

Machine Learning

100 Multiple Choice Questions

Based on First 11 Pages of Curriculum Analysis

(LU1: ML Fundamentals + LU2: Data Preprocessing - Partial)

Nova IMS - Master's Program
Generated: January 15, 2026

MACHINE LEARNING - 100 MULTIPLE CHOICE QUESTIONS

Based on First 11 Pages of Curriculum Analysis

(LU1: ML Fundamentals + LU2: Data Preprocessing - Partial)

LEVEL 1: BASIC ML CONCEPTS (Questions 1-25)

Q1. What is machine learning?

- A) Programming computers with explicit rules for every scenario
- B) Algorithms that improve automatically through experience and data
- C) A database management system
- D) A statistical analysis software

ANSWER: B

Q2. What are 'features' in machine learning?

- A) The final predictions made by the model
- B) Input variables or attributes used for prediction
- C) The training algorithm used
- D) The accuracy score of the model

ANSWER: B

Q3. What are 'labels' in supervised learning?

- A) Names given to datasets
- B) Target outputs or categories for training examples
- C) Feature names
- D) Model hyperparameters

ANSWER: B

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

- A) A dataset
- B) A visualization tool
- C) A mathematical representation learned from data
- D) A programming language

ANSWER: C

Q5. What is 'inference' in machine learning?

- A) Training a model
- B) Making predictions on new, unseen data
- C) Cleaning the data
- D) Selecting features

ANSWER: B

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

- A) Unsupervised learning
- B) Reinforcement learning
- C) Supervised learning
- D) Semi-supervised learning

ANSWER: C

Q7. What is the primary goal of supervised learning?

- A) Discover hidden patterns
- B) Maximize rewards
- C) Learn a mapping function from inputs to outputs
- D) Group similar items

ANSWER: C

Q8. Which of the following is a classification task?

- A) Predicting house prices
- B) Detecting spam emails
- C) Forecasting stock prices
- D) Estimating temperature

ANSWER: B

Q9. Which of the following is a regression task?

- A) Spam detection
- B) Image classification
- C) Predicting house prices
- D) Customer segmentation

ANSWER: C

Q10. What distinguishes regression from classification?

- A) Regression predicts continuous values, classification predicts discrete categories
- B) Regression is unsupervised, classification is supervised
- C) Regression is faster than classification
- D) Regression requires more features

ANSWER: A

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

- A) Only linear regression can do classification
- B) Logistic regression, decision trees, SVMs, and neural networks
- C) Only neural networks can classify
- D) Classification requires unsupervised learning

ANSWER: B

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

- A) Only logistic regression
- B) Linear regression, polynomial regression, ridge/lasso regression
- C) Only neural networks
- D) Regression doesn't use algorithms

ANSWER: B

Q13. What is the primary goal of unsupervised learning?

- A) Predict future values with labeled data
- B) Classify data into predefined categories
- C) Discover hidden patterns in unlabeled data
- D) Maximize cumulative rewards

ANSWER: C

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

- A) Email spam detection
- B) Customer segmentation through clustering
- C) House price prediction
- D) Medical diagnosis

ANSWER: B

Q15. What is clustering in unsupervised learning?

- A) Predicting continuous values
- B) Grouping similar data points together
- C) Classifying into predefined categories
- D) Reducing model complexity

ANSWER: B

Q16. What are common clustering algorithms?

- A) Linear regression, logistic regression
- B) K-means, hierarchical clustering, DBSCAN
- C) Decision trees only
- D) Only neural networks

ANSWER: B

Q17. What is dimensionality reduction?

- A) Removing outliers
- B) Reducing features while preserving information
- C) Increasing the number of samples
- D) Simplifying model architecture

ANSWER: B

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

- A) Only PCA
- B) PCA, t-SNE, UMAP, autoencoders
- C) Only neural networks
- D) Linear regression

ANSWER: B

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

- A) The training dataset
- B) The decision-maker learning optimal behavior
- C) The loss function
- D) A type of neural network

ANSWER: B

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

- A) The Python IDE
- B) The system the agent interacts with
- C) The training data
- D) The model architecture

ANSWER: B

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

- A) Training accuracy

- B) Number of features
- C) Cumulative rewards over time
- D) Dataset size

ANSWER: C

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

- A) The final model accuracy
- B) Feedback signal indicating action quality
- C) The number of training epochs
- D) Feature importance score

ANSWER: B

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

- A) Data cleaning procedure
- B) Strategy mapping states to actions
- C) Feature selection method
- D) Loss function

ANSWER: B

Q24. What are applications of reinforcement learning?

- A) Only spam detection
- B) Game playing, robotics, autonomous vehicles, resource optimization
- C) Only image classification
- D) Only regression tasks

ANSWER: B

Q25. What are common reinforcement learning algorithms mentioned?

- A) Linear regression only
- B) Q-Learning, SARSA, Deep Q-Networks, Policy Gradients
- C) Only K-means
- D) Only PCA

ANSWER: B

LEVEL 2: ML TRIBES AND MATHEMATICAL FOUNDATIONS (Questions 26-45)

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

- A) Linear, Logistic, Decision, Neural, Support
- B) Symbolists, Connectionists, Evolutionaries, Bayesians, Analogizers
- C) Supervised, Unsupervised, Reinforcement, Semi-supervised, Self-supervised
- D) Classification, Regression, Clustering, Dimensionality, Association

ANSWER: B

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

- A) Connectionists
- B) Evolutionaries
- C) Symbolists
- D) Bayesians

ANSWER: C

Q28. What algorithm family do Symbolists primarily use?

- A) Neural networks
- B) Decision trees and rule-based systems
- C) Genetic algorithms
- D) Bayesian networks

ANSWER: B

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

- A) Symbolists
- B) Connectionists
- C) Evolutionaries
- D) Analogizers

ANSWER: B

Q30. What algorithms do Connectionists use?

- A) Decision trees
- B) Deep learning, CNNs, RNNs, transformers
- C) Genetic algorithms
- D) K-nearest neighbors

ANSWER: B

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

- A) Maximum interpretability
- B) Handling unstructured data and learning complex representations
- C) Requires minimal data
- D) Always faster than other methods

ANSWER: B

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

- A) Symbolists
- B) Connectionists
- C) Evolutionaries
- D) Bayesians

ANSWER: C

Q33. What algorithms do Evolutionaries use?

- A) Neural networks
- B) Genetic programming and evolutionary strategies
- C) Decision trees
- D) Linear regression

ANSWER: B

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

- A) Symbolists
- B) Connectionists
- C) Evolutionaries
- D) Bayesians

ANSWER: D

Q35. What algorithms do Bayesians use?

- A) Neural networks only
- B) Naive Bayes, Bayesian networks, Gaussian processes
- C) Decision trees only
- D) Genetic algorithms

ANSWER: B

Q36. What is a strength of Bayesian approaches?

- A) Fastest training time
- B) Uncertainty quantification and prior knowledge integration
- C) No assumptions required
- D) Works without any data

ANSWER: B

Q37. Which ML tribe learns by recognizing similarities?

- A) Symbolists
- B) Connectionists
- C) Evolutionaries
- D) Analogizers

ANSWER: D

Q38. What algorithms do Analogizers use?

- A) Neural networks only
- B) K-nearest neighbors, SVMs, kernel methods
- C) Only genetic algorithms
- D) Only decision trees

ANSWER: B

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

- A) Calculus
- B) Linear algebra
- C) Topology
- D) Number theory

ANSWER: B

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

- A) Linear algebra
- B) Calculus
- C) Probability theory
- D) Statistics

ANSWER: B

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

- A) Linear algebra
- B) Calculus
- C) Probability theory
- D) Discrete mathematics

ANSWER: C

Q42. Which statistical concepts are important for ML?

- A) Only mean
- B) Mean, variance, hypothesis testing, confidence intervals
- C) Only standard deviation
- D) None - statistics not needed

ANSWER: B

Q43. What is a common misconception about machine learning?

- A) It requires math knowledge
- B) It's a black box that magically solves all problems
- C) It needs training data
- D) Models need evaluation

ANSWER: B

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

- A) Yes, always true
- B) No, simpler models can outperform complex ones
- C) Yes, if you have enough data
- D) Complex models are always worse

ANSWER: B

Q45. Can machine learning replace domain expertise entirely?

- A) Yes, ML doesn't need domain knowledge
- B) No, domain knowledge remains essential for success
- C) Yes, if the model is complex enough
- D) Only for simple problems

ANSWER: B

LEVEL 3: DATA PREPROCESSING FUNDAMENTALS (Questions 46-70)

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

- A) 10-20%
- B) 30-40%
- C) 60-80%
- D) 90-95%

ANSWER: C

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

- A) Models need expensive hardware
- B) Data quality directly determines model success
- C) Complex algorithms fix bad data
- D) More data always helps

ANSWER: B

Q48. What is MCAR in missing data terminology?

- A) Missing at Random
- B) Missing Completely at Random
- C) Missing Continuously at Random
- D) Missing Constantly at Random

ANSWER: B

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

- A) Missing values have patterns
- B) No relationship between missing values and any variables
- C) Missing values depend on observed data
- D) Missing values depend on unobserved data

ANSWER: B

Q50. What is MAR in missing data terminology?

- A) Missing at Random
- B) Missing as Required
- C) Missing and Removed
- D) Missing after Regression

ANSWER: A

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

- A) No relationship with any variables
- B) Missingness related to observed but not unobserved data
- C) Missingness related to the missing values themselves
- D) Complete randomness

ANSWER: B

Q52. What is MNAR in missing data terminology?

- A) Missing Not at Random
- B) Missing Near at Random
- C) Missing Naturally at Random
- D) Missing Never at Random

ANSWER: A

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

- A) No relationship with variables
- B) Related to observed data only
- C) Missingness related to the missing values themselves
- D) Complete randomness

ANSWER: C

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

- A) Multiple imputation
- B) KNN imputation
- C) Listwise deletion (removing rows with missing values)
- D) Model-based imputation

ANSWER: C

Q55. What is listwise deletion?

- A) Deleting random rows
- B) Removing entire rows with any missing values
- C) Deleting only missing values
- D) Removing columns

ANSWER: B

Q56. What is a disadvantage of listwise deletion?

- A) Too complex
- B) Wasteful - loses potentially valuable data
- C) Too slow
- D) Requires advanced math

ANSWER: B

Q57. What is mean imputation?

- A) Deleting missing values
- B) Replacing missing values with the feature mean
- C) Using the mode
- D) Random value assignment

ANSWER: B

Q58. What is median imputation typically used for?

- A) Categorical variables
- B) When the distribution has outliers
- C) Only for small datasets
- D) Never used in practice

ANSWER: B

Q59. What is mode imputation used for?

- A) Continuous numerical variables
- B) Categorical variables
- C) Only for large datasets
- D) Time series data only

ANSWER: B

Q60. What is forward fill imputation?

- A) Using future values
- B) Using previous values to fill missing values (for time series)
- C) Random imputation
- D) Mean imputation

ANSWER: B

Q61. What is KNN imputation?

- A) Deleting missing values
- B) Using weighted average of K similar instances to fill missing values
- C) Always using the mean
- D) Random value assignment

ANSWER: B

Q62. What is multiple imputation?

- A) Filling with mean only
- B) Generating multiple complete datasets and combining results
- C) Deleting multiple rows
- D) Using only one method

ANSWER: B

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

- A) Feature scaling
- B) Outlier detection (values beyond $\pm 3\sigma$)
- C) Missing value imputation
- D) Feature selection

ANSWER: B

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

- A) Values beyond ± 1 standard deviation
- B) Values beyond ± 2 standard deviations
- C) Values beyond ± 3 standard deviations
- D) Any value above mean

ANSWER: C

Q65. What does IQR stand for?

- A) Interquartile Regression
- B) Interquartile Range
- C) Internal Quality Ratio
- D) Iterative Quality Ranking

ANSWER: B

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

- A) Beyond ± 1 standard deviation
- B) Beyond ± 2 standard deviations
- C) Below $Q1 - 1.5 \times IQR$ or above $Q3 + 1.5 \times IQR$
- D) Any value above median

ANSWER: C

Q67. What is Isolation Forest?

- A) A tree-based ensemble method
- B) An ML-based outlier detection method using random partitioning
- C) A feature selection technique
- D) A regression algorithm

ANSWER: B

Q68. What is LOF (Local Outlier Factor)?

- A) A regression method
- B) An outlier detection method comparing local density to neighbors
- C) A clustering algorithm
- D) A feature scaling technique

ANSWER: B

Q69. What is DBSCAN used for?

- A) Regression only
- B) Clustering and identifying outliers in low-density regions
- C) Feature selection only
- D) Data imputation

ANSWER: B

Q70. What are common strategies for treating outliers?

- A) Always delete them
- B) Remove (if errors), transform (log), cap (winsorize), or keep (if genuine)
- C) Always keep them
- D) Ignore them completely

ANSWER: B

LEVEL 4: FEATURE SCALING (Questions 71-85)

Q71. Why is feature scaling important?

- A) Makes data look better
- B) Ensures features contribute proportionally to distance-based algorithms
- C) Speeds up data collection
- D) Reduces dataset size

ANSWER: B

Q72. Which algorithms require feature scaling?

- A) Only decision trees
- B) PCA, SVM, logistic regression, neural networks, K-means
- C) All algorithms equally
- D) No algorithms need it

ANSWER: B

Q73. Which algorithms do NOT require feature scaling?

- A) K-nearest neighbors
- B) Support Vector Machines
- C) Tree-based models (decision trees, random forests, XGBoost)
- D) Neural networks

ANSWER: C

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

- A) Scaling to [0,1]
- B) Transforming to mean=0, standard deviation=1
- C) Scaling to [-1,1]
- D) Dividing by maximum value

ANSWER: B

Q75. What is the formula for standardization?

- A) $(x - \min) / (\max - \min)$
- B) $(x - \mu) / \sigma$
- C) x / \max
- D) $x - \text{median}$

ANSWER: B

Q76. What does standardization transform data to?

- A) Range [0, 1]
- B) Mean = 0, Standard deviation = 1
- C) Range [-1, 1]
- D) Sum = 1

ANSWER: B

Q77. Does standardization preserve outlier information?

- A) No, it removes outliers
- B) Yes, it preserves outlier information
- C) Only sometimes
- D) It creates new outliers

ANSWER: B

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

- A) $(x - \mu) / \sigma$
- B) $(x - \text{min}) / (\text{max} - \text{min})$
- C) x / median
- D) $(x - \text{median}) / \text{IQR}$

ANSWER: B

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

- A) Mean = 0, Std = 1
- B) Range [0, 1]
- C) Range $[-\infty, +\infty]$
- D) Median = 0

ANSWER: B

Q80. Is Min-Max scaling sensitive to outliers?

- A) No, very robust to outliers
- B) Yes, outliers compress other values
- C) Never affected by outliers
- D) Only in small datasets

ANSWER: B

Q81. What is Robust Scaling?

- A) Using mean and standard deviation
- B) Using median and IQR (robust to outliers)
- C) Using min and max
- D) Using only the mean

ANSWER: B

Q82. What is the formula for Robust Scaling?

- A) $(x - \text{mean}) / \text{std}$
- B) $(x - \text{median}) / \text{IQR}$
- C) $(x - \text{min}) / (\text{max} - \text{min})$
- D) x / max

ANSWER: B

Q83. When should you use Robust Scaling?

- A) When data has no outliers
- B) When data contains many outliers
- C) Never in practice
- D) Only for categorical data

ANSWER: B

Q84. What is MaxAbs Scaling?

- A) $(x - \text{mean}) / \text{std}$
- B) $x / |\text{max}|$
- C) $(x - \text{min}) / (\text{max} - \text{min})$
- D) $(x - \text{median}) / \text{IQR}$

ANSWER: B

Q85. What is an advantage of MaxAbs Scaling?

- A) Always better than other methods
- B) Preserves sparsity in sparse matrices
- C) Removes all outliers
- D) Works only with negative values

ANSWER: B

LEVEL 5: CRITICAL DATA PREPROCESSING CONCEPTS (Questions 86-100)

Q86. When should you fit a scaler?

- A) On the entire dataset including test set
- B) On training data only, then transform both train and test
- C) On test data only
- D) Separately on train and test

ANSWER: B

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

- A) It's too slow
- B) It causes data leakage - test information influences training
- C) It's more accurate
- D) No problem at all

ANSWER: B

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

- A) Delete them
- B) Store them for transforming test data and production deployment
- C) Recalculate for each new dataset
- D) Only use once

ANSWER: B

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

- A) They're too simple
- B) They make splits based on feature values, not distances
- C) They're always worse
- D) They require special scaling

ANSWER: B

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

- A) Model trains faster
- B) Features with larger scales dominate distance calculations
- C) Model becomes more accurate
- D) Nothing - no impact

ANSWER: B

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

- A) Breaks the model
- B) No improvement - wastes computation, doesn't help
- C) Always improves accuracy
- D) Required for it to work

ANSWER: B

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

- A) No scaling needed
- B) Standardization or Min-Max scaling
- C) Only MaxAbs scaling
- D) Only Robust scaling

ANSWER: B

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

- A) Preprocess before splitting
- B) Split data first, then fit preprocessing only on training set
- C) Doesn't matter - can do either
- D) Always preprocess test set separately

ANSWER: B

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

- A) Engineers are slow
- B) Data quality issues require extensive cleaning and transformation
- C) It's unnecessarily complex
- D) Tools are inefficient

ANSWER: B

Q95. Raw data typically contains which issues?

- A) None - usually perfect
- B) Noise, missing values, inconsistencies, irrelevant features
- C) Only missing values
- D) Only outliers

ANSWER: B

Q96. What degrades model accuracy the most?

- A) Simple algorithms
- B) Poor quality data (noise, errors, missing values)
- C) Small datasets always
- D) Fast training

ANSWER: B

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

- A) No relationship
- B) Direct relationship - better data quality leads to better models
- C) Inverse relationship
- D) Random relationship

ANSWER: B

Q98. Can sophisticated algorithms compensate for poor data quality?

- A) Yes, always
- B) No, "garbage in, garbage out" principle applies
- C) Yes, if model is complex enough
- D) Only sometimes

ANSWER: B

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

- A) It's not important
- B) Different mechanisms require different handling strategies
- C) All handled the same way
- D) Only for documentation

ANSWER: B

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

- A) Always delete outliers
- B) Domain context - whether outliers are errors or genuine extreme cases
- C) Random selection
- D) Always keep all outliers

ANSWER: B

END OF QUESTIONS

SCORE INTERPRETATION GUIDE:

These 100 questions cover ONLY the first 11 pages of the ML Curriculum Analysis:

Content NOT covered (appears on pages 12+):

SCORING TARGETS:

STUDY FOCUS BY SCORE RANGE: