MACHINE LEARNING

1. Which of the following in sk-learn library is used for hyper parameter tuning?
   A) GridSearchCV()
   B) RandomizedCV()
   C) K-fold Cross Validation
   D) All of the above
2. In which of the below ensemble techniques trees are trained in parallel?
   A) Random forest
   B) Adaboost
   C) Gradient Boosting
   D) All of the above
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?
   A) The regularization will increase
   B) The regularization will decrease
   C) No effect on regularization
   D) kernel will be changed to linear
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)*
   Which of the following is true regarding max_depth hyper parameter?
   A) It regularizes the decision tree by limiting the maximum depth up to which a tree can be grown.
   B) It denotes the number of children a node can have.
   C) both A & B
   D) None of the above
5. Which of the following is true regarding Random Forests?
   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.
   D)None of the above
6. What can be the disadvantage if the learning rate is very high in gradient descent?
   A) Gradient Descent algorithm can diverge from the optimal solution.
   B) Gradient Descent algorithm can keep oscillating around the optimal solution and may not settle.
   C) Both of them
   D) None of them
7. As the model complexity increases, what will happen?
   A) Bias will increase, Variance decrease
   B) Bias will decrease, Variance increase
   C)both bias and variance increase
   D) Both bias and variance decrease.
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?
   A) model is underfitting
   B) model is overfitting
   C) model is performing good
   D) None of the above

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.
10. What are the advantages of Random Forests over Decision Tree?
11. What is the need of scaling all numerical features in a dataset? Name any two techniques used for scaling.

MACHINE LEARNING

12. Write down some advantages which scaling provides in optimization using gradient descent algorithm.
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?
14. What is “f-score" metric? Write its mathematical formula.
15. What is the difference between fit(), transform() and fit_transform()?

Q9 Ans: The Gini index can be calculated as follows:

Gini index = 1 - (probability of class A)^2 - (probability of class B)^2
= 1 - (0.4)^2 - (0.6)^2
= 0.48
The entropy can be calculated as follows:

Entropy = - (probability of class A) * log2(probability of class A) - (probability of class B) * log2(probability of class B)
= - 0.4 * log2(0.4) - 0.6 * log2(0.6)
= 0.9710

Therefore, the Gini index of the dataset is 0.48 and the entropy is 0.9710.

Q10 Ans:
Reduced overfitting
Better accuracy:
Robustness to noise:
Feature importance:

Q11 Ans: Scaling ensures that all features are on a similar scale, which helps the machine learning model to learn the correct weights for each feature. When the features are on different scales, the model may assign too much weight to the features with larger values, even if they are not more important than other features.

Scaling can improve the convergence of certain machine learning algorithms, such as gradient descent, by ensuring that the step sizes are similar across all dimensions.

Q12 Ans: Faster convergence: Scaling ensures that the step sizes are similar across all dimensions, which can help the algorithm to converge faster. This is because the algorithm can take larger steps in directions where the gradient is steeper and smaller steps in directions where the gradient is flatter.
Prevents oscillation: If the dimensions of the input features are on different scales, then the optimization can oscillate back and forth between the different dimensions, which can slow down convergence.
Avoids getting stuck in local minima: Scaling can help the algorithm to avoid getting stuck in local minima by ensuring that the gradient is pointing in the right direction.

Q13 Ans: In the case of a highly imbalanced dataset for a classification problem, accuracy is not a good metric to measure the performance of the model. This is because accuracy can be misleading in the presence of class imbalance.
For example, suppose we have a dataset with 95% of the samples belonging to one class and 5% belonging to another class. If we build a model that predicts all samples as belonging to the majority class, it will have an accuracy of 95%. However, this model is not useful in practice because it fails to detect the minority class.
In such cases, it is important to use other metrics that take into account the class imbalance, such as precision, recall, and F1-score.

Q14 Ans: F-Measure = (2 * Precision * Recall) / (Precision + Recall)

Q15 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.