$, where μ is the mean and σ is the standard deviation of the feature.

Q76. What does standardization transform data to?

ANSWER: B

Explanation: After standardization, each feature has a mean of 0 and standard deviation of 1, centering the distribution around zero.

Q77. Does standardization preserve outlier information?

ANSWER: B

Explanation: Yes, standardization preserves the relative relationships between values, including outliers. It doesn't remove outliers but transforms them proportionally.

Q78. What is Min-Max normalization (scaling)?

ANSWER: B

Explanation: Min-Max scaling formula is (x - min) / (max - min), which linearly transforms values to the [0, 1] range.

Q79. What range does Min-Max scaling transform data to?

ANSWER: B

Explanation: Min-Max scaling transforms all values to the range [0, 1], where the minimum becomes 0 and the maximum becomes 1.

Q80. Is Min-Max scaling sensitive to outliers?

ANSWER: B

Explanation: Yes, Min-Max scaling is sensitive to outliers because extreme values determine the min and max, causing all other values to be compressed into a narrow range.

Q81. What is Robust Scaling?

ANSWER: B

Explanation: Robust Scaling uses the median and IQR instead of mean and standard deviation, making it robust to outliers since these statistics aren't affected by extreme values.

Q82. What is the formula for Robust Scaling?

ANSWER: B

Explanation: Robust Scaling formula is (x - median) / IQR, using statistics that are resistant to outliers.

Q83. When should you use Robust Scaling?

ANSWER: B

Explanation: Use Robust Scaling when your data contains many outliers that you want to preserve but not let dominate the scaling process.

Q84. What is MaxAbs Scaling?

ANSWER: B

Explanation: MaxAbs Scaling divides each value by the maximum absolute value: x / |max|, scaling to the range [-1, 1] while preserving zero entries.

Q85. What is an advantage of MaxAbs Scaling?

ANSWER: B

Explanation: MaxAbs Scaling doesn't shift/center the data, so sparse matrices (with many zeros) remain sparse, which is important for memory efficiency with high-dimensional data.

LEVEL 5: CRITICAL DATA PREPROCESSING CONCEPTS

Q86. When should you fit a scaler?

ANSWER: B

Explanation: You must fit the scaler only on training data, then use those same parameters to transform both training and test sets. This prevents information leakage from test data.

Q87. Why is fitting a scaler on the entire dataset (including test) a problem?

ANSWER: B

Explanation: Fitting on the entire dataset causes data leakage because test set statistics influence the transformation applied to training data, leading to overly optimistic performance estimates.

Q88. What should you do with the scaler parameters after fitting on training data?

ANSWER: B

Explanation: Store the scaler parameters (mean, std, min, max, etc.) after fitting on training data. Use these same parameters for all future data transformations, including test and production data.

Q89. Why must tree-based models NOT require feature scaling?

ANSWER: B

Explanation: Tree-based models make splits based on feature thresholds, which are invariant to monotonic transformations like scaling. The relative ordering of values doesn't change with scaling.

Q90. What happens if you forget to scale features for KNN?

ANSWER: B

Explanation: Without scaling, features with larger numerical ranges (e.g., income in thousands) will dominate distance calculations over features with smaller ranges (e.g., age), creating biased predictions.

Q91. What happens if you scale features for a Random Forest?

ANSWER: B

Explanation: Scaling provides no benefit for Random Forest and wastes computation time. The model's performance is unchanged because splits are based on thresholds, not distances.

Q92. For neural networks, which scaling is typically recommended?

ANSWER: B

Explanation: Neural networks use gradient descent optimization, which works best when all features are on similar scales. Standardization or Min-Max scaling are both commonly used.

Q93. What is a critical rule about train/test splitting and preprocessing?

ANSWER: B

Explanation: Split data first into train/test, then fit all preprocessing (including scalers) only on the training set. This prevents test information from leaking into the training process.

Q94. Why does data preprocessing consume 60-80% of ML workflow time?

ANSWER: B

Explanation: Real-world data contains missing values, outliers, inconsistencies, wrong formats, duplicates, and irrelevant information. Extensive cleaning and transformation is required before modeling.

Q95. Raw data typically contains which issues?

ANSWER: B

Explanation: Raw data typically contains noise (random errors), missing values, inconsistencies (formatting issues), outliers, and irrelevant or redundant features that must be addressed.

Q96. What degrades model accuracy the most?

ANSWER: B

Explanation: Poor data quality (noisy, missing, incorrect, biased) has a far greater negative impact than algorithm choice. Clean, relevant data is more important than sophisticated algorithms.

Q97. What is the relationship between data quality and model performance?

ANSWER: B

Explanation: Better data quality enables models to learn true patterns rather than noise, leading to better generalization. The relationship is direct and strong.

Q98. Can sophisticated algorithms compensate for poor data quality?

ANSWER: B

Explanation: No. The 'garbage in, garbage out' principle applies - sophisticated algorithms cannot overcome fundamental data quality problems. They'll just learn the garbage more effectively.

Q99. Why is understanding missing data mechanisms (MCAR, MAR, MNAR) important?

ANSWER: B

Explanation: Different missing data mechanisms (MCAR, MAR, MNAR) require different handling approaches. Understanding the mechanism helps choose appropriate imputation or deletion strategies.

Q100. What should guide your choice of outlier treatment strategy?

ANSWER: B

Explanation: Outlier treatment should be guided by domain knowledge. True errors should be removed, but genuine extreme values (like legitimate high incomes) should often be kept or transformed, not automatically deleted.