MACHINE LEARNING- Assignment 07

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

ANS: A) GridSearchCV()

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

ANS: 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

ANS: 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

ANS: 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

ANS: 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. C) Both of them D) None of them.

ANS: 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.

ANS: 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

ANS: 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.

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

ANS:

Sr. No	Random Forest	Decision Tree
1.	While building a random forest the number of rows are selected randomly.	Whereas, it built several decision trees and find out the output.
2.	It combines two or more decision trees together.	Whereas the decision is a collection of variables or data set or attributes.
3.	It gives accurate results.	Whereas it gives less accurate results.
4.	By using multiple trees it reduces the chances of overfitting.	On the other hand, decision trees, it has the possibility of overfitting, which is an error that occurs due to variance or due to bias.
5.	Random forest is more complicated to interpret.	Whereas, the decision tree is simple so it is easy to read and understand.
6.	In a random forest, we need to generate, process, and analyze trees so that this process is slow, it may take one hour or even days.	The decision tree is not accurate but it processes fast which means it is fast to implement.
7.	It has more computation because it has n number of decision trees, so more decision trees more computation.	Whereas it has less computation.
8.	It has complex visualization, but it plays an important role to show hidden patterns behind the data.	On the other hand, it is simple to visualize because we just need to fit the decision tree model.
9.	The classification and regression problems can be solved by using random forest.	Whereas a decision tree is used to solve the classification and regression problems.
10.	It uses the random subspace method and bagging during tree construction, which has built-in feature importance.	Whereas a decision is made based on the selected sample's feature, this is usually a feature that is used to

		make a decision, decision tree learning is a process to find the optimal value for each internal tree node.

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

ANS: In Data Processing, 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.

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

ANS: Some advantages of batch gradient descent are its computational efficiency: it produces a stable error gradient and a stable convergence. Some disadvantages are that the stable error gradient can sometimes result in a state of convergence that isn't the best the model can achieve.

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: 4 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'.

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()?

ANS: The fit(data) method is used to compute the mean and std dev for a given feature to be used further for scaling. The transform(data) method is used to perform scaling using mean and std dev calculated using the. fit () method. The fit_transform () method does both fits and transform.

THE END