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.

Ans: Entropy = -P(A)*log(P(A)- P(B)*logP(B)

P(A)=4/10 = 2/5 ; P(B)=6/10=3/5

Entropy = - 2/5*log(2/5) - 3/5*log(3/5)

Entropy = 0.9

Gini Index: 1-[P(A)^2 +P(B)^2] = 1 - 0.6 = 0.4

Gini Index = 0.4

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

Ans: The advantages of Random Forests over Decision Tree are:

1. **Improved accuracy:** Random forests typically produce more accurate predictions than individual decision trees because they combine the predictions of multiple trees.
2. **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.
3. **Handling missing values:** Random forests can handle missing values in the input data without the need for imputation.
4. **Handling outliers:** Random forests are less sensitive to outliers in the input data than decision trees.
5. **Feature importance:** Random forests can provide feature importance measures, which can be used to identify the most important features in the input data.
6. **Handle high dimensional data:** Random forests are less prone to the curse of dimensionality, they handle high dimensional data well.
7. **Robustness:** Random forests are more robust to noise in the data and can generalize well from noisy data.

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

Ans: Real-world datasets often contain features that are varying in degrees of magnitude, range and units. Therefore, in order for machine learning models to interpret these features on the same scale, we need to perform feature scaling.

The two techniques used for scaling are:

1. Min-Max Scaler
2. Standard Scaler

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

Ans: Because scaling (or normalizing) inputs makes gradient descent converge faster.

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: When working with imbalanced data, the minority class is our interest most of the time. Like when detecting "spam" emails, they number quite a few compared to "not spam" emails. So, the machine learning algorithms favor the larger class and sometimes even ignore the smaller class if the data is highly imbalanced. This can lead to erroneous conclusions.

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

Ans: F-score or F1 Score is a metric to evaluate a binary classification model on the basis of predictions that are made for the positive class. It is calculated with the help of Precision and Recall. It is a type of single score that represents both Precision and Recall. So, ***the F1 Score can be calculated as the harmonic mean of both precision and Recall, assigning equal weight to each of them.***

The formula for calculating the F1 score is given below:

$$F1 - score = 2 * \frac{precision * recall}{precision + recall}$$

As F-score make use of both precision and recall, so it should be used if both of them are important for evaluation, but one (precision or recall) is slightly more important to consider than the other. For example, when False negatives are comparatively more important than false positives, or vice versa.

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

Ans: The **fit(data)** method is used to compute the mean and std dev for a given feature to be used further for scaling.

The **transform(data)** method is used to perform scaling using mean and std dev calculated using the .fit() method.

The **fit_transform()** method does both fits and transform.