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

ANS- All of the above

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

ANS-Random forest

1. 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-The regularization will decrease

1. 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)*

Which of the following is true regarding max\_depth hyper parameter?

ANS-It regularizes the decision tree by limiting the maximum depth up to which a tree can be grown.

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

ANS-It's an ensemble of weak learners.

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

ANS-Gradient Descent algorithm can diverge from the optimal solution.

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

ANS-Bias will decrease, Variance increase

1. 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-model is overfitting

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

1. 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-Let p\_A = 0.4 be the proportion of class A in the dataset, and p\_B = 0.6 be the proportion of class B in the dataset.

Gini index: The Gini index is a measure of impurity or diversity in a dataset. It is calculated as follows:

Gini = 1 - (p\_A^2 + p\_B^2)

Gini = 1 - (0.4^2 + 0.6^2)

Gini = 0.48

Entropy: Entropy is another measure of impurity or diversity in a dataset. It is calculated as follows:

Entropy = -p\_Alog2(p\_A) - p\_Blog2(p\_B)

Entropy = -0.4log2(0.4) - 0.6log2(0.6)

Entropy = 0.971

Therefore, the Gini index of the dataset is 0.48 and the entropy of the dataset is 0.971

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

ANS-Random forest algorithm avoids and prevents overfitting by using multiple trees. The results are not accurate. This gives accurate and precise results. Decision trees require low computation, thus reducing time to implement and carrying low accuracy.

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

ANS-The most common techniques of feature scaling are Normalization and Standardization. Normalization is used when we want to bound our values between two numbers, typically, between [0,1] or [-1,1]. While Standardization transforms the data to have zero mean and a variance of 1, they make our data unitless.

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

ANS- Following is the some advantages-

Improved Convergence

Avoidance of local minima

Improved Accuracy

Regularization

1. 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-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.

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

ANS-An F-score is the harmonic mean of a system's precision and recall values. It can be calculated by the following formula: 2 x [(Precision x Recall) / (Precision + Recall)].

1. 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