MACHINE LEARNING ASSIGNMENT - 7

1. Which of the following in sk-learn library is used for hyper parameter tuning?
(D) All of the above

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?
(A) The regularization will increase

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?
(A) It's an ensemble of weak learners.
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?
(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.

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.
Ans - 0.5

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

→ A **random forest** is simply a collection of **decision trees** whose results are aggregated into one final result. Their ability to limit overfitting without substantially increasing error due to bias is why they are such powerful models. One way **Random Forests** reduce variance is by training on different samples of the data.

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

- 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.
- It has unbiased estimate of gradients. The more the examples, the lower the standard error.

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?
→ If you choose the wrong metric to evaluate your models, you are likely to choose a poor model, or in the worst case, be misled about the expected performance of your model.

Choosing an appropriate metric is challenging generally in applied machine learning, but is particularly difficult for imbalanced classification problems. Firstly, because most of the standard metrics that are widely used assume a balanced class distribution, and because typically not all classes, and therefore, not all prediction errors, are equal for imbalanced classification.

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 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.

The F-score is commonly used for evaluating information retrieval systems such as search engines, and also for many kinds of machine learning models, in particular in natural language processing.

It is possible to adjust the F-score to give more importance to precision over recall, or vice-versa. Common adjusted F-scores are the F0.5-score and the F2-score, as well as the standard F1-score.

15. What is the difference between fit(), transform() and fit_transform()?
→ The **fit()** function calculates the values of these parameters. The **transform** function applies the values of the parameters on the actual data and gives the normalized value. The **fit_transform()** function performs both **in the** same step. Note that the same value is got whether we perform in 2 steps or **in a** single step.