ASSIGNMENT-7

MACHINE LEARNING- 7

1. Which of the following in sk-learn library is used for hyper parameter tuning?

D) All of the above

2. In which of the below ensemble techniques trees are trained in parallel?

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?

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.

5. Which of the following is true regarding Random Forests?

A) It's an ensemble of weak learners.

6. What can be the disadvantage if the learning rate is very high in gradient descent?

C) Both of them

7. As the model complexity increases, what will happen?

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?

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.

The Gini index and entropy are both measures of impurity or disorder in a dataset. They are used to evaluate the quality of a split in decision tree algorithms and other classification algorithms.

To calculate the Gini index of a dataset, we use the following formula: Gini = 1 - (p(A)^2 + p(B)^2) where p(A) and p(B) are the probability of class A and class B respectively.

In this case, p(A) = 40% and p(B) = 60%. Substituting these values into the formula we get: Gini = 1 - (0.4^2 + 0.6^2) = 1 - (0.16 + 0.36) = 1 - 0.52 = 0.48

To calculate the entropy of a dataset, we use the following formula: Entropy = - p(A)*log2(p(A)) - p(B)*log2(p(B)) where p(A) and p(B) are the probability of class A and class B respectively.

In this case, p(A) = 40% and p(B) = 60%. Substituting these values into the formula we get: Entropy = -0.4*log2(0.4) - 0.6*log2(0.6) = -0.4*(-0.602) - 0.6*(-0.778) = 0.602 + 0.778 = 0.98

Therefore, the Gini index of the dataset is 0.48 and the entropy of the dataset is 0.98

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

1. Improved Performance: Random Forests generally have better predictive performance than a single decision tree because they reduce overfitting by averaging the predictions of multiple trees.
2. Reduced Variance: Random Forests reduce the variance of the model by averaging the predictions of multiple trees. This makes the model more robust and less likely to be affected by outliers or noise in the data.
3. Improved Generalization: Random Forests improve the generalization of the model by averaging the predictions of multiple trees. This makes the model less sensitive to the specific training data and more likely to perform well on new data.

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

1. Algorithms that use distance-based measures, such as k-nearest neighbors and support vector machines, are sensitive to the scale of the features. Scaling the features ensures that all features are on a similar scale, which can improve the performance of these algorithms.
2. Some machine learning algorithms, such as gradient descent and linear regression, can be sensitive to the scale of the features. Scaling the features can help to reduce the time required to converge and improve the performance of these algorithms.
3. Scaling the features can also help to reduce the impact of outliers on the model.

The two techniques are normalization and standardization

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

1. Faster convergence: Scaling the features can help to reduce the time required for the algorithm to converge. By scaling the features, the optimization process will have to take smaller steps and make fewer iterations to reach the optimal solution.

2. Improved optimization: Scaling the features can help to improve the optimization process by reducing the curvature of the objective function. This can make the optimization process more efficient, as the optimizer will be able to make larger steps in the direction of the gradient.
3. Robustness: Scaling the features can make the optimization process more robust by reducing the impact of outliers or large values. This can help to prevent the optimizer from getting stuck in local minima or diverging from the optimal solution.

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?

Accuracy is not always a good metric to measure the performance of a model on a highly imbalanced dataset, because accuracy can be misleading in such cases.

When the dataset is highly imbalanced, meaning that one class has significantly more examples than the other class, a model can still achieve a high accuracy by simply predicting the majority class all the time, without ever identifying the minority class. This is known as the "accuracy paradox".

Instead, other metrics such as precision, recall, F1-score, and the area under the Receiver Operating Characteristic (ROC) curve are often used to evaluate the performance of a model on an imbalanced dataset. These metrics take into account the imbalance in the dataset and give a better indication of how well the model is performing on the minority class.

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

The F1-score is defined as the harmonic mean of precision and recall, where precision is the number of true positive predictions divided by the number of true positive and false positive predictions, and recall is the number of true positive predictions divided by the number of true positive and false negative predictions.

The mathematical formula for F1-score is:

F1-score = 2 * (precision * recall) / (precision + recall)

It ranges between 0 and 1, where 1 represents perfect precision and recall, and 0 represents the worst possible performance.

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

- **fit()** is used to estimate the parameters of a model from training data. For example, when using a preprocessing method such as StandardScaler, the **fit()** method is used to compute the mean and standard deviation of the training data, which are then used to standardize the data.
- **transform()** is used to apply a preprocessing method to a dataset. For example, after fitting a StandardScaler to the training data, the **transform()** method can be used to standardize the test data using the same mean and standard deviation values.
- **fit_transform()** is a convenience method that combines the **fit()** and **transform()** methods into a single function. It first fits the preprocessing method to the training data, and then applies the transformation to the training data and returns the transformed data. This is useful when you want to fit and transform the data in one step