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

ANS: D

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

ANS: A

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

ANS: B

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

ANS: A

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

ANS: C

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

ANS:A

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.

Ans: B

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

ANS: B

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.

Gini index:

Gini(t) = 1 - ∑(i=1 to c) (p(i|t))^2

where t is a node in the decision tree, c is the number of classes, and p(i|t) is the proportion of class i at node t.

Entropy:

Entropy(t) = -∑(i=1 to c) p(i|t) * log2(p(i|t))

where t is a node in the decision tree, c is the number of classes, and p(i|t) is the proportion of class i at node t.

Gini(A) = 1 - (0.4)^2 - (0.6)^2 = 0.48

Gini(B) = 1 - (0.6)^2 - (0.4)^2 = 0.48

Gini index of the dataset = 0.4 * Gini(A) + 0.6 * Gini(B) = 0.48

Entropy:

Entropy(A) = -0.4 * log2(0.4) - 0.6 * log2(0.6) = 0.971

Entropy(B) = -0.6 * log2(0.6) - 0.4 * log2(0.4) = 0.971

Entropy of the dataset = 0.4 * Entropy(A) + 0.6 * Entropy(B) = 0.971

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

MACHINE LEARNING

The advantages of Random Forests over Decision Tree are:

Improved accuracy: Random Forests can improve the accuracy of the model compared to a single Decision Tree by combining the predictions of multiple trees.

Reduced overfitting: Random Forests are less prone to overfitting than Decision Trees, especially when the number of trees in the forest is large.

Robustness to outliers: Random Forests are robust to outliers and noise in the data, since the predictions are made based on the average of multiple trees.

Feature importance: Random Forests can provide a measure of feature importance, which can help in feature selection and understanding the underlying relationships between features and the target variable.

Efficient for large datasets: Random Forests can handle large datasets efficiently by performing parallel processing and subsetting of data.

Easy to use: Random Forests are easy to use and require minimal data preprocessing and hyperparameter tuning compared to other complex models

.

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

Scaling all numerical features in a dataset is important because it helps to ensure that all features are scale, which can be beneficial for many machine learning algorithms. If the features are not on the same scale, some algorithms may be biased towards the features with larger values, which can lead to poor performance.

Two common techniques for scaling numerical features in a dataset are:

Standardization (Z-score normalization): In this technique, the data is transformed so that it has zero mean and unit variance. This is done by subtracting the mean of the feature from each data point and then dividing by the standard deviation.

Min-Max scaling: In this technique, the data is scaled to a fixed range, usually [0, 1]. This is done by subtracting the minimum value of the feature from each data point and then dividing by the range (i.e., the difference between the maximum and minimum values.

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

Scaling provides several advantages when using the gradient descent algorithm for optimization:

Faster convergence: Scaling can help the gradient descent algorithm to converge faster by making the optimization landscape smoother and less steep.

Prevents numerical overflow or underflow: Scaling can prevent numerical overflow or underflow that can occur when multiplying very large or very small numbers in the optimization process.

Prevents oscillations: Scaling can prevent oscillations around the optimal solution that can occur when the optimization landscape is steep and the learning rate is high.

Helps with regularization: Scaling can help with regularization by ensuring that all features are on the same scale, which can prevent the optimization process from being biased towards features with larger values.

More accurate updates: Scaling can ensure that the gradient updates are more accurate by reducing the magnitude of the gradient, which can prevent the optimization process from overshooting the optimal solution.

Overall, scaling can improve the stability, speed, and accuracy of the gradient descent algorithm, leading to better optimization performance.

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?

In case of a highly imbalanced dataset for a classification problem, accuracy may not be a good metric to measure the performance of the model. This is because accuracy only takes into account the total number of correct predictions made by the model, without considering the distribution of the classes in the dataset.

For example, consider a binary classification problem with 99% of the samples belonging to class A and only 1% belonging to class B. If a model simply predicts all samples as class A, it would achieve an accuracy of 99%, which seems very high. However, such a model would be practically useless because it fails to correctly identify any samples from class B.

Therefore, when dealing with imbalanced datasets, it is often more useful to consider other metrics such as precision, recall, F1 score, or area under the ROC curve (AUC-ROC), which take into account the distribution of the classes and provide a more meaningful measure of the model's performance. These metrics focus on the performance of the model on the minority class, which is often of greater interest in imbalanced datasets.

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

F-score is a statistical metric used to evaluate the performance of a classification model. It is also known as F1 score or harmonic mean.

The F-score is the harmonic mean of the precision and recall of a classification model, where precision is the proportion of correctly identified positive instances out of the total predicted positive instances, and recall is the proportion of correctly identified positive instances out of the total actual positive instances.

The mathematical formula for F-score is as follows:

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

precision = true positives / (true positives + false positives)
recall = true positives / (true positives + false negatives)

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

In machine learning, fit(), transform(), and fit_transform() are methods used in the process of feature scaling and data preprocessing.

The fit() method is used to calculate the parameters needed for scaling or preprocessing the data. It is called on the training data to estimate the statistics of the data, such as the mean and standard deviation, which are used to transform the data. This method does not transform the data; it only calculates the parameters.

The transform() method applies the calculated parameters obtained by the fit() method to the data. It is used to scale or preprocess the data based on the calculated parameters. The transform() method is applied to both the training and testing data to ensure consistency in the scaling or preprocessing.

The fit_transform() method combines the fit() and transform() methods. It calculates the parameters needed for scaling or preprocessing the data on the training data and then applies the calculated parameters to the training data.

The main difference between the three methods is that fit() calculates the parameters and transform() applies the parameters to the data. On the other hand, fit_transform() does both operations at once. It is important to note that fit() should only be applied to the training data, while transform() and fit_transform() can be applied to both the training and testing data.