Machine learning Worksheet-7

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 : 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

Answer: 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

Answer: 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

Answer: It regularizes the decision tree by limiting the maximum depth up to which a tree can be grown

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: It's an ensemble of weak learners

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 : Gradient Descent algorithm can diverge from the optimal solution

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: 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

Answer: model is overfitting

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 : Gini index - 1 - (p_A^2 + p_B^2)

where p_A and p_B are the proportions of class A and B, respectively ,i.e p_A=40/100 = 0.4 & p_B= 60/100 = 0.6

Gini Index = 1 - (0.4^2 + 0.6^2) = 1 - 0.52 = 0.48

So , Gini index = 0.48

Entropy = - p_A log2(p_A) - p_B log2(p_B)

Entropy = -0.4 log2(0.4) - 0.6 log2(0.6) = 0.971

So , Entropy = 0.971.

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

Answer : Random Forests can provide information about which features are important for the classification. By looking at the importance scores of each

feature, we can identify the most relevant variables and use them for feature selection.

By combining multiple trees, Random Forests can provide higher accuracy than a single Decision Tree. The individual trees may have errors, but the ensemble of trees reduces these errors, leading to improved performance.

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

Answer : Scaling all numerical features in a dataset is important because it ensures that all features have a comparable scale and range, which can be important for certain machine learning algorithms that are sensitive to the scale of the input features. Scaling also prevents features with larger magnitudes from dominating over other features.

Two techniques used for scaling are :

1. Standardization : This technique scales the features to have zero mean and unit variance, which means that the values are centered around 0 with a standard deviation of 1. The formula for standardization is:

x_scaled = (x - mean(x)) / std(x)

2. Min_Max Scaling : his technique scales the features to a fixed range, usually between 0 and 1, by subtracting the minimum value and dividing by the range of the feature. The formula for Min-Max scaling is:

x_scaled = (x - min(x)) / (max(x) - min(x))

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

Answer : Scaling the features can improve the condition number of the cost function, which is a measure of how sensitive the output is to changes in the

input. This can make the optimization more stable and less sensitive to small changes in the input.

Scaling can make it easier to visualize the data by bringing all the features to a similar scale. This can help in identifying patterns and relationships between the features.

Overall, scaling the features can improve the performance of optimization algorithms such as gradient descent, leading to faster convergence and more accurate results.

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 : In the case of a highly imbalanced dataset, accuracy may not be a good metric to measure the performance of a classification model. This is because accuracy measures the proportion of correctly classified instances out of the total number of instances, without considering the class distribution.

Precision measures the proportion of true positives among all the positive predictions, recall measures the proportion of true positives among all actual positive instances, and F1 score is a harmonic mean of precision and recall. AUC-ROC measures the performance of the model across all possible decision thresholds and is more robust to imbalanced class distribution.

Therefore, it is important to select the appropriate metric based on the nature of the dataset and the problem being solved, especially when dealing with imbalanced datasets.

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

Answer : The F-score is a metric commonly used in binary classification problems that takes into account both precision and recall. It is the harmonic mean of precision and recall and provides a balanced measure of the two.

The mathematical formula for F-score is:

F-score = 2 * (precision * recall) / (precision + recall)

where precision is the ratio of true positives to the sum of true positives and false positives, and recall is the ratio of true positives to the sum of true positives and false negatives.

Precision = true positives / (true positives + false positives)

Recall = true positives / (true positives + false negatives)

The F-score ranges from 0 to 1, with a value of 1 indicating perfect precision and recall. A higher F-score indicates better performance of the classification model

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

Answer : **fit()**: This method is used to learn the parameters of a transformation on the data. It takes in the input data as its argument and learns the parameters needed to transform the data. For example, in the case of scaling the data, fit() will learn the mean and standard deviation of each feature in the data.

transform(): This method is used to apply a learned transformation to the data. It takes in the input data as its argument and applies the learned transformation to it. For example, in the case of scaling the data, transform() will apply the previously learned mean and standard deviation scaling to the data.

fit_transform(): This method is used to both learn the parameters of a transformation and apply that transformation to the data in a single step. It takes in the input data as its argument, learns the parameters needed to transform the data, and applies the transformation to the data.