

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:- A

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:- B

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

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:-D

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

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

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

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

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.

Answer:- Calculation. The Gini Index or Gini Impurity is calculated by subtracting the sum of the squared probabilities of each class from one. It favours mostly the larger partitions and are very simple to implement.

Gini index (a criteria to minimize the probability of misclassification):

$$\text{Gini} = 1 - \sum p_j^2$$

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

Answer:- Random forest algorithm **avoids and prevents overfitting by using multiple trees**. The results are not accurate. This gives accurate and precise results. Decision trees require low computation, thus reducing time to implement and carrying low accuracy.

A decision tree combines some decisions, whereas a random forest combines several decision trees. Thus, **it is a long process, yet slow. Whereas, a decision tree is fast and operates easily on large data sets, especially the linear one. The random forest model needs rigorous training.**

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

Answer:- Feature Scaling is a technique to standardize the independent features present in the data in a fixed range. It is performed during the data pre-processing to handle highly varying magnitudes or values or units. If feature scaling is not done, then a machine learning algorithm tends to weigh greater values, higher and consider smaller values as the lower values, regardless of the unit of the values.

Techniques to perform Feature Scaling
consider the two most important ones:

- **Min-Max Normalization:** This technique re-scales a feature or observation value with distribution value between 0 and 1.
- **Standardization:** It is a very effective technique which re-scales a feature value so that it has distribution with 0 mean value and variance equals to 1.

MACHINE LEARNING

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

Answer:- Gradient descent is an optimization algorithm which is commonly-used to train machine learning models and neural networks. **Training data helps these models learn over time**, and the cost function within gradient descent specifically acts as a barometer, gauging its accuracy with each iteration of parameter updates.

Gradient Descent is an iterative optimization algorithm, used to find the minimum value for a function. The general idea is to initialize the parameters to random values, and then take small steps in the direction of the "slope" at each iteration. Gradient descent is highly used in supervised learning to minimize the error function and find the optimal values for the parameters.

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:- 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.

Answer:- 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'. 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(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.**