ASSIGNMENT – 7

MACHINE LEARNING

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

A) GridSearchCV()

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

A) Random forest

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?

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.

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

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

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

C) Both of them

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

B) Bias will decrease, Variance increase

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?

B) model is overfitting

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?

Random forests typically perform better than decision trees due to the following reasons: Random forests solve the problem of overfitting because they combine the output of multiple decision trees to come up with a final prediction.

Random forests consist of multiple single trees each based on a random sample of the training data. **They are typically more accurate than single decision trees**. The following figure shows the decision boundary becomes more accurate and stable as more trees are added

Random forest algorithm avoids and prevents overfitting by using multiple trees. The results are not accurate. This gives accurate and precise results. Decision trees require low computation, thus reducing time to implement and carrying low accuracy

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

Feature Scaling is a method to transform the numeric features in a dataset to a standard range **so that the performance of the machine learning algorithm improves**. It can be achieved by normalizing or standardizing the data values

we can say that the scaling is used for **making data points generalized so that the distance between them will be lower**.

The most common techniques of feature scaling are **Normalization and Standardization**. Normalization is used when we want to bound our values between two numbers, typically, between [0,1] or [-1,1]. While Standardization transforms the data to have zero mean and a variance of 1, they make our data unitless.

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

Gradient descent is an optimization algorithm which is commonly-used to train machine learning models and neural networks. **Training data helps these models learn over time**, and the cost function within gradient descent specifically acts as a barometer, gauging its accuracy with each iteration of parameter updates.

Advanages:

1. Gradient descent is an efficient algorithm that can handle large datasets and high-dimensional parameter spaces.
2. Gradient descent can be used with a variety of loss functions and machine learning models, including linear regression, logistic regression, and neural networks.
3. Gradient descent updates the model parameters incrementally, which means it can converge to the optimal solution faster than other optimization algorithms.
4. Gradient descent can be easily parallelized, allowing for faster optimization on multiple processors.

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?

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**

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

The F-score, also called the F1-score, is **a measure of a model's accuracy on a dataset**. It is used to evaluate binary classification systems, which classify examples into 'positive' or 'negative'.

The traditional F measure is calculated as follows: **F-Measure = (2 * Precision * Recall) / (Precision + Recall)**

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

Fit() :

In the **fit()** method, where we use the required formula and perform the calculation on the feature values of input data and fit this calculation to the transformer. For applying the fit() method (fit transform in python), we have to use **fit()** in frontof the transformer object.

Transform ():

For changing the data, we probably do transform in the transform() method, where we apply the calculations that we have calculated in fit() to every data point in feature F. We have to use **.transform()** in front of a fit object because we transform the fit calculations.

Fit_transform():

The fit_transform() method is basically the combination of the fit method and the transform method. This method simultaneously performs fit and transform operations on the input data and converts the data points.Using fit and transform separately when we need them both decreases the efficiency of the model. Instead, fit_transform() is used to get both works done.