

MACHINE LEARNING

- 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
- 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
- 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
- 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
- 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
- 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
- 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.
- 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.

Ans: Gini index: $p(A)(1-p(A)) + p(B)(1-p(B)) = 40\%(1-40\%) + 60\%(1-60\%) = 0.24$

Entropy: $-p(A) \cdot \log_2(p(A)) - p(B) \cdot \log_2(p(B)) = -40\% \cdot \log_2(40\%) - 60\% \cdot \log_2(60\%) = 0.97$

MACHINE LEARNING

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

Ans: The advantages of Random Forests over Decision Tree are:

- It overcomes the problem of overfitting by averaging or combining the results of different decision trees.
- Random forests work well for a large range of data items than a single decision tree does.
- Random forest has less variance than single decision tree.
- Random forests are very flexible and possess very high accuracy.
- Scaling of data does not require in random forest algorithm. It maintains good accuracy even after providing data without scaling.
- Random Forest algorithms maintains good accuracy even a large proportion of the data is missing.

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

Ans: Feature scaling is the final step of data preprocessing in machine learning. It is a technique to standardize the independent variables of the dataset in a specific range. In feature scaling, we put our variables in the same range and in the same scale so that no any variable dominate the other variable.

The two techniques use for scaling are:

- Min-Max Scaling
 - Normalization
-

MACHINE LEARNING

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

Ans: The advantages which scaling provides in optimization using gradient descent algorithm are:

- It helps gradient descent move smoothly towards the minima and the steps for gradient descent are updated at the same rate for all the features.
- It helps 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?

Ans: In case of a highly imbalanced dataset for a classification problem, accuracy is not a good metric to measure the performance of the model because accuracy is computed by dividing the number of correct predictions to total predictions. As the majority class is over-represented, the classifier may predict the majority class most of the time and still have a high accuracy.

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

Ans: f-score is a harmonic mean between recall and precision. Its range is [0,1]. This metric usually tells us how precise and robust our classifier is.

It is used to measure the test's accuracy.

The mathematical formula for f-score is:

$$F\text{-score} = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}).$$

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

Ans: In machine learning, fit(), transform() and fit_transform() are methods of the scikit-learn library used for preprocessing data:

- fit() method is used to fit the data to the model, it is used to calculate the internal parameters of the model.
- transform() method is used to transform the data according to the internal parameters calculated during the fit() method.
- fit_transform() method is used to fit the data to the model and then transform it in one step