

# Machine Learning - 100 Questions

Pages 12+ Part 2: Metrics, Algorithms

Nova IMS

Generated: January 15, 2026

## MACHINE LEARNING - 100 MULTIPLE CHOICE QUESTIONS (PART 2)

Based on Pages 12+ of Curriculum Analysis - Continued

(Advanced Metrics, Algorithms, Ensemble Methods, Neural Networks, Projects)

### LEVEL 1: ADVANCED CLASSIFICATION METRICS (Questions 1-20)

**Q1. Which metric should you optimize when false negatives are very costly (e.g., disease diagnosis)?**

- A) Precision
- B) Accuracy
- C) Recall
- D) Specificity

**ANSWER: C**

**Q2. Which metric should you optimize when false positives are very costly (e.g., spam filtering important emails)?**

- A) Recall
- B) Precision
- C) Accuracy
- D) F2-score

**ANSWER: B**

**Q3. What does the F1-score represent?**

- A) Arithmetic mean of precision and recall
- B) Harmonic mean of precision and recall
- C) Geometric mean of precision and recall
- D) Maximum of precision and recall

**ANSWER: B**

**Q4. What is the formula for F1-score?**

- A)  $(\text{Precision} + \text{Recall}) / 2$
- B)  $2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$
- C)  $\text{Precision} \times \text{Recall}$
- D)  $\sqrt{(\text{Precision} \times \text{Recall})}$

**ANSWER: B**

**Q5. When is F1-score most useful?**

- A) Balanced datasets only
- B) Imbalanced datasets needing balance between precision and recall
- C) Never useful
- D) Only for regression

**ANSWER: B**

**Q6. What is the F-Beta score?**

- A) Same as F1-score
- B) Weighted version allowing emphasis on precision or recall
- C) Only for binary classification
- D) A regression metric

**ANSWER: B**

**Q7. What does  $\beta > 1$  in F-Beta score emphasize?**

- A) Precision
- B) Recall

- C) Accuracy
- D) Specificity

**ANSWER: B**

**Q8. What does  $\beta < 1$  in F-Beta score emphasize?**

- A) Recall
- B) Precision
- C) Accuracy
- D) F1-score

**ANSWER: B**

**Q9. What is specificity (True Negative Rate)?**

- A)  $TP / (TP + FP)$
- B)  $TN / (TN + FP)$  - Of actual negatives, how many correctly identified
- C)  $TP / (TP + FN)$
- D)  $(TP + TN) / \text{Total}$

**ANSWER: B**

**Q10. What does ROC stand for?**

- A) Rate of Classification
- B) Receiver Operating Characteristic
- C) Regression Operating Curve
- D) Random Output Classification

**ANSWER: B**

**Q11. What does the ROC curve plot?**

- A) Precision vs Recall
- B) True Positive Rate vs False Positive Rate at various thresholds
- C) Accuracy vs Threshold
- D) Training vs Test error

**ANSWER: B**

**Q12. What does each point on the ROC curve represent?**

- A) A different model
- B) A different classification threshold
- C) A different dataset
- D) A different feature

**ANSWER: B**

**Q13. What does the diagonal line in ROC space represent?**

- A) Perfect classifier
- B) Random classifier
- C) Worst classifier
- D) Optimal classifier

**ANSWER: B**

**Q14. What does AUC stand for?**

- A) Average Under Curve
- B) Area Under (ROC) Curve
- C) Accuracy Under Calculation
- D) Algorithm Using Classification

**ANSWER: B**

**Q15. What does an AUC of 0.5 indicate?**

- A) Perfect classification
- B) Random classification (no discriminative power)
- C) Good performance
- D) Worst possible performance

**ANSWER: B**

**Q16. What does an AUC of 1.0 indicate?**

- A) Random classifier
- B) Perfect classification
- C) Average performance
- D) Failed model

**ANSWER: B**

**Q17. What is considered "good" AUC performance?**

- A) 0.5-0.6
- B) 0.8-0.9
- C) 0.3-0.4
- D) 1.5-2.0

**ANSWER: B**

**Q18. What is Matthews Correlation Coefficient (MCC)?**

- A) Same as accuracy
- B) Balanced measure even for imbalanced classes, range [-1, 1]
- C) Only for binary classification
- D) A regression metric

**ANSWER: B**

**Q19. What MCC value indicates perfect prediction?**

- A) 0
- B) 1
- C) -1
- D) 0.5

**ANSWER: B**

**Q20. What is the Precision-Recall curve useful for?**

- A) All classification problems equally
- B) Highly imbalanced datasets (more informative than ROC)
- C) Only balanced datasets
- D) Regression only

**ANSWER: B**

## LEVEL 2: REGRESSION METRICS (Questions 21-35)

**Q21. What does MAE stand for?**

- A) Maximum Absolute Error
- B) Mean Absolute Error
- C) Median Average Error
- D) Minimum Absolute Error

**ANSWER: B**

**Q22. What is the formula for MAE?**

- A)  $(1/n) \sum (y_{\text{actual}} - y_{\text{predicted}})^2$
- B)  $(1/n) \sum |y_{\text{actual}} - y_{\text{predicted}}|$
- C)  $\sqrt{[(1/n) \sum (y_{\text{actual}} - y_{\text{predicted}})^2]}$
- D)  $1 - (SS_{\text{residual}} / SS_{\text{total}})$

**ANSWER: B**

**Q23. What is an advantage of MAE?**

- A) Penalizes large errors heavily
- B) Easy interpretation, robust to outliers, same units as target
- C) Always better than MSE
- D) Most complex metric

**ANSWER: B**

**Q24. What does MSE stand for?**

- A) Mean Standard Error
- B) Mean Squared Error
- C) Maximum Squared Error
- D) Median Squared Error

**ANSWER: B**

**Q25. What is the formula for MSE?**

- A)  $(1/n) \sum |y - \hat{y}|$
- B)  $(1/n) \sum (y - \hat{y})^2$
- C)  $\sqrt{[(1/n) \sum (y - \hat{y})^2]}$
- D)  $(y - \hat{y}) / y$

**ANSWER: B**

**Q26. What is a characteristic of MSE?**

- A) Linear penalty
- B) Heavily penalizes large errors (squared)
- C) Robust to outliers
- D) Same units as target

**ANSWER: B**

**Q27. What does RMSE stand for?**

- A) Relative Mean Squared Error
- B) Root Mean Squared Error
- C) Random Mean Squared Error
- D) Reduced Mean Squared Error

**ANSWER: B**

**Q28. What is the formula for RMSE?**

- A)  $(1/n) \sum (y - \hat{y})^2$

- B)  $\sqrt{[(1/n) \sum (y - \hat{y})^2]}$
- C)  $(1/n) \sum |y - \hat{y}|$
- D)  $1 - \text{MSE}$

**ANSWER: B**

**Q29. Why is RMSE preferred over MSE?**

- A) Always more accurate
- B) Returns to original units (more interpretable)
- C) Faster to calculate
- D) Ignores outliers

**ANSWER: B**

**Q30. What does  $R^2$  (R-squared) represent?**

- A) Root squared error
- B) Proportion of variance in target explained by model
- C) Random error
- D) Regression coefficient

**ANSWER: B**

**Q31. What is the range of  $R^2$ ?**

- A)  $[0, 1]$
- B)  $(-\infty, 1]$
- C)  $[0, \infty)$
- D)  $[-1, 1]$

**ANSWER: B**

**Q32. What does  $R^2 = 1$  indicate?**

- A) Worst fit
- B) Perfect fit
- C) Random predictions
- D) Average fit

**ANSWER: B**

**Q33. What does  $R^2 = 0$  indicate?**

- A) Perfect fit
- B) Model performs as well as predicting the mean
- C) Worst possible fit
- D) Random performance

**ANSWER: B**

**Q34. What does negative  $R^2$  indicate?**

- A) Good fit
- B) Model is worse than simply predicting the mean
- C) Perfect fit
- D) Not mathematically possible

**ANSWER: B**

**Q35. What is adjusted  $R^2$ ?**

- A) Same as  $R^2$
- B) Adjusts  $R^2$  for number of predictors (penalizes unnecessary features)
- C) Only for classification
- D) Always higher than  $R^2$

**ANSWER: B**

### LEVEL 3: BIAS-VARIANCE TRADEOFF & OVERFITTING (Questions 36-55)

**Q36. What is bias in machine learning?**

- A) Data collection error
- B) Error from wrong assumptions in learning algorithm
- C) Variance in predictions
- D) Training time

**ANSWER: B**

**Q37. What does high bias indicate?**

- A) Overfitting
- B) Underfitting - model too simple
- C) Perfect fit
- D) High variance

**ANSWER: B**

**Q38. What is variance in machine learning?**

- A) Standard deviation of data
- B) Error from sensitivity to small fluctuations in training data
- C) Always better than bias
- D) Dataset size

**ANSWER: B**

**Q39. What does high variance indicate?**

- A) Underfitting
- B) Overfitting - model too complex
- C) Perfect fit
- D) Low bias

**ANSWER: B**

**Q40. What is the total error decomposition?**

- A) Error = Bias + Variance
- B) Expected Error = Bias<sup>2</sup> + Variance + Irreducible Error
- C) Error = Bias × Variance
- D) Error = Bias - Variance

**ANSWER: B**

**Q41. What is irreducible error?**

- A) Can be eliminated with better models
- B) Inherent noise in data that cannot be eliminated
- C) Always zero
- D) Same as bias

**ANSWER: B**

**Q42. What characterizes underfitting?**

- A) Excellent training, poor test performance
- B) Poor performance on both training and test sets
- C) Perfect fit to training data
- D) High variance

**ANSWER: B**

**Q43. What are symptoms of underfitting?**

- A) Large gap between train and test scores

- B) Learning curves plateau at poor performance, curves close together
- C) Perfect training accuracy
- D) High variance in predictions

**ANSWER: B**

**Q44. What are solutions to underfitting?**

- A) Add regularization, reduce features
- B) Increase model complexity, add features, reduce regularization
- C) Get less data
- D) Use simpler algorithm

**ANSWER: B**

**Q45. What characterizes overfitting?**

- A) Poor performance on both sets
- B) Excellent training performance, poor test performance
- C) Moderate performance on both sets
- D) Low bias

**ANSWER: B**

**Q46. What are symptoms of overfitting?**

- A) Both curves plateau at poor performance
- B) Large gap between training and validation curves
- C) Perfect test performance
- D) Low training accuracy

**ANSWER: B**

**Q47. What are solutions to overfitting?**

- A) Increase complexity, remove regularization
- B) More training data, reduce complexity, add regularization
- C) Remove all features
- D) Stop using validation sets

**ANSWER: B**

**Q48. What is regularization?**

- A) Making data regular
- B) Adding penalty to model complexity to prevent overfitting
- C) Removing data
- D) Feature scaling

**ANSWER: B**

**Q49. What does L1 regularization (Lasso) do?**

- A) Shrinks all coefficients equally
- B) Drives some coefficients to exactly zero (feature selection)
- C) Increases complexity
- D) Only for neural networks

**ANSWER: B**

**Q50. What does L2 regularization (Ridge) do?**

- A) Drives coefficients to zero
- B) Shrinks all coefficients but keeps all features
- C) Removes features randomly
- D) Only for decision trees

**ANSWER: B**



**Q51. What does the regularization parameter ( $\alpha$  or  $\lambda$ ) control?**

- A) Learning rate
- B) Strength of penalty (higher = more regularization)
- C) Number of features
- D) Training time

**ANSWER: B**

**Q52. What is early stopping?**

- A) Stopping training early randomly
- B) Stopping when validation performance stops improving
- C) Stopping at fixed epochs
- D) Never stopping training

**ANSWER: B**

**Q53. What is data leakage?**

- A) Missing data
- B) Test data information influencing training process
- C) Data stored incorrectly
- D) Too much data

**ANSWER: B**

**Q54. What is an example of data leakage?**

- A) Using cross-validation correctly
- B) Fitting scaler on entire dataset before splitting
- C) Using regularization
- D) Proper train/test split

**ANSWER: B**

**Q55. Why is data leakage critical to avoid?**

- A) Slows training
- B) Causes overly optimistic performance estimates (invalidates results)
- C) Improves accuracy
- D) No real impact

**ANSWER: B**

## LEVEL 4: LINEAR & LOGISTIC REGRESSION (Questions 56-70)

**Q56. What does linear regression predict?**

- A) Categories
- B) Continuous numerical values
- C) Probabilities only
- D) Clusters

**ANSWER: B**

**Q57. What is the hypothesis function for linear regression?**

- A)  $h(x) = e^{(\beta^T x)}$
- B)  $h(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$
- C)  $h(x) = 1 / (1 + e^{-x})$
- D)  $h(x) = \max(0, x)$

**ANSWER: B**

**Q58. What does linear regression minimize?**

- A) Absolute error
- B) Mean Squared Error (MSE)
- C) Cross-entropy
- D) Hinge loss

**ANSWER: B**

**Q59. What is the Normal Equation in linear regression?**

- A)  $\beta = X^T y$
- B)  $\beta = (X^T X)^{-1} X^T y$
- C)  $\beta = y^T X$
- D)  $\beta = X / y$

**ANSWER: B**

**Q60. What are key assumptions of linear regression?**

- A) No assumptions needed
- B) Linearity, independence, homoscedasticity, normality of residuals
- C) Only linearity
- D) Only normality

**ANSWER: B**

**Q61. What does logistic regression predict?**

- A) Continuous values
- B) Probabilities of class membership (0 to 1)
- C) Integers only
- D) Clusters

**ANSWER: B**

**Q62. What function does logistic regression use?**

- A) Linear function
- B) Sigmoid (logistic) function
- C) ReLU function
- D) Step function

**ANSWER: B**

**Q63. What is the sigmoid function formula?**

- A)  $\sigma(z) = z$

- B)  $\sigma(z) = 1 / (1 + e^{(-z)})$
- C)  $\sigma(z) = e^z$
- D)  $\sigma(z) = \max(0, z)$

**ANSWER: B**

**Q64. What loss function does logistic regression use?**

- A) Mean Squared Error
- B) Log loss (binary cross-entropy)
- C) Hinge loss
- D) Absolute error

**ANSWER: B**

**Q65. What is the decision boundary in logistic regression?**

- A) Non-linear curve
- B)  $\beta^T x = 0$  (linear)
- C) Random boundary
- D) Always circular

**ANSWER: B**

**Q66. What does the softmax function do?**

- A) Reduces dimensionality
- B) Extends logistic regression to multiclass (outputs probability distribution)
- C) Only for binary classification
- D) Scales features

**ANSWER: B**

**Q67. What is an advantage of logistic regression?**

- A) Handles non-linear boundaries well
- B) Fast training, calibrated probabilities, interpretable
- C) Always most accurate
- D) No assumptions

**ANSWER: B**

**Q68. What is a limitation of logistic regression?**

- A) Too complex
- B) Assumes linear decision boundary
- C) Cannot output probabilities
- D) Requires billions of samples

**ANSWER: B**

**Q69. When is logistic regression appropriate?**

- A) Highly non-linear problems
- B) When interpretability critical and decision boundary roughly linear
- C) Never in practice
- D) Only for regression problems

**ANSWER: B**

**Q70. What regularization is common in logistic regression?**

- A) No regularization used
- B) L1 (Lasso) or L2 (Ridge) or ElasticNet
- C) Only dropout
- D) Only batch normalization

**ANSWER: B**

## LEVEL 5: NAIVE BAYES & KNN (Questions 71-85)

**Q71. What theorem is Naive Bayes based on?**

- A) Central Limit Theorem
- B) Bayes' Theorem
- C) Pythagorean Theorem
- D) Fermat's Last Theorem

**ANSWER: B**

**Q72. What is Bayes' Theorem formula?**

- A)  $P(y|X) = P(X) \times P(y)$
- B)  $P(y|X) = P(X|y) \times P(y) / P(X)$
- C)  $P(y|X) = P(X) + P(y)$
- D)  $P(y|X) = P(X) - P(y)$

**ANSWER: B**

**Q73. What "naive" assumption does Naive Bayes make?**

- A) All classes equally likely
- B) Features are independent given the class
- C) Data is normally distributed
- D) No assumptions

**ANSWER: B**

**Q74. What Naive Bayes variant is used for continuous features?**

- A) Multinomial Naive Bayes
- B) Gaussian Naive Bayes
- C) Bernoulli Naive Bayes
- D) Complement Naive Bayes

**ANSWER: B**

**Q75. What Naive Bayes variant is used for text classification (word counts)?**

- A) Gaussian Naive Bayes
- B) Multinomial Naive Bayes
- C) Bernoulli Naive Bayes
- D) Continuous Naive Bayes

**ANSWER: B**

**Q76. What is Laplace smoothing in Naive Bayes?**

- A) Removes noise
- B) Adds small constant to avoid zero probabilities
- C) Scales features
- D) Removes features

**ANSWER: B**

**Q77. What is an advantage of Naive Bayes?**

- A) Always most accurate
- B) Very fast, works with small datasets, handles high dimensions
- C) No assumptions
- D) Perfect for all problems

**ANSWER: B**

**Q78. What is a limitation of Naive Bayes?**

- A) Too slow

- B) Strong independence assumption rarely holds in reality
- C) Requires huge datasets
- D) Cannot handle missing data

**ANSWER: B**

**Q79. What does KNN stand for?**

- A) K-Nearest Networks
- B) K-Nearest Neighbors
- C) K-Numbered Nodes
- D) K-Neural Neighbors

**ANSWER: B**

**Q80. How does KNN make predictions?**

- A) Fits a model during training
- B) Finds K nearest training samples and uses majority vote or average
- C) Uses decision tree
- D) Uses neural network

**ANSWER: B**

**Q81. Why is KNN called "lazy learning"?**

- A) It's slow
- B) No explicit training phase - memorizes training data
- C) It doesn't learn anything
- D) Requires manual work

**ANSWER: B**

**Q82. What is the most common distance metric in KNN?**

- A) Manhattan distance
- B) Euclidean distance
- C) Cosine similarity
- D) Hamming distance

**ANSWER: B**

**Q83. How do you choose optimal K in KNN?**

- A) Always use  $K=1$
- B) Use cross-validation to find best K
- C) Always use  $K=100$
- D) Random selection

**ANSWER: B**

**Q84. What happens with very small K (e.g.,  $K=1$ )?**

- A) High bias, low variance
- B) Low bias, high variance - sensitive to noise
- C) Perfect performance
- D) No predictions possible

**ANSWER: B**

**Q85. What is a major limitation of KNN?**

- A) Too simple conceptually
- B) Slow prediction  $O(n)$ , memory-intensive, curse of dimensionality
- C) Always overfits
- D) Cannot handle classification

**ANSWER: B**

## LEVEL 6: DECISION TREES (Questions 86-100)

**Q86. How do decision trees make predictions?**

- A) Linear equations
- B) Recursive binary partitioning of feature space using if-then-else rules
- C) Distance calculations
- D) Probability distributions

**ANSWER: B**

**Q87. What does Gini impurity measure?**

- A) Tree depth
- B) Probability of incorrect classification (impurity of node)
- C) Number of splits
- D) Training time

**ANSWER: B**

**Q88. What is the Gini impurity range?**

- A)  $[0, 1]$
- B)  $[0, 0.5]$  where 0 = pure, 0.5 = maximally impure
- C)  $[-1, 1]$
- D)  $[0, \infty)$

**ANSWER: B**

**Q89. What does entropy measure in decision trees?**

- A) Tree size
- B) Disorder or uncertainty in the node
- C) Accuracy
- D) Speed

**ANSWER: B**

**Q90. What is Information Gain?**

- A) Total entropy
- B) Parent Entropy - Weighted Child Entropy (reduction in uncertainty)
- C) Tree depth
- D) Number of features

**ANSWER: B**

**Q91. What splitting criterion is sklearn's default for classification trees?**

- A) Entropy
- B) Gini impurity
- C) MSE
- D) MAE

**ANSWER: B**

**Q92. What splitting criterion is used for regression trees?**

- A) Gini impurity
- B) MSE (Mean Squared Error) or MAE
- C) Entropy
- D) Information Gain

**ANSWER: B**

**Q93. What is the max\_depth hyperparameter?**

- A) Maximum number of features

- B) Maximum depth (levels) the tree can grow
- C) Maximum samples per leaf
- D) Maximum training time

**ANSWER: B**

**Q94. What is pruning in decision trees?**

- A) Adding branches
- B) Removing branches to reduce overfitting
- C) Feature selection
- D) Data cleaning

**ANSWER: B**

**Q95. What are advantages of decision trees?**

- A) Always best accuracy
- B) Highly interpretable, no scaling needed, handles mixed types
- C) Never overfit
- D) Fastest training

**ANSWER: B**

**Q96. What are limitations of decision trees?**

- A) Too simple to use
- B) Very prone to overfitting (high variance), unstable
- C) Require too much data
- D) Cannot handle missing values

**ANSWER: B**

**Q97. What does "greedy algorithm" mean for decision trees?**

- A) Uses too much memory
- B) Makes locally optimal splits, not globally optimal
- C) Always finds best solution
- D) Requires all data at once

**ANSWER: B**

**Q98. Why are decision trees unstable?**

- A) Poor implementation
- B) Small data changes cause large tree differences
- C) Too stable actually
- D) Only for classification

**ANSWER: B**

**Q99. Do decision trees require feature scaling?**

- A) Yes, always required
- B) No, tree splits don't depend on scale
- C) Only sometimes
- D) Required for accuracy

**ANSWER: B**

**Q100. What prediction does a regression tree leaf node contain?**

- A) A linear equation
- B) The mean (or median) of training samples in that leaf
- C) The mode
- D) A probability distribution

**ANSWER: B**

**END OF SECOND 100 QUESTIONS (Pages 12+)**



**REMAINING CONTENT TO COVER (Would require additional questions):**

## **SCORE INTERPRETATION:**

90-100: Excellent understanding of metrics, algorithms, and concepts

80-89: Very good grasp of core ML concepts

70-79: Good foundation, review bias-variance and specific algorithms

60-69: Fair understanding, focus on metrics and regression concepts

Below 60: Review classification metrics and algorithm fundamentals

## **STUDY RECOMMENDATIONS:**