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

   Ans: D.All the above.

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*sklearn.svm.*SVC *(C=1.0, kernel='rbf', degree=3)* we increasing the C hyper parameter, what will happen?**

   Ans: B) The regularization will decrease

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 and B.

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

   Ans: A) It's an ensemble of weak learners.
   B) The component trees are trained in series
   C) In case of classification problem, the prediction is made by taking mode of the class labels predicted by the component trees

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

   Ans: A) Gradient Descent algorithm can diverge from the optimal solution.

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: Train accuracy=0.95 and Test accuracy=0.75 Which of the following is true regarding the model

Ans: A) model is underfitting

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.

Gini Index:

Probability of Class A=0.4/2=0.2, squared probability(A)=0.2*0.2=0.04

Probability of Class B=0.6/2=0.3, squared probability(B)=0.3*0.3=0.09

Sum of squared probabilities=0.04+0.09=0.13

Gini Index= 1-sum of squared probability of each class

So Gini Index=1-0.13=**0.87.**

The Gini Index or Gini Impurity is calculated by **subtracting the sum of the squared probabilities of each class from one**

Entropy:

Probability of Class A p(A)=0.4/2=0.2

Log of p(A)= -0.69

Probability of Class B p(B)=0.6/2=0.3

Log of P(B)= -0.52

Entropy = -((p(0) * log(P(0)) + p(1) * log(P(1)))

Entropy=-((0.2*-0.69)+(0.3*-0.52))

Entropy= -(-0.49+-0.22)=-(-0.71)=0.71

Entropy =0.71

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

Decision Tree Random Forest.

Always scope for overfitting	It avoids and prevent overfitting by using multiple trees.
Results are not accurate.	This Gives accurate and precise results.

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

In Data Processing, we try to change the data in such a way that the model can process it without any problems. And Feature Scaling is one such process in which we transform the data into a better version. Feature Scaling is done to normalize the features in the dataset into a finite range.

In standardization, we calculate the z-value for each of the data points and replaces those with these values.

$$X_{new} = \frac{X - X_{mean}}{\sigma}$$

In min-max you will subtract the minimum value in the dataset with all the values and then divide this by the range of the dataset(maximum-minimum). In this case, your dataset will lie between 0 and 1 in all cases

Xnew =x-xmin / xmax-xmin

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

To ensure that the gradient descent moves smoothly towards the minima and that the steps for gradient descent are updated at the same rate for all the features, we scale the data before feeding it to the model. Having features on a similar scale can help the gradient descent converge more quickly towards the minima

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?

Accuracy is not a good metric for imbalanced datasets.

This model would receive a very good accuracy score as it predicted correctly for the majority of observations, but this hides the true performance of the model which is objectively not good as it only predicts for one class.

14. What is "f-score" metric? Write its mathematical formula.

In statistical analysis of binary classification, the **F-score** or **F-measure** is a measure of a test's accuracy. It is calculated from the precision and recall of the test, where the precision is the number of true positive results divided by the number of all positive results, including those not identified correctly, and the recall is the number of true positive results divided by the number of all samples that should have been identified as positive. Precision is also known as positive predictive value, and recall is also known as sensitivity in diagnostic binary classification.

F Score= 2*(Precision*recall)/ precision+Recall)

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

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