

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 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:- D) All of 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

Answer:-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.
 - 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:- 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

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 increases

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: B) Model is overfitting

MACHINE LEARNING

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?

Answer:- The random forest has complex visualization and accurate predictions, but the decision tree has simple visualization and less accurate predictions. The advantages of Random Forest are that it prevents overfitting and is more accurate in predictions. Decision trees require low computation, thus reducing time to implement and carrying low accuracy where as the random forest consumes more computation. The process of generation and analyzing is time-consuming. The decision tree is easy to visualize. The only task is to fit the decision tree model where as random forest This has complex visualization as it determines the pattern behind the data. Decision tree The results are not accurate where as the random forest gives accurate and precise results. Decision tree there is always a scope for overfitting, caused due to the presence of variance. Random forest algorithm avoids and prevents overfitting by using multiple trees.

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

Answer:- Scaling is required to rescale the data and it's used when we want features to be compared on the same scale for our algorithm. And, when all features are in the same scale, it also helps algorithms to understand the relative relationship better. This means that you're transforming your data so that it fits within a specific scale, like 0-100 or 0-1. 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.

Answer:- To ensure that the gradient descent moves smoothly towards the minima and that the steps for gradient descent are updated at the same rate for all the features, we scale the data before feeding it to the model. Having features on a similar scale can help the gradient descent converge more quickly towards the minima.

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:- No. 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:- An F-score is the harmonic mean of a system's precision and recall values. It can be calculated by the following formula: $2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}$.

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

Answer:- 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.