

## 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. B) Support Vector Machines
7. B) CART can create multiway trees (more than two children for a node), and CHAID can only create binary trees (a maximum of two children for a node).
8. D) Lasso will cause some of the coefficients to become 0, A) Ridge will lead to some of the coefficients to be very close to 0
9. B) remove only one of the features, D) use Lasso regularization
10. A) Overfitting, C) Underfitting
11. For a feature having a large number of unique feature values or categories one-hot encoding must be avoided. Label encoding can be used in such cases.
12. When we are using an imbalanced dataset, we can oversample the minority class using replacement. This technique is called oversampling. Similarly, we can randomly delete rows from the majority class to match them with the minority class which is called undersampling.
13. The key difference between ADASYN and SMOTE is that the former uses a density distribution.
14. GridSearchCV is a technique for finding the optimal parameter values from a given set of parameters in a grid. It's essentially a cross-validation technique. The model as well as the parameters must be entered. After extracting the best parameter values, predictions are made.
15. The evaluation metric used to evaluate a regression model are:
  - a. Mean Squared Error – It is an important loss function for algorithms fit or optimized using the least squares framing of a regression problem.
  - b. Root Mean Squared Error – The Root Mean Squared Error is an extension of the Mean Squared Error.
  - c. Mean Absolute Error – It is an error that the units of error score match the units of the target value that is being predicted.

## STATISTICS WORKSHEET-8

1. a. The probability of rejecting  $H_0$  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. a. accepting  $H_0$  when it is false
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. b.  $z < z_\alpha$
10. c. the level of significance
11. a. level of significance
12. d. All of the Above
13. ANOVA is the way to find out if experimental results are significant. One-way analysis of variance involves one independent variable which has a number of different levels. The dependent variable is a continuous variable.
14. There are three primary assumptions in ANOVA:
  - a. The responses for each factor level have a normal population distribution.
  - b. These distributions have the same variance.
  - c. The data are independent.
15. The only difference between one-way and two-way ANOVA is the number of independent variables. A one-way ANOVA has one independent variable, while a two-way ANOVA has two.