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

Answer-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

Answer-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
   B) The regularization will decrease
   C) No effect on regularization
   D) kernel will be changed to linear

Answer- B) The regularization will decrease

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

Answer-C)both A & B

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

Answer-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

Answer-C)Both 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.

   Answer- B) Bias will decrease, Variance increase

MACHINE LEARNING

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

Answer-) 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.

Answer-

The **Gini index** of the given dataset is 0.28 & its **Entropy** is 0.97

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

Answer-

The advantages of Random Forests over Decision Tree are:

a. It reduces overfitting in decision trees and helps to improve the accuracy.
b. It works well with both categorical and continuous values.
c. It automates missing values present in the data.
d. Normalizing of data is not required as it uses a rule-based approach.
e. It is Robust to outliers.
f. It Works well with non-linear data.
g. It runs efficiently on a large dataset.
h. Better accuracy than Decision tree.

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

Answer- **Scaling** is essential for machine learning algorithms that calculate distances between data. If not scale, the feature with a higher value range starts dominating when calculating distances.

Two techniques used for scaling:

1. Min Max Scaler
2. Standard Scaler

MACHINE LEARNING

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

Answer-

Advantages of scaling in optimization using gradient descent algorithm are:

1. It makes the training faster. It prevents the optimization from getting stuck in local optima.
2. It gives a better error surface shape.
3. Weight decay and Bayes optimization can be done more conveniently.
4. It's also important to apply feature scaling if regularization is used as part of the loss function so that coefficients are penalized appropriately

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?

Answer-

In case of a highly imbalanced dataset for a classification problem accuracy is not at all good metric to measure the performance of the model achieving 90 percent classification accuracy, or even 99 percent classification accuracy, may be trivial on an imbalanced classification problem.

This means that intuitions for classification accuracy developed on balanced class distributions will be applied and will be wrong, misleading the practitioner into thinking that a model has good or even excellent performance when it, in fact, does not.

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

Answer-

In statistical analysis of binary classification, the **F-score** or **F-measure** is a measure of a test's accuracy. It is calculated from the precision and recall of the test, where the precision is the number of true positive results divided by the number of all positive results, including those not identified correctly, and the recall is the number of true positive results divided by the number of all samples that should have been identified as positive.,

F score = 2 * (precision * recall) / (precision + recall)

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

Answer-

fit() : In this method, 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 we have to use ***.fit()*** in front of the transformer object.

Suppose we initialize the Standard Scaler object O and we do .fit() then what will it do that, it takes the feature F and it will just compute the mean (μ) and standard deviation (σ) of feature F. That has happened in the fit method.

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.

We use the example that is used above section when we create an object of the fit method then we just put it in front of the .transform and transform method uses those calculations to transform the scale of the data points, and the output will we get is always in the form of sparse matrix or array.

MACHINE LEARNING

fit_transform() : This method is the combination of fit method and transform method, it is equivalent to fit().transform(). This method performs fit and transform on the input data at a single time and converts the data points. If we use fit and transform separate when we need both then it will decrease the efficiency of the model so we use **fit_transform()** which will do both the work.

Suppose, we create the Standar Scaler object, and then we perform .fit_transform() then it will calculate the mean(μ) and standard deviation(σ) of the feature F at a time it will transform the data points of the feature F.