Machine Learning

100 Multiple Choice Questions

Ranked by Ascending Complexity

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

MACHINE LEARNING - 100 MULTIPLE CHOICE QUESTIONS

Ranked by Ascending Complexity

LEVEL 1: BASIC CONCEPTS (Questions 1-15)

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. Which type of learning uses labeled training data?

A) Unsupervised learning

B) Reinforcement learning

C) Supervised learning

D) Semi-supervised learning

ANSWER: C

Q3. In machine learning terminology, what are 'features'?

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

Q4. 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 rewards through trial and error

ANSWER: C

Q5. 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

Q6. What distinguishes regression from classification?

A) Regression predicts continuous values, classification predicts discrete categories

B) Regression is unsupervised, classification is supervised

C) Regression uses neural networks, classification uses decision trees

D) Regression requires more data than classification

ANSWER: A

Q7. In reinforcement learning, what does an agent learn to maximize?

A) Training accuracy

B) Number of features

C) Cumulative rewards

D) Data preprocessing speed

ANSWER: C

Q8. What is a 'label' in supervised learning?

A) The name of the dataset

B) The target output or category for training examples

C) The number of features in the data

D) The algorithm's hyperparameters

ANSWER: B

Q9. Which ML approach groups similar data points together?

A) Regression

B) Classification

C) Clustering

D) Dimensionality reduction

ANSWER: C

Q10. 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

Q11. What does the principle 'garbage in, garbage out' emphasize?

A) Models need expensive hardware

B) Data quality directly determines model success

C) Complex algorithms always perform better

D) More data is always better

ANSWER: B

Q12. What is the purpose of splitting data into training and test sets?

A) To increase the amount of data

B) To evaluate model performance on unseen data

C) To speed up training

D) To reduce overfitting during training

ANSWER: B

Q13. What is the typical ratio for a train/test split?

A) 50/50

B) 60/40

C) 80/20

D) 95/5

ANSWER: C

Q14. Which type of missing data has no relationship with any variables?

A) MAR (Missing at Random)

B) MCAR (Missing Completely at Random)

C) MNAR (Missing Not at Random)

D) MFR (Missing for Reasons)

ANSWER: B

Q15. What is the simplest method to handle missing data?

A) Multiple imputation

B) KNN imputation

C) Deletion (removing rows with missing values)

D) Model-based imputation

ANSWER: C

LEVEL 2: LOW-MEDIUM COMPLEXITY (Questions 16-40)

Q16. What does standardization (Z-score normalization) transform features to?

A) Range [0, 1]

B) Mean = 0, Standard deviation = 1

C) Range [-1, 1]

D) Sum = 1

ANSWER: B

Q17. Which imputation method replaces missing values with the average?

A) Forward fill

B) Mode imputation

C) Mean imputation

D) Hot-deck imputation

ANSWER: C

Q18. What is the IQR method used for?

A) Feature scaling

B) Outlier detection

C) Missing value imputation

D) Feature selection

ANSWER: B

Q19. In the IQR method, outliers are typically defined as values:

A) Beyond ±1 standard deviation

B) Beyond ±2 standard deviations

C) Below Q1 - 1.5×IQR or above Q3 + 1.5×IQR

D) In the top or bottom 5% of data

ANSWER: C

Q20. What does Min-Max scaling transform features to?

A) Mean = 0, Std = 1

B) Range [0, 1]

C) Range [-∞, +∞]

D) Median = 0

ANSWER: B

Q21. Which type of encoding is appropriate for nominal categorical variables with few categories?

A) Label encoding

B) Target encoding

C) One-hot encoding

D) Frequency encoding

ANSWER: C

Q22. What is the main disadvantage of one-hot encoding with high-cardinality features?

A) It's computationally slow

B) It creates many new columns, increasing dimensionality

C) It only works for binary variables

D) It requires the target variable

ANSWER: B

Q23. When should you use stratified sampling in train/test split?

A) When you have very large datasets

B) When you have imbalanced classes

C) When you have missing values

D) When you have continuous target variables

ANSWER: B

Q24. What is K in K-Fold cross-validation?

A) The number of features

B) The number of folds (subsets) the data is divided into

C) The number of training epochs

D) The number of hyperparameters

ANSWER: B

Q25. What is a common value for K in K-Fold cross-validation?

A) K = 2

B) K = 5 or K = 10

C) K = 50

D) K = 100

ANSWER: B

Q26. What does accuracy measure in classification?

A) Only the correctly predicted positive cases

B) The percentage of all correct predictions

C) Only the correctly predicted negative cases

D) The balance between precision and recall

ANSWER: B

Q27. What does MAE stand for in regression metrics?

A) Maximum Absolute Error

B) Mean Accuracy Estimation

C) Mean Absolute Error

D) Median Average Error

ANSWER: C

Q28. Which metric is less sensitive to outliers?

A) Mean Squared Error (MSE)

B) Root Mean Squared Error (RMSE)

C) Mean Absolute Error (MAE)

D) R-squared

ANSWER: C

Q29. In linear regression, what does the coefficient β represent?

A) The error term

B) The change in output for a unit change in the corresponding feature

C) The total number of features

D) The training accuracy

ANSWER: B

Q30. What function does logistic regression use to produce probabilities?

A) Linear function

B) Exponential function

C) Sigmoid function

D) Logarithmic function

ANSWER: C

Q31. What assumption does Naive Bayes make about features?

A) Features are linearly related

B) Features are independent of each other given the class

C) Features are normally distributed

D) Features have equal variance

ANSWER: B

Q32. Which Naive Bayes variant is used for continuous features?

A) Multinomial Naive Bayes

B) Bernoulli Naive Bayes

C) Gaussian Naive Bayes

D) Complement Naive Bayes

ANSWER: C

Q33. In KNN, what does the 'K' represent?

A) The number of features

B) The number of nearest neighbors to consider

C) The number of clusters

D) The number of iterations

ANSWER: B

Q34. Why is feature scaling critical for KNN?

A) It speeds up training

B) It reduces memory usage

C) KNN uses distance metrics which are sensitive to feature scales

D) It improves interpretability

ANSWER: C

Q35. What is the most commonly used distance metric in KNN?

A) Manhattan distance

B) Cosine similarity

C) Euclidean distance

D) Hamming distance

ANSWER: C

Q36. What does a confusion matrix display?

A) The correlation between features

B) The actual vs predicted classifications

C) The distribution of feature values

D) The learning curve over time

ANSWER: B

Q37. What are True Positives (TP) in a confusion matrix?

A) Predicted positive, actually negative

B) Predicted negative, actually positive

C) Predicted positive, actually positive

D) Predicted negative, actually negative

ANSWER: C

Q38. What transformation is commonly applied to right-skewed data?

A) Square transformation

B) Log transformation

C) Exponential transformation

D) Polynomial transformation

ANSWER: B

Q39. Which algorithm grows a tree by recursively splitting data?

A) K-Nearest Neighbors

B) Naive Bayes

C) Decision Tree

D) Logistic Regression

ANSWER: C

Q40. What does Gini impurity measure in decision trees?

A) The depth of the tree

B) The probability of incorrect classification

C) The number of features used

D) The training time

ANSWER: B

LEVEL 3: MEDIUM COMPLEXITY (Questions 41-65)

Q41. What is precision in classification metrics?

A) TP / (TP + FP) - Of predicted positives, how many are actually positive

B) TP / (TP + FN) - Of actual positives, how many were predicted

C) (TP + TN) / Total - Overall accuracy

D) TN / (TN + FP) - Of actual negatives, how many were predicted

ANSWER: A

Q42. What is recall (sensitivity) in classification metrics?

A) TP / (TP + FP)

B) TP / (TP + FN) - Of actual positives, how many were correctly identified

C) TN / (TN + FP)

D) (TP + TN) / Total

ANSWER: B

Q43. When false negatives are very costly (e.g., disease diagnosis), which metric should you optimize?

A) Precision

B) Accuracy

C) Recall

D) Specificity

ANSWER: C

Q44. When false positives are very costly (e.g., spam detection of important emails), which metric should you optimize?

A) Recall

B) Precision

C) Accuracy

D) F2-score

ANSWER: B

Q45. What does the F1-score represent?

A) The arithmetic mean of precision and recall

B) The harmonic mean of precision and recall

C) The geometric mean of precision and recall

D) The maximum of precision and recall

ANSWER: B

Q46. What does an AUC (Area Under ROC Curve) of 0.5 indicate?

A) Perfect classification

B) Random classification (no discriminative power)

C) Good classification performance

D) The model predicts everything as negative

ANSWER: B

Q47. What does an R^2 score of 0 indicate in regression?

A) Perfect fit

B) The model performs as well as predicting the mean

C) The model has no predictive power

D) The model is worse than random

ANSWER: B

Q48. What is the bias-variance tradeoff?

A) The balance between training time and accuracy

B) The balance between model simplicity (high bias) and complexity (high variance)

C) The balance between precision and recall

D) The balance between training and test set size

ANSWER: B

Q49. What characterizes overfitting?

A) Poor performance on both training and test sets

B) Excellent training performance but poor test performance

C) Poor training performance but good test performance

D) Moderate performance on both training and test sets

ANSWER: B

Q50. What characterizes underfitting?

A) Excellent training performance, poor test performance

B) Poor performance on both training and test sets

C) Perfect fit to training data

D) High variance in predictions

ANSWER: B

Q51. What is the primary solution to overfitting?

A) Increase model complexity

B) Remove regularization

C) Get more training data or reduce model complexity

D) Decrease the number of training epochs

ANSWER: C

Q52. What is the primary solution to underfitting?

A) Reduce the number of features

B) Add more regularization

C) Increase model complexity or add more relevant features

D) Use a simpler algorithm

ANSWER: C

Q53. What is data leakage in machine learning?

A) Missing values in the dataset

B) When test data information influences the training process

C) When the model is too complex

D) When features are highly correlated

ANSWER: B

Q54. Which is an example of data leakage?

A) Using cross-validation

B) Fitting a scaler on the entire dataset before splitting

C) Using regularization

D) Removing outliers from training data only

ANSWER: B

Q55. What does L1 regularization (Lasso) do?

A) Shrinks all coefficients equally

B) Drives some coefficients to exactly zero, performing feature selection

C) Only applies to neural networks

D) Increases model complexity

ANSWER: B

Q56. What does L2 regularization (Ridge) do?

A) Sets some coefficients to exactly zero

B) Shrinks all coefficients but keeps all features

C) Increases model capacity

D) Only works with decision trees

ANSWER: B

Q57. What is target encoding?

A) Encoding the target variable as 0 and 1

B) Replacing categories with the mean of the target variable for that category

C) Creating binary columns for each category

D) Sorting categories alphabetically

ANSWER: B

Q58. What is the main risk of target encoding?

A) It's computationally expensive

B) It only works for binary classification

C) It can cause target leakage and overfitting

D) It requires too much memory

ANSWER: C

Q59. What is feature engineering?

A) Selecting the best subset of existing features

B) Creating new features from existing data using domain knowledge

C) Removing outliers from features

D) Scaling features to the same range

ANSWER: B

Q60. Which feature selection method is model-independent?

A) Recursive Feature Elimination (RFE)

B) Tree-based feature importance

C) Correlation analysis and chi-squared test (Filter methods)

D) Lasso regularization

ANSWER: C

Q61. What is the curse of dimensionality?

A) Having too little data

B) When having too many features causes problems like sparsity and overfitting

C) When features are highly correlated

D) When the model is too simple

ANSWER: B

Q62. What does PCA (Principal Component Analysis) do?

A) Selects the most important original features

B) Creates new uncorrelated features that capture maximum variance

C) Removes outliers from the dataset

D) Predicts the target variable

ANSWER: B

Q63. In decision trees, what is pruning?

A) Removing features from the dataset

B) Removing branches from the tree to reduce overfitting

C) Adding more splits to improve accuracy

D) Scaling the feature values

ANSWER: B

Q64. What is the max_depth hyperparameter in decision trees?

A) The maximum number of features to consider

B) The maximum number of samples per leaf

C) The maximum depth (levels) the tree can grow

D) The maximum training time

ANSWER: C

Q65. Why are decision trees prone to overfitting?

A) They require feature scaling

B) They can grow very deep and memorize training data noise

C) They only work with numerical features

D) They cannot handle missing values

ANSWER: B

LEVEL 4: MEDIUM-HIGH COMPLEXITY (Questions 66-85)

Q66. What is the main advantage of Random Forest over a single decision tree?

A) Faster training time

B) Better interpretability

C) Reduces overfitting through ensemble averaging

D) Requires less memory

ANSWER: C

Q67. How does Random Forest create diversity among trees?

A) Using different algorithms for each tree

B) Bootstrap sampling of data and random feature subsampling at splits

C) Using different loss functions

D) Training on different target variables

ANSWER: B

Q68. What is bootstrap sampling in Random Forest?

A) Sampling without replacement

B) Sampling with replacement to create datasets of same size

C) Splitting data sequentially

D) Removing outliers before sampling

ANSWER: B

Q69. What is Out-of-Bag (OOB) evaluation in Random Forest?

A) Validation using samples not included in bootstrap sample for each tree

B) Testing on a separate holdout set

C) Cross-validation with 10 folds

D) Evaluation on the training set

ANSWER: A

Q70. In Random Forest, what is the default max_features for classification?

A) All features (n_features)

B) Half the features (n_features/2)

C) Square root of features (sqrt(n_features))

D) Log of features (log(n_features))

ANSWER: C

Q71. What is the key difference between bagging and boosting?

A) Bagging trains models in parallel, boosting trains sequentially

B) Bagging only works with trees, boosting works with any algorithm

C) Bagging reduces bias, boosting reduces variance

D) Bagging requires more memory than boosting

ANSWER: A

Q72. In Gradient Boosting, what does each new tree learn to predict?

A) The original target variable

B) The residual errors of the previous ensemble

C) A random subset of the data

D) The feature importances

ANSWER: B

Q73. What is the learning rate in Gradient Boosting?

A) How fast the algorithm trains

B) The shrinkage factor applied to each tree's contribution

C) The rate at which data is sampled

D) The speed of cross-validation

ANSWER: B

Q74. Why use a low learning rate in Gradient Boosting?

A) Faster training

B) Fewer trees needed

C) Better generalization with more trees

D) Reduced memory usage

ANSWER: C

Q75. What is Grid Search in hyperparameter tuning?

A) Randomly sampling hyperparameter combinations

B) Exhaustively trying all combinations in a predefined grid

C) Using gradient descent to find optimal hyperparameters

D) Selecting hyperparameters based on expert knowledge

ANSWER: B

Q76. What is the main advantage of Random Search over Grid Search?

A) It always finds the global optimum

B) It's more thorough

C) It can explore a wider range with the same budget

D) It requires less memory

ANSWER: C

Q77. What is Bayesian Optimization?

A) Random sampling of hyperparameters

B) Using a probabilistic model to guide search toward promising regions

C) Grid search with cross-validation

D) Manual hyperparameter selection

ANSWER: B

Q78. In SVM, what is the margin?

A) The error on the training set

B) The distance between the hyperplane and the nearest data points

C) The number of support vectors

D) The regularization parameter

ANSWER: B

Q79. What are support vectors in SVM?

A) All training data points

B) The misclassified points

C) The data points closest to the decision boundary that define it

D) The feature vectors

ANSWER: C

Q80. What does the kernel trick in SVM do?

A) Speeds up training

B) Implicitly maps data to higher dimensions for non-linear separation

C) Reduces the number of features

D) Handles missing values

ANSWER: B

Q81. What is the most popular kernel for SVM?

A) Linear kernel

B) Polynomial kernel

C) RBF (Radial Basis Function) or Gaussian kernel

D) Sigmoid kernel

ANSWER: C

Q82. In SVM, what does the C parameter control?

A) The kernel width

B) The tradeoff between margin size and classification errors

C) The number of support vectors

D) The polynomial degree

ANSWER: B

Q83. What is nested cross-validation used for?

A) Faster training

B) Unbiased evaluation: inner loop for tuning, outer loop for evaluation

C) Feature selection only

D) Reducing overfitting during training

ANSWER: B

Q84. What is SHAP (SHapley Additive exPlanations)?

A) A new machine learning algorithm

B) A method for explaining individual predictions using game theory

C) A hyperparameter tuning technique

D) A type of neural network layer

ANSWER: B

Q85. Why is feature importance from Random Forest sometimes misleading?

A) It's too slow to compute

B) It can be biased toward high-cardinality and continuous features

C) It only works for binary classification

D) It requires the target variable

ANSWER: B

LEVEL 5: HIGH COMPLEXITY (Questions 86-100)

Q86. What is backpropagation in neural networks?

A) Forward pass through the network

B) The algorithm for computing gradients using the chain rule

C) A regularization technique

D) A method for selecting features

ANSWER: B

Q87. What is the purpose of activation functions in neural networks?

A) To scale inputs

B) To introduce non-linearity enabling learning of complex patterns

C) To reduce overfitting

D) To speed up training

ANSWER: B

Q88. What is the most common activation function for hidden layers in modern neural networks?

A) Sigmoid

B) Tanh

C) ReLU (Rectified Linear Unit)

D) Softmax

ANSWER: C

Q89. What is dropout in neural networks?

A) Removing features from input

B) Randomly dropping neurons during training as regularization

C) Stopping training early

D) Removing outliers from data

ANSWER: B

Q90. What problem does batch normalization address?

A) Overfitting

B) Internal covariate shift, stabilizing and speeding up training

C) Class imbalance

D) Missing values

ANSWER: B

Q91. What is the vanishing gradient problem?

A) When gradients become too large

B) When gradients become very small in deep networks, preventing learning

C) When the model has no gradients

D) When validation error is too high

ANSWER: B

Q92. What is the Adam optimizer?

A) A simple gradient descent variant

B) An adaptive learning rate optimizer combining momentum and RMSprop

C) A regularization technique

D) A neural network architecture

ANSWER: B

Q93. In Gradient Boosting, why are shallow trees (depth 3-5) often preferred?

A) They train faster

B) They prevent individual trees from overfitting; ensemble handles complexity

C) They require less memory

D) They work better with categorical features

ANSWER: B

Q94. What is the key innovation of XGBoost over standard Gradient Boosting?

A) It uses decision trees

B) Regularization, parallel processing, and handling missing values

C) It's written in Python

D) It only works for classification

ANSWER: B

Q95. What is early stopping in iterative algorithms?

A) Starting training early

B) Stopping training when validation performance stops improving

C) Stopping at a fixed number of iterations

D) Stopping when training error is zero

ANSWER: B

Q96. What is concept drift in ML deployment?

A) When the model architecture changes

B) When the statistical properties of the target variable change over time

C) When features are added or removed

D) When the training data is shuffled

ANSWER: B

Q97. In an end-to-end ML project, when should you use the test set?

A) Throughout hyperparameter tuning

B) After every experiment to check progress

C) Only once at the very end for final evaluation

D) Before training to check data quality

ANSWER: C

Q98. What is the purpose of A/B testing in ML deployment?

A) To test two different training algorithms

B) To compare new model performance against baseline in production

C) To split data into train and test

D) To test for missing values

ANSWER: B

Q99. What is transfer learning in neural networks?

A) Transferring data between computers

B) Using a pre-trained model's knowledge for a new but related task

C) Moving a model to production

D) Converting between different frameworks

ANSWER: B

Q100. What is the most critical mistake that invalidates ML results?

A) Not using the most complex algorithm

B) Data leakage: information from test set influencing training

C) Using too few features

D) Training for too few epochs

ANSWER: B

END OF QUESTIONS

SCORE INTERPRETATION:

BY LEVEL TARGETS: