MACHINE LEARNING

ASSIGNMENT - 7

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?
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?

C) No effect on regularization

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?
C) both A & B

5. Which of the following is true regarding Random Forests?
A) It's an ensemble of weak learners.

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?
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?
ANS-
1. Random forest algorithm avoids and prevents overfitting by using multiple trees.
   WHILE
There is always a scope for overfitting, caused due to the presence of variance.in decision trees.

2. This gives accurate and precise results in random forests.
   WHILE
The results are not accurate.

11. What is the need of scaling all numerical features in a dataset? Name any two techniques used for scaling.
ANS-
Scaling can make a difference between a weak machine learning model and a better one.
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].

12. Write down some advantages which scaling provides in optimization using gradient descent algorithm.
ANS-
- We can use fixed learning rate during training without worrying about learning rate decay.
- It has straight trajectory towards the minimum and it is guaranteed to converge in theory to the global minimum if the loss function is convex and to a local minimum if the loss function is not convex.

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-

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 F-score is a way of combining the precision and recall of the model, and it is defined as the harmonic mean of the model's precision and recall.

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

ANS-

In simple language, the fit() method will allow us to get the parameters of the scaling function.

The transform() method will transform the dataset to proceed with further data analysis steps.

The fit_transform() method will determine the parameters and transform the dataset.