**Machine Learning**

1. What is the advantage of hierarchical clustering over K-means clustering?

Ans: B) In hierarchical clustering you don’t need to assign number of clusters in beginning

2. Which of the following hyper parameter(s), when increased may cause random forest to over fit the data?

Ans: A) max\_depth

3. Which of the following is the least preferable resampling method in handling imbalance datasets?

Ans: C) RandomUnderSampler

4. Which of the following statements is/are true about “Type-1” and “Type-2” errors?

Ans: C) 1 and 3

5. Arrange the steps of k-means algorithm in the order in which they occur:

Ans: D) 1-3-2

6. Which of the following algorithms is not advisable to use when you have limited CPU resources and time, and when the data set is relatively large?

Ans: C) K-Nearest Neighbors

7. What is the main difference between CART (Classification and Regression Trees) and CHAID (Chi Square Automatic Interaction Detection) Trees?

Ans: 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. In Ridge and Lasso regularization if you take a large value of regularization constant(lambda), which of the following things may occur?

Ans: A) Ridge will lead to some of the coefficients to be very close to 0

C) Ridge will cause some of the coefficients to become 0

9. Which of the following methods can be used to treat two multi-collinear features?

Ans: B) remove only one of the features

D) use Lasso regularization

10. After using linear regression, we find that the bias is very low, while the variance is very high. What are the possible reasons for this?

Ans: A) Overfitting

B) Multicollinearity

D) Outliers

11. In which situation One-hot encoding must be avoided? Which encoding technique can be used in such a case?

Ans: One hot encoding is not advisable in case the variable to be encoded has very high or very low number of categories. In case of high number of categories (say 15), one hot encoding will create 14 new features if we drop the first feature, which will result in low feature importance of each of these 14 features. And in case of low number of categories (say 3 or 4), one hot encoding will create 2 to 3 number of features when drop first is true, this will result in highly correlated features and while building the model some of these features will be eliminated even if they are important.To avoid these problems one can use frequency encoding or mean encoding. In frequency encoding each of the category is coded based on its frequency and in mean it is encoded as per mean.

12. In case of data imbalance problem in classification, what techniques can be used to balance the dataset? Explain them briefly.

Ans: In case of data imbalance, we can either under sample the major class or over sample the minor class. under sampling is often unadvisable because while under sampling the problem of data loss occurs, which further results in less generic models. Also, data is quite expensive and must not be wasted. Following are some methods for dealing with data imbalance.

RANDOM UNDERSAMPLER: In random undersampler, the data points of major class are dropped randomly so that the data points of major class in sample becomes equal to data points of minor class. This method has a problem of data loss associated with it as most of the data is dropped while fitting it to dataset. Data is expensive and it is the key to a data science algorithm. Therefore, it is not preferred in actual practice.

RANDOM OVERSAMPLER: Random Oversampler makes random replications of data points of minor class by multiplying them to a random weight so as to make the number of data points of minor class become equal to data points of major class. It is less preferred over SMOTE and ADASYN because, it merely makes duplication of data and hence the model built has no added value.

SMOTE (Synthetic Minority Oversampling Technique): Synthetic Minority Oversampling Technique works by selecting examples that are close in the feature space, drawing a line between the examples in the feature space and drawing a new sample at a point along that line. Specifically, a random example from the minority class is first chosen. Then k of the nearest neighbors for that example are found (typically k=5). A randomly selected neighbor is chosen and a synthetic example is created at a randomly selected point between the two examples in feature space.

ADASYN (Adaptive Synthetic): ADASYN also generates synthetic samples as done by SMOTE. But it takes care of the distribution of the data points. It generates synthetic samples where the crowding of the minority samples is comparatively less. In this case the final dataset has similar distribution of all the samples in the feature space.

13. What is the difference between SMOTE and ADASYN sampling techniques?

Ans:

SMOTE (Synthetic Minority Oversampling Technique): Synthetic Minority Oversampling Technique works by selecting examples that are close in the feature space, drawing a line between the examples in the feature space and drawing a new sample at a point along that line. Specifically, a random example from the minority class is first chosen. Then k of the nearest neighbors for that example are found (typically k=5). A randomly selected neighbor is chosen and a synthetic example is created at a randomly selected point between the two examples in feature space.

ADASYN (Adaptive Synthetic): ADASYN also generates synthetic samples as done by SMOTE. But it takes care of the distribution of the data points. It generates synthetic samples where the crowding of the minority samples is comparatively less. In this case the final dataset has similar distribution of all the samples in the feature space.

14. What is the purpose of using GridSearchCV? Is it preferable to use in case of large datasets? Why or why not?

Ans: Grid search CV stands for Grid Search Cross Validation. It is used for hyper parameter tuning the machine learning models. In grid search cv we make a params grid which is a dictionary containing different values of each parameter we want to tune. We also provide number of folds of cross validation and the ML model on which the grid search is being performed. Grid search fits all the possible combinations of each and every parameter mentioned in params grid with number of folds of cross validation. This gives the best possible score and hence the best parameters for the model to be fit. In this way we get the best performing set and we build our final model on these parameters. Gridsearch CV is disadvantageous over large datasets as it does a large number of fits which takes a lot of time and is computationally very heavy. But it is very helpful tool if the dataset is not very large.

15. List down some of the evaluation metric used to evaluate a regression model. Explain each of them in brief.

Ans: The evaluation metrics used to evaluate a regression model are as follows:

R2 Score: Also known as coefficient of determination. R2 signifies that what fraction of total variance is explained by the best fit line.

Adjusted R2 Score: Adjusted R2 and R2 both represent that how well the model fits the data points. But adjusted R2 penalizes the model for using more features. In case we increase the number of features in training data the R2 will increase but adjusted R2 will only increase if the new feature adds value to our model. Due to this reason adjusted R2 is considered as a better evaluation metric than R2. Adjusted R2 is always less than or equal to R2.

Mean Squared Error (MSE): Mean Squared error is the mean of all the squares of the distances of each data point from its predictions by the best fit line

Root Mean Squared Error (RMSE): RMSE is the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are; RMSE is a measure of how spread out these residuals are. In other words, it tells you how concentrated the data is around the line of best fit. It is nothing but square root of MSE.

**Python**

1. Which of the following operators is used to calculate remainder in a division?

Ans: C) %

2. In python 2//3 is equal to?

Ans: B) 0

3. In python, 6<<2 is equal to?

Ans: B) 10

4. In python, 6&2 will give which of the following as output?

Ans: A) 2

5. In python, 6|2 will give which of the following as output?

Ans: D) 6

6. What does the finally keyword denotes in python?

Ans: C) the finally block will be executed no matter if the try block raises an error or not.

7. What does raise keyword is used for in python?

Ans: A) It is used to raise an exception.

8. Which of the following is a common use case of yield keyword in python?

Ans: C) in defining a generator

9. Which of the following are the valid variable names?

Ans: A) \_abc, C) abc2

10. Which of the following are the keywords in python?

Ans: A) yield, B) raise

**Statistics**

1. In hypothesis testing, type II error is represented by β and the power of the test is 1−β then β is:

Ans: a. The probability of rejecting H0 when H1 is true

2. In hypothesis testing, the hypothesis which is tentatively assumed to be true is called the

Ans: b. null hypothesis

3. When the null hypothesis has been true, but the sample information has resulted in the rejection of the null, a \_\_\_\_\_\_\_\_\_ has been made

Ans: d. Type I error

4.For finding the p-value when the population standard deviation is unknown, if it is reasonable to assume that the population is normal, we use

Ans: b. the t distribution with n - 1 degrees of freedom

5. A Type II error is the error of

Ans: a. accepting Ho when it is false

6. A hypothesis test in which rejection of the null hypothesis occurs for values of the point estimator in either tail of the sampling distribution is called

Ans: d. a two-tailed test

7. In hypothesis testing, the level of significance is

Ans: b. the probability of committing a Type I error

8. In hypothesis testing, b is

Ans: a. the probability of committing a Type II error

9. When testing the following hypotheses at an α level of significance

H0: p = 0.7

H1: p > 0.7

Ans: a. z > zα

10. Which of the following does not need to be known in order to compute the P-value?

Ans: c. the level of significance

11. The maximum probability of a Type I error that the decision maker will tolerate is called the

Ans: a. level of significance

12. For t distribution, increasing the sample size, the effect will be on

Ans: d. All of the Above

13. What is Anova in SPSS?

Ans: The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of two or more independent (unrelated) groups (although you tend to only see it used when there are a minimum of three, rather than two groups).

14. What are the assumptions of Anova?

Ans: There are three primary assumptions in ANOVA: The responses for each factor level have a normal population distribution. These distributions have the same variance. The data are independent.

15. What is the difference between one way Anova and two way Anova?

Ans: A one-way ANOVA only involves one factor or independent variable, whereas there are two independent variables in a two-way ANOVA. In a one-way ANOVA, the one factor or independent variable analysed has three or more categorical groups. A two-way ANOVA instead compares multiple groups of two factors