MACHINE LEARNING

ASSIGNMENT - 7

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:- Gini index = 1 - (p_A^2 + p_B^2)

Entropy = -p_A log2(p_A) - p_B log2(p_B)

p_A = 0.4

p_B = 0.6

Using these values, we can calculate the Gini index and entropy as follows:

Gini index = 1 - (0.4^2 + 0.6^2) = 0.48

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

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

ANS:- 1.Improved accuracy: Random Forests typically provide higher accuracy than Decision Trees, especially when dealing with complex datasets. By combining multiple decision trees, Random Forests reduce the risk of overfitting and improve the model's ability to generalize to new, unseen data.

2.Robustness to noise: Random Forests are less susceptible to noise and outliers in the dataset, as they aggregate the results of multiple decision trees. This reduces the impact of individual noisy data points on the model's predictions.

3.Feature importance estimation: Random Forests can estimate the importance of features in the dataset, which can be used for feature selection or to gain insights into the underlying patterns and relationships in the data.

4.Scalability: Random Forests are highly scalable and can handle large datasets with many features and instances.

5.Parallelizability: Random Forests are easily parallelizable, as the decision trees in the ensemble can be trained independently and in parallel, which makes them well-suited for distributed computing environments.

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

scaling.

ANS:- Scaling of numerical features is necessary to ensure that all features are on a similar scale or range. This is important because many machine learning algorithms are based on distance metrics, and features that are on different scales can have a disproportionate impact on the results. Scaling also helps to improve the convergence speed of some optimization algorithms, and can improve the performance of some machine learning models.

1.Min-Max scaling or normalization

2.Standardization

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

algorithm.

ANS:-

- Faster convergence
- More accurate results
- Better conditioning
- Easier to choose hyperparameters

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:- Accuracy is not a good metric for imbalanced datasets.

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.

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

ANS:- F-score, also known as the F1 score, is a measure of a model's accuracy that takes into account both precision and recall. It is a way of combining precision and recall into a single metric that balances their importance.

The mathematical formula for F-score is:

F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

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

ANS:- fit(): This method is used to estimate the parameters of a machine learning model based on the training data. When we call fit() on a training dataset, the model learns the patterns in the data and determines the values of the model's parameters that minimize the error between the predicted output and the true output. The fit() method is typically called on the training data only, and the learned parameters are stored for future use.

transform(): This method is used to transform the input data using the parameters learned by the fit() method. When we call transform() on a dataset, the model applies the learned transformation to the data. For example, in the case of feature scaling, the transform() method can be used to scale the input features to a similar range using the parameters learned by the fit() method.

fit_transform(): This method combines the fit() and transform() methods into a single step. When we call fit_transform() on a dataset, the model first learns the parameters using the fit() method and then applies the learned transformation to the data using the transform() method. This can be a useful shortcut when we want to apply a single transformation to the data.