**MACHINE LEARNING**

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

Ans. A) GridSearchCV()

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

Ans. A) 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. A) The regularization will increase

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. D) None of the above

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

Ans. C) In case of classification problem, the prediction is made by taking mode of the class labels predicted by the component trees.

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

Ans. C) Both of them

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

Ans. C)both bias and 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. B) model is overfitting

1. 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. Calculating of entropy:

E (S)=

E= - 0.4 log 0.4log – 0.6 log 0.6 log

E=0.9710

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

Ans. There are several advantages of using a random forest over a decision tree

Improved accuracy: Random forests typically produce more accurate predictions than individual decision trees because they combine the predictions of multiple trees.

Reduced overfitting: Because a random forest is composed of multiple decision trees, each tree is less likely to overfit the training data, resulting in improved generalization to new data.

Handling missing values: Random forests can handle missing values in the input data without the need for imputation.

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

Ans. The F-score is a metric used to evaluate the performance of a Machine Learning model. It combines precision and recall into a single score.

F-score formula**F-score = 2 \* (precision \* recall) / (precision + recall)**

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

Ans. Fit()- This method goes through the training data, calculates the parameters and saves them as internal objects.

Transform()-The parameters generated using the fit() method are now used and applied to the training data to update them.

Fit\_transform()-This method may be more convenient and efficient for modelling and transforming the training data simultaneously.