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.
   Entropy is 0.97 bits
   Gini index = A/A+B = 0.4

MACHINE LEARNING

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

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. Decision trees have a low bias and are non-parametric, they suffer from a high variance which makes them less useful for most practical applications.

By aggregating multiple decision trees, one 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 for scaling.

In many machine learning algorithms, to bring all features in the same standing, we need to do scaling so that one significant number doesn't impact the model just because of their large magnitude.

Feature scaling in machine learning is one of the most critical steps during the pre-processing of data before creating a machine learning model. Scaling can make a difference between a weak machine learning model and a better one.

Two most common techniques of feature scaling are Normalization and Standardization.

MACHINE LEARNING

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

    Gradient descent is an optimization algorithm used to minimize the cost function in machine learning algorithms like Logistic Regression, SVM, Neural Networks etc. If features are on different scale, certain weights are updated faster than others in Gradient Descent. However, feature scaling helps in causing Gradient Descent to converge much faster as standardizing all the variables on to the same scale.

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?

    Accuracy Metric is one the simplest and widely used metric to measure the performance of a classification predictive model. The reason for its wide use is because it is easy to calculate, easy to interpret, and is a single number to summarize the model's capability. However, accuracy metric fails to perform on an imbalanced dataset as it gives misleading conclusions. In an imbalanced dataset getting an accuracy score of 90 or 99 are trivial as model might have considered the less numbered observation as error or outliers and could have ignored them in the prediction.

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

    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'.

    $F1=2 * \frac{*precision*recall}{precision+recall}$

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

    Fit() method is used to fit the transformer like MinMaxSCaler to the input data and perform the required computations to the specific transformer we apply.

    Transform() method of sklearn transformers, will transform the input data into some transformed spaced. The output is usually an array matrix with equal number of samples as the input data. The transformation will be performed based on the parameters that were computed during fit.

    Fit_transform() method is basically 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.