Machine Learning

1) C) Both are equally proficient
2) A) max_depth
3) A) SMOTE
4) C) 1 and 3
5) D) 1-3-2
6) A) Decision Trees
7) C) CART can only create binary trees (a maximum of two children for a node), and CHAID can create multiway trees (more than two children for a node)

8)

A) Ridge will lead to some of the coefficients to be very close to 0
B) Lasso will lead to some of the coefficients to be very close to 0

9) B) remove only one of the features

C) Use ridge regularization

10) A) Overfitting

11) One-hot encoding should be avoided when the number of unique categories in a categorical variable is very high, as it can lead to the curse of dimensionality. In such situations, an alternative encoding technique called target encoding or mean encoding can be used. Target encoding replaces each category with the mean of the target variable for that category

12) In the case of data imbalance in classification, where one class has significantly fewer observations than another

Resampling: Resampling is a technique used to either increase or decrease the number of instances in a dataset.

1) Oversampling
2) Undersampling

13) SMOTE and ADASYN are both oversampling techniques used to address the problem of imbalanced datasets

14) GridSearchCV is a technique used in machine learning to perform hyperparameter tuning by searching through a predefined grid of hyperparameters for the best combination that maximizes a specified evaluation metric

GridSearchCV is not preferred to use in case of large datasets because it can be computationally expensive and time-consuming

In such cases, alternative techniques such as randomized search or Bayesian optimization can be used

15) Regression models are used to predict a continuous output variable and To evaluate the performance of a regression model, several evaluation metrics can be used

1) Mean Absolute Error
2) Mean Squared Error
3) Root Mean Squared Error
4) R-squared
5) Adjusted R-squared
6) Mean Absolute Percentage Error
7) Mean Percentage Error

Stats

1) b. The probability of failing to reject H_0 when H_1 is true
2) b. null hypothesis
3) d. Type I error
4) b. the t distribution with n - 1 degrees of freedom
5) d. rejecting Ho when it is true
6) d. a two-tailed test
7) b. the probability of committing a Type I error
8) a. the probability of committing a Type II error
9) a. $z > z\alpha$
10) c. the level of significance
11) a. level of significance
12) a. Degrees of Freedom

13)

ANOVA (Analysis of Variance) is a statistical technique used to compare the means of two or more groups.
In SPSS, ANOVA can be performed using the "General Linear Model" command in the "Analyze" menu.

14)

Independence: The observations within each group should be independent of each other.

Normality: The data within each group should follow a normal distribution. Violation of this assumption may cause the test to be less powerful or may lead to Type I errors.

Homogeneity of variance: The variance of the data within each group should be equal. This assumption is also called homoscedasticity. Violation of this assumption may cause the test to be less powerful or may lead to Type I errors.

Random sampling: The groups should be formed by randomly sampling from the population

15) One-way ANOVA is used when there is one independent variable with two or more levels. Two-way ANOVA is used when there are two independent variables, also called factors, and each independent variable has two or more levels