Machine Learning - 100 Questions

Pages 12+ Part 1

Nova IMS
Generated: January 16, 2026

LEVEL 1: CATEGORICAL ENCODING BASICS

Q1. Why do machine learning algorithms require numerical inputs?

A) For visual appeal

B) Algorithms perform mathematical operations requiring numbers

C) To save memory

D) No specific reason

Q2. What is label encoding?

A) Creating binary columns

B) Assigning integers to each category (0, 1, 2, ...)

C) Using category names as-is

D) Removing categories

Q3. What problem does label encoding create for nominal variables?

A) Too slow

B) Implies ordinal relationship when none exists

C) Uses too much memory

D) Always accurate

Q4. When is label encoding appropriate?

A) Always for all categorical variables

B) For ordinal variables or tree-based models

C) Never in any situation

D) Only for binary variables

Q5. What is one-hot encoding?

A) Assigning integers to categories

B) Creating binary column for each category

C) Removing categories

D) Using category frequencies

Q6. What is the main advantage of one-hot encoding?

A) Creates fewer features

B) No ordinal assumption for nominal data

C) Faster computation

D) Works only with trees

Q7. What is the main disadvantage of one-hot encoding?

A) Too slow to compute

B) Increases dimensionality (creates many columns)

C) Only works for binary variables

D) Requires target variable

Q8. For how many categories is one-hot encoding typically appropriate?

A) Unlimited categories

B) Few categories (typically <15-20)

C) Only 2 categories

D) 100+ categories

Q9. What is the "dummy variable trap" in one-hot encoding?

A) Using too few categories

B) Multicollinearity when using all dummy variables

C) Missing data problem

D) Outlier detection issue

Q10. How do you avoid the dummy variable trap?

A) Use all dummy variables

B) Drop one dummy variable (first or last)

C) Never use one-hot encoding

D) Add more categories

Q11. What is target encoding (mean encoding)?

A) Encoding the target as 0 and 1

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

C) Creating binary columns

D) Using category counts

Q12. What is the main advantage of target encoding?

A) Simplest to implement

B) Captures relationship with target (high information)

C) Never causes problems

D) Works without target variable

Q13. What is the main risk of target encoding?

A) Too slow

B) Target leakage and overfitting

C) Reduces accuracy

D) Only works for regression

Q14. What is frequency encoding?

A) Encoding based on alphabetical order

B) Replacing category with its frequency/count

C) Creating binary columns

D) Using target mean

Q15. What is binary encoding used for?

A) Only binary variables

B) High-cardinality features (converts to binary code)

C) Continuous variables

D) Ordinal variables only

LEVEL 2: FEATURE ENGINEERING & TRANSFORMATION

Q16. What is feature engineering?

A) Selecting existing features

B) Creating new features from existing data using domain knowledge

C) Removing outliers

D) Scaling features

Q17. What can be extracted from date/time features?

A) Nothing useful

B) Year, month, day, day_of_week, hour, is_weekend, season

C) Only the year

D) Only the day

Q18. What features can be extracted from text data?

A) Nothing quantifiable

B) Length, word count, sentiment score, TF-IDF vectors

C) Only word count

D) Only length

Q19. What are ratio features in feature engineering?

A) Random ratios

B) Meaningful ratios like debt_to_income, price_per_sqft

C) All possible ratios

D) Only percentages

Q20. What are aggregation features?

A) Random groupings

B) Group-based statistics (mean, median, std by category)

C) Sum of all features

D) Only counts

Q21. What are interaction features?

A) Random combinations

B) Multiplying or combining features to capture relationships

C) Adding all features

D) Deleting features

Q22. What is polynomial feature transformation?

A) Only linear features

B) Creates polynomial combinations like x^2, x^3, x■×x■

C) Removes features

D) Sorts features

Q23. What is a disadvantage of polynomial features?

A) Too simple

B) Dramatically increases dimensionality

C) Always improves accuracy

D) Only works for regression

Q24. What is binning (discretization)?

A) Removing features

B) Converting continuous to categorical ranges

C) Only for categorical data

D) Scaling method

Q25. When is log transformation useful?

A) For all data types

B) Right-skewed distributions and reducing outlier impact

C) Only for negative values

D) Normal distributions

Q26. What does log transformation do to multiplicative relationships?

A) Removes them

B) Converts to additive relationships

C) Makes them exponential

D) No effect

Q27. What is Box-Cox transformation?

A) A fixed transformation

B) Automatically finds optimal power transformation

C) Only for categorical data

D) A feature selection method

Q28. What limitation does Box-Cox have?

A) Too complex

B) Only for strictly positive values

C) Only for negative values

D) Works for all data

Q29. What is Yeo-Johnson transformation?

A) Same as Box-Cox

B) Similar to Box-Cox but handles zero and negative values

C) Only for positive values

D) A scaling method

Q30. What does automated feature engineering using Featuretools do?

A) Manual feature creation

B) Automatic deep feature synthesis

C) Only scales features

D) Removes features

Q31. What role does domain knowledge play in feature engineering?

A) Not important

B) Essential for identifying meaningful interactions

C) Only needed for simple problems

D) Can be completely ignored

Q32. What is the BMI formula as an example of feature engineering?

A) weight × height

B) weight / height^2

C) weight + height

D) weight - height

Q33. What are lagged features in time series?

A) Future values

B) Previous values used as predictors

C) Random values

D) Mean values

Q34. What are rolling window features?

A) Fixed values

B) Statistics computed over moving time windows

C) Only for images

D) Random samples

Q35. Why is feature engineering considered both art and science?

A) It's completely random

B) Requires creativity and domain knowledge plus technical skill

C) Only uses algorithms

D) No creativity needed

LEVEL 3: FEATURE SELECTION METHODS

Q36. Why is feature selection important?

A) Makes data look better

B) Reduces dimensionality, overfitting, and improves performance

C) Always worse for models

D) Only for visualization

Q37. What are the three main categories of feature selection methods?

A) Fast, medium, slow

B) Filter, Wrapper, Embedded methods

C) Simple, medium, complex

D) Linear, nonlinear, mixed

Q38. What are filter methods in feature selection?

A) Model-dependent methods

B) Model-independent statistical tests

C) Only for neural networks

D) Require training models

Q39. What does variance threshold do?

A) Adds variance

B) Removes low-variance (near-constant) features

C) Increases features

D) Scales features

Q40. Why remove low-variance features?

A) They're too important

B) Near-constant values provide little information

C) To increase complexity

D) Required by law

Q41. What does correlation analysis identify?

A) Independent features

B) Highly correlated features (multicollinearity)

C) Important features only

D) Outliers

Q42. What is the chi-squared test used for in feature selection?

A) Numerical features only

B) Categorical features with categorical target

C) Only continuous variables

D) Model training

Q43. What is the ANOVA F-test used for?

A) Categorical features only

B) Numerical features with categorical target

C) Only for regression

D) Only for trees

Q44. What does mutual information measure?

A) Feature size

B) Dependency between features and target

C) Only linear relationships

D) Feature count

Q45. What is an advantage of filter methods?

A) Considers interactions

B) Fast, model-agnostic, no overfitting risk

C) Always most accurate

D) Requires many models

Q46. What is a disadvantage of filter methods?

A) Too slow

B) Ignores feature interactions (univariate)

C) Always causes overfitting

D) Requires target variable

Q47. What are wrapper methods?

A) Model-independent methods

B) Model-dependent methods that search feature subsets

C) Only statistical tests

D) Scaling methods

Q48. What is forward selection?

A) Remove all features

B) Start empty, iteratively add best feature

C) Start with all features

D) Random selection

Q49. What is backward elimination?

A) Add features one by one

B) Start with all, iteratively remove worst feature

C) Random removal

D) Keep all features

Q50. What is Recursive Feature Elimination (RFE)?

A) Adds features recursively

B) Recursively removes least important features

C) Random selection

D) One-time removal

Q51. What is an advantage of wrapper methods?

A) Very fast

B) Considers feature interactions

C) Never overfits

D) No model needed

Q52. What is a disadvantage of wrapper methods?

A) Too simple

B) Computationally expensive, overfitting risk

C) Always fast

D) Never accurate

Q53. What are embedded methods?

A) Separate from training

B) Built-in feature selection during model training

C) Only filter methods

D) Manual selection

Q54. How does Lasso (L1) perform feature selection?

A) Keeps all features equally

B) Drives some coefficients to exactly zero

C) Only scales features

D) Adds features

Q55. How do tree-based models provide feature importance?

A) Random scores

B) Based on split improvement and usage frequency

C) All features equal

D) External calculation

LEVEL 4: DIMENSIONALITY REDUCTION & TRAIN/TEST SPLITTING

Q56. What is the difference between feature selection and feature extraction?

A) No difference

B) Selection keeps original features, extraction creates new ones

C) Selection always better

D) Extraction always better

Q57. What does PCA (Principal Component Analysis) do?

A) Selects best features

B) Creates new uncorrelated features capturing maximum variance

C) Removes outliers

D) Scales features

Q58. Are PCA components the same as original features?

A) Yes, exactly the same

B) No, linear combinations of original features

C) Only sometimes

D) Only for small datasets

Q59. What does LDA (Linear Discriminant Analysis) maximize?

A) Variance only

B) Class separability

C) Feature count

D) Training speed

Q60. What is t-SNE used for?

A) Model training

B) Non-linear dimensionality reduction for visualization

C) Feature scaling

D) Outlier detection

Q61. What is UMAP?

A) Same as PCA

B) Faster than t-SNE, preserves global structure

C) Only for categorical data

D) A classification algorithm

Q62. What are autoencoders?

A) Manual encoders

B) Neural network-based dimensionality reduction

C) Only for images

D) Statistical methods

Q63. Why split data into training and test sets?

A) To increase data volume

B) To evaluate model performance on unseen data

C) To speed up training

D) Required by Python

Q64. What is the typical train/test split ratio?

A) 50/50

B) 80/20 or 70/30

C) 95/5

D) 99/1

Q65. What is stratified sampling?

A) Random sampling only

B) Maintains class distribution in all splits

C) Only for large datasets

D) Removes rare classes

Q66. When is stratified sampling essential?

A) Always, for all problems

B) For imbalanced classification datasets

C) Never needed

D) Only for regression

Q67. What is a train/validation/test split?

A) Two-way split

B) Three-way split (e.g., 60/20/20) with validation for tuning

C) No splitting needed

D) Only train and test

Q68. What is the validation set used for?

A) Final evaluation

B) Hyperparameter tuning

C) Initial training

D) Data cleaning

Q69. What is the test set used for?

A) Hyperparameter tuning

B) Final unbiased evaluation (use only once)

C) Training the model

D) Feature selection

Q70. What is time series splitting?

A) Random splitting

B) Sequential validation respecting temporal order

C) Any split method

D) Only for images

LEVEL 5: CROSS-VALIDATION TECHNIQUES

Q71. What is cross-validation?

A) Single train/test split

B) Multiple train/test splits for robust performance estimates

C) Only for final testing

D) Data cleaning method

Q72. What is K-Fold cross-validation?

A) One split

B) Split data into K equal folds, train on K-1, test on 1, repeat K times

C) Random splits

D) No splitting

Q73. What are typical values for K in K-Fold CV?

A) K = 2

B) K = 5 or K = 10

C) K = 100

D) K = 1

Q74. How is the final cross-validation score calculated?

A) Use only best fold

B) Average performance across all K folds

C) Use only worst fold

D) Random selection

Q75. What is stratified K-Fold?

A) Same as regular K-Fold

B) Maintains class distribution in each fold

C) Only for regression

D) Random folding

Q76. When is stratified K-Fold essential?

A) Never needed

B) For imbalanced classification problems

C) Only for large datasets

D) All regression problems

Q77. What is Leave-One-Out Cross-Validation (LOOCV)?

A) Leave out 50% of data

B) K = n (each sample is one fold)

C) Random leaving out

D) No leaving out

Q78. When is LOOCV appropriate?

A) Large datasets (10,000+ samples)

B) Very small datasets (<100 samples)

C) Always the best choice

D) Never appropriate

Q79. What is repeated K-Fold CV?

A) Single K-Fold run

B) Repeat K-Fold multiple times with different random splits

C) No repetition

D) Only two repetitions

Q80. What is time series cross-validation?

A) Random CV

B) Forward-chaining with progressively increasing training window

C) Standard K-Fold

D) No CV for time series

Q81. Why can't you use standard K-Fold for time series?

A) Too slow

B) Would use future data in training (look-ahead bias)

C) Works perfectly fine

D) Only technical limitation

Q82. What is Group K-Fold?

A) Random grouping

B) Ensures groups (e.g., patients) not split across folds

C) Only for classification

D) Same as stratified

Q83. When is Group K-Fold needed?

A) Never

B) When multiple samples from same source (prevent leakage)

C) Always for all problems

D) Only for regression

Q84. What is nested cross-validation?

A) Single CV loop

B) Outer loop for evaluation, inner loop for hyperparameter tuning

C) No nesting

D) Only for neural networks

Q85. Why is nested CV considered the gold standard?

A) Fastest method

B) Provides unbiased performance estimate

C) Easiest to implement

D) Requires no computation

LEVEL 6: CLASSIFICATION METRICS - CONFUSION MATRIX

Q86. What does a confusion matrix show?

A) Feature correlations

B) Actual vs predicted classifications

C) Training time

D) Feature importance

Q87. What are True Positives (TP)?

A) Predicted positive, actually negative

B) Predicted positive, actually positive (correct)

C) Predicted negative, actually positive

D) Predicted negative, actually negative

Q88. What are False Positives (FP)?

A) Predicted positive, actually positive

B) Predicted positive, actually negative (Type I Error)

C) Predicted negative, actually positive

D) Predicted negative, actually negative

Q89. What are False Negatives (FN)?

A) Predicted positive, actually negative

B) Predicted negative, actually negative

C) Predicted negative, actually positive (Type II Error)

D) Predicted positive, actually positive

Q90. What are True Negatives (TN)?

A) Predicted positive, actually negative

B) Predicted negative, actually positive

C) Predicted positive, actually positive

D) Predicted negative, actually negative (correct)

Q91. What is a Type I Error?

A) False Negative

B) False Positive

C) True Positive

D) True Negative

Q92. What is a Type II Error?

A) False Positive

B) False Negative

C) True Positive

D) True Negative

Q93. What is accuracy in classification?

A) TP / (TP + FP)

B) (TP + TN) / (TP + TN + FP + FN)

C) TP / (TP + FN)

D) TN / (TN + FP)

Q94. Why is accuracy misleading for imbalanced datasets?

A) It's always accurate

B) Can achieve high accuracy by predicting majority class

C) Only works for balanced data technically

D) Too complex to calculate

Q95. If 95% of emails are not spam, what accuracy can you get by always predicting "not spam"?

A) 50%

B) 95% (without learning anything useful)

C) 5%

D) 0%

Q96. What is precision in classification?

A) TP / (TP + FN)

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

C) TN / (TN + FP)

D) (TP + TN) / Total

Q97. What question does precision answer?

A) Of actual positives, how many did we catch?

B) When model says positive, how often is it right?

C) What's the overall accuracy?

D) How many negatives are there?

Q98. What is recall (sensitivity)?

A) TP / (TP + FP)

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

C) TN / (TN + FP)

D) (TP + TN) / Total

Q99. What question does recall answer?

A) When model says positive, how often right?

B) Of all actual positive cases, how many did model catch?

C) What's the overall accuracy?

D) How many false positives?

Q100. Which metric focuses on minimizing false positives?

A) Recall

B) Precision

C) Accuracy

D) Specificity