

Machine Learning Fall 2024

School of Economics, Quaid-i-Azam University, Islamabad

2025-01-31

Encircle/Cross the correct answer.

1. Which of the following is NOT a resampling method?

- a) Cross-validation
- b) Bootstrapping
- c) Ridge Regression
- d) Leave-One-Out Cross-Validation (LOOCV)

2. What is the primary purpose of cross-validation?

- a) To improve the training accuracy of a model
- b) To estimate the prediction error of a model
- c) To increase the number of training samples
- d) To test different hyperparameter values

3. In k-fold cross-validation, how is the data divided?

- a) Into k random subsets where each subset is used once as a validation set
- b) Into two halves, one for training and one for testing
- c) Into a single training set with multiple iterations of random sampling
- d) Into multiple training sets but with no validation set

4. What is the key advantage of Leave-One-Out Cross-Validation (LOOCV) over k-fold cross-validation?

- a) It is computationally faster
- b) It provides an unbiased estimate of the test error
- c) It requires fewer computations
- d) It reduces the bias in parameter estimation

5. What is the main drawback of using LOOCV compared to k-fold cross-validation?

- a) It provides a biased estimate of test error
- b) It uses too little data for training
- c) It is computationally expensive
- d) It cannot be used for non-linear models

6. The bootstrap method is primarily used to:

- a) Reduce overfitting in machine learning models
- b) Estimate the accuracy of a statistic by resampling with replacement
- c) Improve model accuracy by adding more data
- d) Reduce the number of predictor variables

7. When performing the bootstrap, how is the dataset resampled?

•

- a) By randomly selecting data points without replacement
- b) By dividing the data into k-folds and selecting one for validation
- c) By randomly selecting data points with replacement
- d) By using 80% of the data for training and 20% for validation

8. In k-fold cross-validation, what happens when k increases?

- a) The computational cost increases
- b) The training set size decreases
- c) The variance of the estimate decreases
- d) The bias of the estimate increases

9. Which of the following is true about bias-variance tradeoff in cross-validation?

- a) LOOCV has lower variance but higher bias compared to 10-fold CV
- b) LOOCV has higher variance but lower bias compared to 10-fold CV
- c) 10-fold CV has higher variance and bias than LOOCV
- d) Both LOOCV and 10-fold CV have the same bias and variance

10. The bootstrap is particularly useful when:

- a) The dataset is very large
- b) The dataset is small and obtaining additional data is expensive
- c) There are no missing values in the dataset
- d) The model has a high number of predictor variables

11. Which of the following is NOT a reason for using subset selection in regression models?

- a) Reducing model complexity
- b) Improving prediction accuracy
- c) Handling missing values
- d) Enhancing interpretability

12. In best subset selection, how are models compared to select the best one?

- a) By using the Adjusted R^2 or C_p statistic
- b) By selecting the model with the highest training accuracy
- c) By using only cross-validation
- d) By minimizing the number of predictor variables

13. What is the primary difference between forward and backward stepwise selection?

- a) Forward selection starts with no predictors and adds one at a time, while backward selection starts with all predictors and removes one at a time
- b) Forward selection is computationally expensive, while backward selection is not
- c) Backward selection can handle categorical predictors, while forward selection cannot
- d) Forward selection can only be used for classification problems

14. Why is Ridge Regression preferred over ordinary least squares (OLS) when dealing with multicollinearity?

- a) It forces some coefficients to be exactly zero
- b) It reduces variance by introducing a penalty on large coefficients
- c) It increases bias and variance equally
- d) It performs subset selection automatically

15. What is the primary difference between Ridge Regression and Lasso Regression?

- a) Ridge Regression allows some coefficients to be exactly zero, while Lasso does not
- b) Lasso forces some coefficients to be exactly zero, performing variable selection
- c) Ridge Regression is only used for classification, while Lasso is for regression
- d) Lasso Regression increases the number of predictors in the model

16. The tuning parameter λ in Ridge and Lasso regression controls:

- a) The number of predictor variables included in the model
- b) The bias-variance tradeoff
- c) The number of observations used for training
- d) The maximum likelihood estimation

17. What happens when $\lambda = 0$ in Ridge Regression?

- a) The Ridge model performs the same as OLS regression
- b) All coefficients are forced to zero
- c) The model performs best with no variance
- d) The Ridge penalty increases exponentially

18. Which of the following methods can be used to determine the optimal value of λ in Ridge and Lasso Regression?

- a) Adjusted R^2
- b) Akaike Information Criterion (AIC)
- c) Cross-validation
- d) Using the F-statistic

19. Why is the Elastic Net method useful?

- a) It combines Ridge and Lasso to balance feature selection and shrinkage

- b) It only selects the most important variable in the model
- c) It works only with categorical variables
- d) It is a faster alternative to both Ridge and Lasso

20. When using Lasso Regression, what happens to correlated predictor variables?

- a) Both variables are retained with reduced coefficients
- b) One variable is usually shrunk to zero while the other remains in the model
- c) Lasso increases the coefficients of correlated variables
- d) Lasso does not work if predictor variables are correlated

21. What is the primary advantage of decision trees over linear regression?

- a) They can model non-linear relationships and interactions between variables
- b) They always provide better predictions than linear regression
- c) They require large amounts of data to work effectively
- d) They are more computationally efficient than regression models

22. In a regression tree, how is the response variable predicted for a new observation?

- a) By averaging all the target values in the training set
- b) By using a weighted sum of predictor variables
- c) By taking the mean of the target values in the terminal node where the observation falls
- d) By fitting a linear model to each subset of the data

23. What is the criterion used to split nodes in regression trees?

- a) Gini impurity
- b) Entropy
- c) Residual sum of squares (RSS)
- d) Maximum likelihood estimation

24. Which of the following is a major drawback of deep decision trees?

- a) They cannot handle categorical variables
- b) They tend to overfit the training data
- c) They do not work well with missing data
- d) They cannot be used for classification tasks

25. What is the purpose of pruning a decision tree?

- a) To increase model complexity and improve training accuracy
- b) To reduce overfitting and improve generalization to test data
- c) To decrease bias by adding more splits to the tree
- d) To remove categorical variables from the dataset

26. In classification trees, what is commonly used as the splitting criterion?

- a) Sum of squared errors

- b) Gini impurity or entropy
- c) Adjusted R^2
- d) Cross-validation error

27. How does cross-validation help in tree-based methods?

- a) It helps in selecting the best hyperparameters, such as the tree depth
- b) It reduces computational complexity
- c) It eliminates the need for pruning
- d) It always improves the training accuracy

28. Why does a classification tree with too many splits perform poorly on new data?

- a) It increases bias in the model
- b) It does not capture interactions between variables
- c) It memorizes the training data and does not generalize well (overfitting)
- d) It cannot handle non-linear relationships

29. In a decision tree, which of the following measures is NOT commonly used to evaluate split quality for classification problems?

- a) Gini impurity
- b) Entropy
- c) Residual sum of squares (RSS)
- d) Misclassification error

30. How does the Gini impurity metric behave when a node is pure?

- a) It takes the highest possible value
- b) It equals 0
- c) It becomes equal to the entropy
- d) It becomes negative

Part 2

Give brief answer of each of the following :

1. How does cross-validation help estimate model performance, OR why is bootstrapping useful for measuring variability?

2. Why does ridge regression shrink coefficients instead of setting them to zero, OR how does the Lasso regression perform variable selection?

3. Why does a regression tree split based on minimizing variance, OR how does a classification tree decide the best split criterion?