MACHINE LEARNING

1. Which of the following in sk-learn library is used for hyper parameter tuning?

Ans -> A) GridSearchCV() , B) RandomizedCV()

2. In which of the below ensemble techniques trees are trained in parallel?

Ans -> A) Random forest

3. In machine learning, if in the below line of code

Ans -> A) The regularization will increase

4. Check the below line of code and answer the following questions:
sklearn.tree.DecisionTreeClassifier(*criterion='gini',splitter='best',max_depth=None, min_samples_split=2)

Ans -> C) both A & B

5. Which of the following is true regarding Random Forests?

Ans -> A) It's an ensemble of weak learners.

6. What can be the disadvantage if the learning rate is very high in gradient descent?

Ans -> C) Both of them

7. As the model complexity increases, what will happen?

Ans -> B) Bias will decrease, Variance increase

8. Suppose I have a linear regression model which is performing as follows

Ans -> A) model is underfitting

Q9 to Q15 are subjective answer type questions, Answer them briefly.

9. Suppose we have a dataset which have two classes A and B. The percentage of class A is 40% and percentage of class B is 60%. Calculate the Gini index and entropy of the dataset.

Ans ->

10. What are the advantages of Random Forests over Decision Tree?

- Ans -> A decision tree is more simple and interpretable but prone to overfitting, but a random forest is complex and prevents the risk of overfitting.

- Random forest is a more robust and generalized performance on new data, widely used in various domains such as finance, healthcare, and deep learning.

11. What is the need of scaling all numerical features in a dataset? Name any two techniques used for scaling.

Ans -> Scaling -> We are going to unitless all featue and we are going to standardize this data then we called a Scaler.

Suppose -> I have Four Friends and all friends are equally knowledgeble and one friends in very close to me .then i cannot listen him only remaining 3 are eqaully knowlegeble . i have to listen to all my four friends . i cannot biased .i have to listen all my friends .

Biased- We always try to listen one guy.

I have Four Friends and all friends are equally knowledgeable and one friends in very close to me .then i cannot listen him only remaining 3 are equally knowledgeable . i have to listen to all my four friends . i cannot biased .i have to listen all my friends .

Biased- We always try to listen one guy.

And we are going to apply scaler only in feature.

1-> Standard Scaler , 2- Min Max Scaler

12. Write down some advantages which scaling provides in optimization using gradient descent algorithm.

Ans- > Flexibility: Gradient Descent can be used with various cost functions and can handle non-linear regression problems

Scalability: Gradient Descent is scalable to large datasets since it updates the parameters for each training example one at a time.

Convergence: Gradient Descent can converge to the global minimum of the cost function, provided that the learning rate is set appropriately.

13 . In case of a highly imbalanced dataset for a classification problem, is accuracy a good metric to measure the performance of the model. If not, why?

Ans -> So, In any case if we have any imbalanced dataset there are 2 conditions.

1st -> Model is Performing Very Well -> Because Our Model is biased to a Particular Class. Because Our Dataset is Imbalanced .

2nd -> Model is Performing Very Bad -> Because Our dataset in imbalanced dataset .So, any Prediction is may be gone wrong to any class because of all classes is not given in same amount . So for that we have lots a technique Like SMOTE (Synthetic Minority OverSampling Technique)

14. What is “f-score" metric? Write its mathematical formula
Ans ->Both Precision and Recall for evaluating a model one such metric is the F1SCORE .

F1 Score is defined as the harmonic mean of precision and Recall.

Mathematical Formula is = 2*(Precision * Recall) / Precision +Recall

15. What is the difference between fit(), transform() and fit_transform()?

Ans -> The fit() method helps in fitting the data into a model, transform() method helps in transforming the data into a form that is more suitable for the model. Fit_transform() method, on the other hand, combines the functionalities of both fit() and transform() methods in one step