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

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.**

   Answer-We have the formula for GINI INDEX=A/A+B

   =40/100=0.4

FLIP ROBO

MACHINE LEARNING

Entropy = -(p(A) * log(P(B)) + p(B) * log(P(B)))

40*log(60)+60* log(60)=177

10. What are the advantages of Random Forests over Decision Tree?
Answer- **Decision trees** have a low bias / are non-parametric, they suffer from a high variance which makes them less useful for most practical applications.
By aggregating multiple decision trees, we can reduce the variance of the model output significantly, thus improving performance. While this could be archived by simple tree bagging, the fact that each tree is build on a bootstrap sample of the same data gives a lower bound on the variance reduction, due to correlation between the individual trees. **Random Forest** addresses this problem by sub-sampling features, thus de-correlating the trees to a certain extend and therefore allowing for a greater variance reduction / increase in performance.

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

 Answer-
We try to change the data in such a way that the model can process it without any problems. And Feature Scaling is one such process in which we transform the data into a better version. Feature Scaling is done to normalize the features in the dataset into a finite range.
 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.
Answer- Below some advantages which scaling provides in optimization using gradient descentalgorithm.

- It makes the training faster.
- It prevents the optimization from getting stuck in local optima.
- It gives a better error surface shape.
- Weight decay and Bayes optimization can be done more conveniently.

13.In case of a highly imbalanced dataset for a classification problem, is accuracy a good metric tomeasure the performance of the model. If not, why?
Answer-

No,Accuracy **is not a good metric for imbalanced datasets** because we have an imbalanced dataset and a badly performing model which always predicts for the majority class. 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.
 Answer-
The **F-score**, also known as 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.

MACHINE LEARNING

It can be calculated by the following formula: **2 x [(Precision x Recall) / (Precision + Recall)]**.

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

   fit() : used for generating learning model parameters from training data

  transform() : parameters generated from fit() method,applied upon model to generate transformed data set.

  fit_transform() : combination of fit() and transform() api on same data set