MACHINE LEARNING WORKSHEET-8

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

A) Hierarchical clustering is computationally less expensive

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

C) Both are equally proficient D) None of these

ANS1.'B'

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

data?

A) max\_depth B) n\_estimators

C) min\_samples\_leaf D) min\_samples\_splits

ANS2.'B'

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

A) SMOTE B) RandomOverSampler

C) RandomUnderSampler D) ADASYN

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

1. Type1 is known as false positive and Type2 is known as false negative.

2. Type1 is known as false negative and Type2 is known as false positive.

3. Type1 error occurs when we reject a null hypothesis when it is actually true.

A) 1 and 2 B) 1 only

C) 1 and 3 D) 2 and 3

ANS4.'C'

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

1. Randomly selecting the cluster centroids

2. Updating the cluster centroids iteratively

3. Assigning the cluster points to their nearest center

A) 3-1-2 B) 2-1-3

C) 3-2-1 D) 1-3-2

ANS5.'D'

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?

A) Decision Trees B) Support Vector Machines

C) K-Nearest Neighbors D) Logistic Regression

ANS6.'B'

7. What is the main difference between CART (Classification and Regression Trees) and CHAID (Chi

Square Automatic Interaction Detection) Trees?

A) CART is used for classification, and CHAID is used for regression.

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).

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)

D) None of the above

ANS7.'C'

In Q8 to Q10, more than one options are correct, Choose all the correct options:

8. In Ridge and Lasso regularization if you take a large value of regularization constant(lambda), which

of the following things may occur?

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

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

D) Lasso will cause some of the coefficients to become 0

ANS8.'A

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

A) remove both features from the dataset

B) remove only one of the features

C) Use ridge regularization D) use Lasso regularization

ANS9.'D'

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?

A) Overfitting B) Multicollinearity

C) Underfitting D) Outliers

ANS10.'A'

Q10 to Q15 are subjective answer type questions, Answer them briefly.

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

such a case?

ANS11. It leads to dummy variable trap which ultimately leads to the problem of multicoliniarity

.Instead we can use label encoder

12. In case of data imbalance problem in classification, what techniques can be used to balance the

dataset? Explain them briefly.

Ans12.

1. Random Oversampling

2.Random Undersampling

3. SMOTE

4.ADASYN

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

ANS13.he key difference between ADASYN and SMOTE is that the former uses a density distribution, as a criterion to automatically decide the number of synthetic samples that must be generated for each minority sample by adaptively changing the weights of the different minority samples to compensate for the skewed distributions. The latter generates the same number of synthetic samples for each original minority sample.

14. What is the purpose of using GridSearchCV? Is it preferable to use in case of large datasets? Why or

why not?

ANS14.GridSearchCV is a library function that is a member of sklearn’s model\_selection package. It helps to loop through predefined hyperparameters and fit your estimator (model) on your training set. So, in the end, you can select the best parameters from the listed hyperparameters.

Yes it is prefered for large datasets as it helps in hyperparameter tuning

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

in brief

ANS15,

1. Mean absolute error

2. Mean squared error

3. R2 Score